FOXA1 is required for ErbB2 expression and luminal differentiation in HER2-positive breast cancer.

[1]**Running title:** FOXA1 in HER2-positive cancer.

Jaekwang Jeong[1, 10], Jongwon Lee[2, 3, 10], Jaechul Lim[4], Jaehun Shin[5], Kwangmin Yoo[2], Jonghun Kim[6], Yoshiaki Tanaka[7], Hyun Seop Tae[8], Lark Kyun Kim[9], In-Hyun Park[6], John Wysolmerski[1], Jungmin Choi[2]

1. Section of Endocrinology and Metabolism, Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA
2. Department of Biomedical Sciences, Korea University College of Medicine, Seoul, Republic of Korea
3. Brain Korea 21 Plus Project for Biomedical Science, Korea University College of Medicine, Seoul, Korea.
4. College of Veterinary Medicine and Research Institute for Veterinary Science, Seoul National University, Seoul 08826, Republic of Korea
5. Integrated Science Engineering Division, Underwood International College, Yonsei University, Republic of Korea
6. Department of Genetics, Yale Stem Cell Center, Yale School of Medicine, New Haven, CT 06520, USA
7. Department of Medicine, Maisonneuve-Rosemont Hospital Research Centre, University of Montreal, Montreal, Quebec, H1T 2M4, Canada
8. Elgen Therapeutics, 193 Munji-Ro, Yuseong-Gu, Daejeon 34051, South Korea
9. Department of Biomedical Sciences, Graduate School of Medical Science, Brain Korea 21 Project, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul 06230, Republic of Korea
10. These authors contribute equally.

Corresponding authors
Jaekwang Jeong, email: jaekwang.jeong@yale.edu
John Wysolmerski, email: john.wysolmerski@yale.edu
Jungmin Choi, email: jungminchoi@korea.ac.kr

Abstract

Forkhead box protein A1 (FOXA1), a pioneering transcriptional factor known for its critical roles in prostate and ERα–positive breast cancer, is also expressed in human epidermal growth factor receptor-2 (HER2/ErbB2)-positive breast cancers. However, its role in HER2-pos tumors is less well understood. Here we investigate the function of FOXA1 in HER2/ErbB2-positive breast cancers. The loss of FOXA1 was associated with a marked decrease in the viability of HER2-positive and HER2 amplified cell lines, suggesting a pivotal involvement of FOXA1 in these breast cancers. Employing patient-derived single-cell RNA sequencing and spatial transcriptomics, we demonstrate that *FOXA1* is co-expressed with *ErbB2* in HER2-positive breast cancers. Suppression of FOXA1 expression led to the reduction of HER2 expression and signaling. Chromatin Immunoprecipitation Sequencing (ChIP-seq) and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) identified FOXA1 binding motifs in the *ErbB2* promoter and regulatory element regions, which controlled *ErbB2 gene* expression. Notably, FOXA1 knockdown was observed to enhance Epithelial-Mesenchymal Transition (EMT) signaling and impede luminal tumor differentiation. Furthermore, we find that FOXA1 and TRPS1 combine to regulate TEAD/YAP-TAZ activity. Taken together, these findings highlight the essential role of FOXA1 in maintaining HER2 expression and a luminal cell phenotype in HER2-positive breast cancers.

Key words: Breast Cancer, FOXA1, HER2, ErbB2, Epithelial Mesenchymal Transition (EMT), TEAD/YAP, TRPS1

Introduction

Forkhead box protein A1 (FOXA1) is a DNA-binding protein and functions as a pioneering transcription factor activating androgen receptor (AR) expression in prostate cancer and estrogen receptor α (ERα/*ESR1*) expression in breast cancer. FOXA1 binds to DNA and induces an open chromatin conformation that facilitates the binding of tissue-specific transcription factors such as the AR in prostate cancer cells, which has been associated with tumor progression and de-differentiation in prostate cancer (1-3). Similarly, in ER-positive breast tumors, FOXA1 facilitates the accessibility of ERα to gene promoters (3), maintaining an ERα dependent program of gene expression that defines the canonical luminal, ER-positive tumor phenotype (4, 5). Furthermore, although ERα expression usually predicts a better prognosis, high levels of FOXA1 expression in ER-positive tumors can promote cellular proliferation, migration, tumor progression and a poor prognosis (6).

ErbB2/HER2/Neu (HER2) is a receptor tyrosine kinase within the ErbB family that is commonly overexpressed in breast and other cancers. Approximately 20% of invasive breast cancers have traditionally been considered as HER2-positive, and elevated tumor levels of HER2 correlate with more aggressive behavior and increased mortality in breast cancer patients (7-9). Experimental overexpression of HER2 in mammary epithelial cells in MMTV-Neu transgenic mice is sufficient to produce breast cancers *in vivo* demonstrating its transforming ability (10, 11). Interestingly, despite low levels of ERα expression, FOXA1 is expressed in many HER2-positive tumors, where its expression correlates with higher rates of cell proliferation (12). It has been proposed that there is cross-regulation between HER2 and FOXA1 in ERα-negative breast tumors involving several other transcription factors, including RELB, PAK1, and TLE3 (13).

However, it is not known whether FOXA1 interacts directly with the *ErbB2* gene or with HER2 protein.

FOXA1 is also expressed in normal mammary epithelial cells, where it contributes to luminal cell fate determination. Specifically, co-expression of FOXA1 with both ERα and another transcription factor, GATA3, has been shown to be necessary for mammary progenitor cells to differentiate into luminal epithelial cells. Studies suggest that GATA3 upregulates FOXA1 expression to maintain ER-alpha expression and to establish and maintain luminal cell fate (14). Ablation of either FOXA1 or GATA3 impairs luminal epithelial cell differentiation (4, 5, 15). GATA3 expression is also associated with ERα expression and a luminal cell phenotype in ER-positive breast cancers. However, unlike FOXA1, in HER2-positive breast cancers, there is a negative correlation between HER2 and GATA3 expression, and HER2-positive tumors with low GATA3 expression exhibit increased cellular proliferation and more aggressive tumor progression (16). However, the molecular mechanisms regulating GATA3 expression in HER2-positive tumors are not known.

The Hippo pathway regulates normal organ development and plays diverse roles in cancer development (17). The Hippo pathway transcriptional coactivators YAP and TAZ, interact with the TEAD/TEF family of transcription factor and lead to tumor growth and metastasis (18). Recently, it has shown that TRPS1 which is transcription factor belong to the GATA family represses YAP/TAZ-dependent transcription in HER2-negative breast cancer cell lines (19). However, interactions between FOXA1 and hippo pathways in HER2-positive breast cancer remains unclear.

In this manuscript we present evidence that FOXA1 regulates HER2 expression in HER2-positive tumors by binding to the *ErbB2* promoter and regulatory regions. Furthermore,

we demonstrate that FOXA1 contributes to the maintenance of a luminal tumor identity in HER2-positive tumor cells in the setting of low GATA3 expression and functions as an upstream regulator of a TRPS1-TEAD/YAP-TAZ axis.

Results

FOXA1 is co-expressed with *ErbB2* in HER2-positive breast cancer cells.

FOXA1 is well known to promote ERα gene expression and is necessary for the ER-responsive transcriptome of canonical ER-positive luminal breast cancers (3, 6). However, FOXA1 is also expressed in HER2-positive breast cancers with low levels of ERα expression. To assess the effects of knocking out FOXA1 in both ER-positive and HER2-positive breast cancer, we examined a comprehensive series of cell lines (n = 1,225) derived from various cancer types within the Cancer Dependency Map (20, 21). This analysis revealed that both breast and prostate cancer cells rely on FOXA1 for viability, indicated by aggregate gene effect scores around -1 in the absence of FOXA1 (Supplementary Fig. 1A and Supplementary Table 1). This dependency was observed in both ER-positive and/or HER2-positive breast cancer cell lines (Supplementary Fig. 1B and Supplementary Table 2). Notably, FOXA1 inactivation led to reduced cell viability in HER2-positive cell lines, indicating a potentially important function in HER2-positive breast cancer cells (Supplementary Fig. 1C). Therefore, we explored the potential distinct functions of FOXA1 in HER2-positive breast cancer cells as compared to those in ER-positive breast cancer cells.

We first analyzed single-cell RNA sequencing (scRNA-seq) data from 19 human breast tumors, comprising 6 HER2-positive and 13 ER-positive tumors (22). A total of 148,945 cells passed quality control measures, and three major cell types were identified: epithelial cells, immune cells, and stromal cells (Fig. 1A, Supplementary Fig. 2A), with cell type assignments confirmed using lineage-specific marker genes (Supplementary Fig. 2B-D). As expected, we observed high *ESR1* and low *ErbB2* expression in ER-positive cells, whereas HER2-positive

cells displayed the opposite pattern. Subsequently, we compared the expression patterns of *FOXA1* between ER-positive and HER2-positive cells. Interestingly, *FOXA1* expression was observed in both ER-positive and HER2-positive epithelial cells (Fig. 1B). The correlation analysis revealed a statistically significant strong positive correlation between *FOXA1* and *ESR1* in ER-positive cells [R = 0.79 (Pearson), P-value = 9.63e-186], but weak correlation [R = 0.19 (Pearson), P-value = 2.54e-178] in HER2-positive cells. Conversely, there was a strong positive correlation between *FOXA1* and *ErbB2* in HER2-positive epithelial cells [R = 0.66 (Pearson), P-value = 2.10e-276], while their correlation in ER-positive cells was weak [R = 0.44 (Pearson), P-value = 2.40e-193] (Fig. 1C). In addition, 69.3% of HER2-positive cells co-expressed *FOXA1* and *ErbB2*, compared to only 9.8% of ER-positive cells. In contrast, 45.2% of ER-positive cells co-expressed *ESR1* and *FOXA1*, which was true for only 22.1% of HER2-positive cells (Fig. 1D and E).

To elucidate the spatial gene expression patterns of *ErbB2*, ESR1, and *FOXA1* in breast cancer cells, we employed two distinct and publicly available spatial transcriptomics platforms, Visium and Xenium, on formalin-fixed paraffin-embedded (FFPE) human breast cancer sections (Stage II-B, ER-positive/Progesterone receptor(PR)-negative/HER2-positive) (23). The results highlighted the spatial distribution of *ErbB2*, *ESR1*, and *FOXA1* expression (Fig. 1F-G and Supplementary Fig 3A). In the absence of a subset of normal epithelial cells in these tissue sections, myoepithelial cells served as the control group (24). The expression patterns revealed a substantial overlap of *ErbB2* and *FOXA1* in both tumor subtypes, with *ESR1* showing partial expression. Two distinct tumor subtypes were identified and annotated as *ErbB2+*/*ESR1-* and *ErbB2+*/*ESR1+*, respectively (Fig. 1H, Supplementary Fig. 3A). A subsequent comparison of gene expression between the control group and the two tumor subtypes revealed higher FOXA1

expression in tumor cells compared to control cells (Fig. 1I, Supplementary Fig. 3B). A strong positive correlation [R = 0.865 (Pearson), p-value = 1.0e-193] between *FOXA1* and *ErbB2* was observed across all cell types. Notably, *ErbB2*+/*ESR1*- tumor cells exhibited the highest co-expression density for *ErbB2* and *FOXA1*, aligning with the patterns seen in Figure 1C (Fig. 1J).

FOXA1 is required for *ErbB2* expression in HER2-positive breast cancer cells.

To examine whether FOXA1 affects HER2 expression and/or HER2 signaling, we stably knocked down FOXA1 in the HER2-positive SKBR3 breast cancer cell line (FOXA1KD-SKBR3) (Fig. 2A and B). This knockdown led to a significant reduction in *ERBB2* transcripts and protein in FOXA1KD-SKBR3 cells (Fig. 2C and D). Furthermore, we found a dramatic decrease in the level of phosphorylated HER2 (pHER2) and phosphorylated EGFR (pEGFR) (Fig. 2D and E). These findings suggest that FOXA1 plays a role in regulating HER2 expression and signaling in HER2-positive breast cancer.

To identify direct transcriptional targets of FOXA1, we performed chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) in SKBR3 and MCF7 (ER-positive) breast cancer cell lines (25). These cell lines express higher levels of FOXA1 than non-transformed human mammary epithelial cells (MCF10A) (Fig. 2F and G). The ChIP-seq analysis identified 30,500 peaks, corresponding to 10,471 genes in MCF7 cells, and 11,887 peaks, corresponding to 5,155 genes in SKBR3 cells. The fraction of peaks across different genomic features displayed a similar distribution in both cell lines (Fig. 2H). Of the total number of peaks, 3,480 were mapped to promoter-transcription start site (TSS) regions (Fig. 2I). Interestingly, more than 80% of the peaks fell outside the promoter regions, such as in intergenic and intronic regions. This suggests that the regulatory mechanisms are more complex than just

promoter activation, and FOXA1 might influence gene expression through distant regulatory elements, possibly via enhancers or silencers located in the non-promoter regions. An analysis revealed 617 promoter-TSS regions that were shared between SKBR3 and MCF7 cells, but 2,577 peaks in MCF7 cells and 286 peaks in SKBR3 cells were mutually exclusive (Fig. 2I). This analysis highlights the intricate regulatory functions of FOXA1 in breast cancer cell lines and the complexity of transcriptional regulation in different cellular contexts. We conducted an overrepresentation analysis using enrichR to examine the functional relevance of the genes at FOXA1 binding sites identified by ChIP-seq in the breast cancer cell lines. We concentrated on genes associated with promoter-TSS regions and analyzed the 1,987 genes derived from the 2,577 unique peaks found in MCF7 cells. Our analysis targeting genes at FOXA1 binding sites exclusive to MCF7 cells revealed a significant enrichment in specific pathways, including 'Estrogen response,' 'TGF-beta receptor complex,' and the 'Transcriptional activity of SMAD2/SMAD3', underscoring their unique biological functions in these cells. In contrast, no significantly enriched pathways were identified when we assessed the 492 genes that were shared between SKBR3 and MCF7 cell lines, as well as the 216 genes unique to SKBR3. This suggests that while FOXA1 plays a role in both cell lines, its interactions and effects are specific to the cellular context and influenced by the cellular signaling environment.

To deepen our understanding of FOXA1's context-dependent role, we extended our gene selection beyond promoter-transcriptional start site (TSS) to include FOXA1's binding peaks located in introns, exons, and transcription termination sites (TTSs). This comprehensive analysis identified cell line-specific and exclusive peaks alongside their annotated genes based on a score ratio greater than 2. We discovered 6,401 peaks and 3,152 annotated genes in SKBR3 cells, contrasted with 9,211 peaks and 4,241 annotated genes in the MCF7 cells. Notably, an

overlap of 2,246 peaks and 1,446 annotated genes with a score ratio below 2 was observed (Supplementary Fig. 5A). Subsequent overrepresentation analysis, using enrichR against the WikiPathways (26), highlighted EGF/EGFR signaling pathway as the most enriched pathway among genes associated with exclusively identified peaks in SKBR3 cells (Supplementary Fig. 5B). In addition, "EGFR UP.V1 UP" emerged as the most overrepresented gene set within these unique peaks for SKBR3 cells (Supplementary Fig. 5C), suggesting a specific interaction between FOXA1 and receptor tyrosine kinase (RTK) signaling within this cellular context. Interestingly, the "EGFR UP. V1 UP" gene set was also identified among the 2,246 overlapping peaks, incorporating 1,446 annotated genes (Supplementary Fig. 5D). Therefore, the global enrichment analysis of ChIP-seq data delineate FOXA1's specific involvement in modulating cell-signaling pathways according to the cellular context.

Next, we assessed how FOXA1 contributes to the expression of *ErbB2* and *ESR1* in the two cell lines. In the SKBR3 cells, we identified 2 distinct narrow FOXA1 ChIP-seq peaks within the *ErbB2* gene region. These ChIP-seq peaks were not in the canonical *ErbB2* proximal promoter region (Fig. 2J). The first peak (1) was located in an upstream *ErbB2* promoter region, annotated in the eukaryotic promoter database (EPD) as either an *ErbB2* or *PGAP3* promoter region (Fig. 2J). The second peak (2), found downstream of TSS of the canonical *ErbB2* transcript, defines a potential downstream regulatory region (Fig 2J). Chromatin accessibility analysis using ATAC-seq in SKBR3 cells identified two regions within the *ErbB2* gene that overlapped with the FOXA1 ChIP-seq peaks, suggesting these regions are open for transcriptional machinery and may be necessary for the regulation of *ErbB2* expression in these cells. These results suggest that FOXA1 increases the accessibility of the *ErbB2* gene for the

transcriptional machinery, thereby directly regulating the expression of the *ErbB2* gene in SKBR3 cells.

Interestingly, the peaks identified in the ErbB2 gene by both ChIP-seq and ATAC-seq in SKBR3 cells were absent in MCF7 cells (Fig. 2J). In MCF7 cells, we detected two FOXA1 ChIP-seq peaks in the ESR1 gene's promoter regions, coinciding with ATAC-seq peaks that suggest chromatin accessibility (Fig. 2K). However, these specific ChIP-seq and ATAC-seq peaks were not observed in the *ESR1* gene in SKBR3 cells (Fig. 2K). The FOXA1 binding sites identified within the ErbB2 and ESR1 genes through ChIP-seq analysis exhibited JASPAR scores ranging from 304 to 487, supporting the specificity of FOXA1 binding at these sites (Supplementary Fig. 4A and B). To ascertain the role of FOXA1 in enhancing chromatin accessibility at the identified sites within the *ErbB2* promoter, ATAC-seq analyses were performed on the *ErbB2* gene in FOXA1KD-SKBR3 cells compared to controls. The results displayed a significant reduction in the peaks upon FOXA1 knockdown, confirming that FOXA1 directly regulates chromatin accessibility in the promoter and downstream potential regulatory regions of the *ErbB2* gene (Fig. 2L).

We next focused on the differential gene expression in FOXA1KD-SKBR3 cells compared to control SKBR3 cells, utilizing bulk RNA-seq to uncover these variations. Our application of GSEA against Oncogenic signature gene sets (C6) revealed a significant enrichment of down-regulated genes in the FOXA1KD-SKBR3 cells within the HER2 signaling pathways. Notably, the “ERBB2_UP.V1_UP” gene set, encompassing 190 genes up-regulated by constitutively active c-erbB-2, showed a marked enrichment (enrichment score = -0.481, adjusted p-value = 0.027) (Fig. 2M) (27). Additionally, gene sets liked to HER2 signaling activation, including “AKT_UP.V1_UP” (enrichment score = -0.485, adjusted p-value = 0.027)

and “MEK_UP.V1_UP” (enrichment score = -0.465, adjusted p-value = 0.027), were also significantly enriched (28, 29), demonstrating that FOXA1 knockdown inhibited cell growth and survival pathways typically activated downstream of HER2. Conversely, our analysis revealed an enrichment in the “ERBB2_UP.V1_DN” gene set (enrichment score = 0.436, adjusted p-value = 0.021). This particular gene set comprises 200 genes that are down-regulated by constitutive ErbB2 signaling (Fig. 2M) (27). A volcano plot visualized these findings, marking “ERBB2_UP.V1_DN” and “ERBB2_UP.V1_UP” genes in blue and red, respectively (Fig. 2N). Furthermore, the heatmap detailing the expression profile of these leading-edge subset genes confirmed the definitive regulatory role of FOXA1 in HER2-related transcriptional pathways (Fig. 2O). These results are in concordance with previous biochemical data (Fig. 2C-E) and are supported by ChIP-seq and ATAC-seq analyses, affirming FOXA1’s broad, context-dependent influence in modulating HER2 signaling in breast cancer. FOXA1 binding sites’ differences between MCF7 and SKBR3 cells reflect the underlying biological distinctions between these two breast cancer subtypes and pinpoint the tailored therapeutic strategies based on the specific molecular mechanisms in different subtypes.

FOXA1 is required for luminal identity in HER2-positive breast cancer.

FOXA1 has been shown to suppress genes associated with the basal phenotype. Higher levels of FOXA1 are associated with a luminal phenotype in both ER-positive and ER-negative breast cancer cells (4). Similarly, we found that *FOXA1*, *ESR1*, and *ErbB2* were all more highly expressed in luminal tumor cells than in basal tumor cells. Luminal tumor cells also expressed higher levels of gene that are typically found in luminal epithelial cells, such as *KRT7*, *KRT18*,

and *KRT19* (Supplementary Fig. 6A-C). Based on these findings, we hypothesized that FOXA1 might contribute to maintaining the luminal identity of HER2-positive tumors, with the corollary that suppressing *FOXA1* might lead to the development of basal-like characteristics in these tumors.

To examine the above hypothesis, we first examined DEGs identified from the RNAseq analysis of FOXA1KD-SKBR3 cells and controls. Interestingly, among the top five upregulated genes in FOXA1KD (Supplementary table 3), four genes (*ABCG2*, *COL14A1*, *FSCN1*, and *SLC2A3*) have been associated with drug resistance, poor progression, metastasis and EMT in breast cancer, respectively (30-33). Furthermore, overrepresentation analysis of the top 563 upregulated genes in FOXA1KD (log2 fold change > 5 and adjusted p-value < 0.01) using EnrichR revealed a significant enrichment for four Hallmark gene sets from Molecular Signatures Database (MSigDB). In addition to the Estrogen Response Early gene set known to be associated with FOXA1, three additional gene sets were identified: Angiogenesis, KRAS Signaling Up and EMT (Supplementary table 4). We next employed the ChIP-X Enrichment Analysis 3 (ChEA3) tool to identify potential transcription factors (TFs) associated with the upregulated genes within these pathways. The top 10 TFs identified by ChEA3 included SCRT1, KCNIP3, ARNT2, SOX1, SALL3, TWIST2, INSM2, MEIS3, PRRX1 and SOX2, which are associated with EMT, metastasis and aggressiveness (Supplementary table 5) (34-42). Additionally, the upstream regulator analysis function in Ingenuity Pathway Analysis (IPA) indicated activation of several EMT-associated TFs in FOXA1KD-SKBR3 cells as compared to control SKBR3 cells (Supplementary Fig. 6E). Considering the significant enrichment of EMT-related gene sets and TFs within the highly upregulated genes in FOXA1KD, these findings

suggest that FOXA1 may maintain luminal characteristics, in part, by suppressing the expression of EMT-related genes in HER2-positive breast cancer.

To identify FOXA1-specific transcriptional programs for HER2-positive breast cancer, we compared open chromatin regions, defined by ATAC-seq, in FOXA1KD-SKBR3 cells with control cells. FOXA1KD-SKBR3 cells exhibited decreased chromatin accessibility in promoter-TSS sites compared to the controls (Fig. 3A and B), consistent with globally reduced accessibility (Supplementary Fig. 7A and B). As expected, regions associated with ERBB2 UP.V1_UP leading-edge genes in FOXA1KD-SKBR3 cells showed decreased accessibility and peak scores (Supplementary Fig. 8A). We then investigated whether the observed accessibility changes affected luminal cell type-related gene sets. Gene set enrichment analysis of Differential Accessible Regions (DARs)-associated genes using clusterProfiler (43) revealed that most of the enriched gene sets were defined by genes mapping to TSS regions with decreased accessibility in the knockdown cells. In contrast, the “EMT” and “Coagulation” gene sets were defined by genes with increased accessibility in FOXA1KD cells (Fig. 3C). Similarly, when we examined DARs across all genomic regions, we also identified significant enrichment for EMT-related gene sets (Fig. 3D, Supplementary Fig. 8B and C). Genomic regions associated with the EMT leading-edge genes also displayed increased accessibility and peak scores in FOXA1KD-SKBR3 cells (Supplementary Fig. 8D). Even though overall chromatin accessibility was reduced in FOXA1KD-SKBR3 cells, EMT leading-edge genes showed increased accessibility (Fig. 3E). Finally, across the HER2-positive cancer cell lines in the Cancer Dependency Map dataset, we observed negative correlations between FOXA1 expression and the top five EMT leading-edge genes (Supplementary Fig. 8E and F). Interestingly, there was no correlation between these same genes and FOXA1 expression in ER-positive cell lines (Supplementary Fig. 8E and F).

To test for a causal link between identified DARs in the ATAC-seq data and the corresponding gene expression changes, we employed a previously described methodology to correlate promoter accessibility with gene expression (44). This analysis revealed a robust positive correlation across all genes [R = 0.53 (Pearson); P-value < 2.2e-16] (Fig. 3F). Further focusing on EMT-related genes, we selected 73 genes with DARs in their promoter-TSS and regions and annotated them within the Hallmark EMT gene set. This analysis yielded an even stronger positive correlation [R = 0.63 (Pearson); a P-value < 5.6e-09] (Fig. 3G). Moreover, RNA-seq analysis revealed that the EMT related gene set was significantly enriched in the FOXA1KD-SKBR3 cells. (|Log2FC| > 3, adjusted P-value < 0.05). (Fig. 3H).

To further delineate FOXA1's role in HER2-positive breast cancer accounting for the cellular heterogeneity, we analyzed scRNA-seq data. We defined cell populations of *ErbB2*^{high}*FOXA1*^{high} and *ErbB2*^{low}/*FOXA1*^{low} (Fig. 3I and J) and determined DEGs between these groups. Intriguingly, the *ErbB2*^{low}*FOXA1*^{low} group showed enrichment of EMT-related genes (Fig. 3K). This finding was validated by comparing *FOXA1*-high and -low epithelial cells within the same HER2-positive patient scRNA-seq dataset. Again, the *FOXA1*-low group displayed increased expression of EMT-related genes (Supplementary Fig. 8G and H). Furthermore, Visium spatial transcriptomics analysis revealed a decreased EMT module score (see Method section) in a *FOXA1*^{high} *ErbB2*+/*ESR1*- tumor spots compared to *FOXA1*^{low} myoepithelial cell spots (Fig. 3L). In another independent spatial transcriptomics data, Xenium, where 313 genes were probed, *DST* and *OXTR* genes were selected as surrogates for an EMT-associated pathway (45, 46). Notably, both DST and OXTR expression was higher in the *FOXA1*^{low} control myoepithelial cell compared to *FOXA1*^{high} tumor cells (Supplementary Fig. 9A and B). Collectively, these scRNA-seq and two orthogonal spatial transcriptomics analysis results

provided compelling evidence that FOXA1 suppresses EMT and promotes luminal lineage differentiation in HER2-positive breast cancer cells.

FOXA1/TRPS1 regulates TEAD-YAP-TAZ signaling in the HER2-positive breast cancer.

To explore additional TFs potentially involved in HER2-positive breast cancer, we conducted motif discovery analysis on ATAC-seq data. As expected, FOXA1 and related FOX family motifs, including FOXM1 and FOXK1, were the most prevalent binding sites in regions with decreased chromatin accessibility in FOXA1KD-SKBR3 cells (Fig. 4A). In contrast, GATA and TEAD family motifs emerged as top hits in regions with increased chromatin accessibility (Fig. 4B). We first investigated changes of GATA family gene expression and their target genes following FOXA1 knockdown using RNA-seq data and the Dorothea database. We found that GATA 3 and GATA 5 expression was increased in bulk RNAseq analysis of FOXA1KD-SKBR3 cells (Supplementary Fig. 10A). Consistent with this finding, GATA3-target gene expression was significantly altered in FOXA1KD-SKBR3 cells (Supplementary Fig. 10B). GATA5 data were missing from the Dorothea database, so we could not perform a similar analysis for GATA5-target gene expression. Additionally, we validated upregulation of GATA3 mRNA and nuclear localization in FOXA1KD-SKBR3 cells by qRT-PCR and immunofluorescence, respectively (Supplementary Fig. 10C and D). These findings suggested that FOXA1 might suppress GATA3 activity in HER2-positive breast cancer cells. Given GATA3's known function in regulating ESR1 expression and predicting outcomes in breast cancers (47-50), we further explored the relationships between *ErbB2*, *FOXA1*, and *GATA3* expression using human breast cancer scRNA-seq data. We partitioned *ESR1*^{low} epithelial cells into high and low *ErbB2* expressing groups (Supplementary Fig. 10E). Consistent with our

previous findings, a strong positive correlation between *FOXA1* and *ErbB2* was observed in *ErbB2*^{high} expressing cells. This correlation, though weaker, also held true for *ErbB2*^{low} cells. While there was a positive correlation between *FOXA1* and *GATA3* expression in both *ErbB2*^{low} and *ErbB2*^{high} expressing cells (Supplementary Fig. 10F), there was a weak but highly significant negative correlation between *ErbB2* and *GATA3* expression only in the *ErbB2*^{high} group (Supplementary Fig. 10F). In contrast, *ErbB2*^{low} cells demonstrated a positive correlation between *GATA3* and *ErbB2* expression. This result suggested that *ErbB2* expression and activation through FOXA1 regulates the expression and activity of GATA3 in HER2-positive breast cancer. Further exploring the interplay between ErbB2, and GATA3, we applied SCENIC (51) to infer the gene regulatory networks in *ESR1*^{low} epithelial cells with high and low *ErbB2* expression. This analysis aligned with our previous findings, showing higher FOXA1 transcription activity in *ErbB2*^{high} group compared to *ErbB2*^{low} group (Supplementary Fig. 10G). Although GATA3 and GATA2 TF activities were relatively low, another GATA family member, GATA6, exhibited lower activity in the *ErbB2*^{high} group. Interestingly, several TFs implicated in EMT (52, 53), including JUNB, RELB, ATF4, MYBL2, and KLF4/5/6/10, display higher activity in *ErbB2*^{low} group (Supplementary Fig. 10G). Analysis of the Cancer Dependency Map revealed a negative correlation between the expression of *ErbB2* and *GATA3* in HER2-positive breast cancer cell lines, not in others (Supplementary Fig. 10H). These finding highlight the intricate and subtype-specific relationship between ErbB2, FOXA1 and GATA3 in breast cancer cells. While definitive conclusions require further investigation, they support the hypothesis that HER2 activation through FOXA1 regulates the gene expression and activity of GATA families in the HER2-positive breast cancer cells.

ATAC-seq analysis also revealed the TEAD family motif as prominently represented in regions of increased chromatin accessibility in FOXA1KD-SKBR3 cells (Fig. 4B). Supporting this finding, TEAD expression was higher in these cells (Fig. 4C and D). We next identified 245 genes with enriched TEAD binding motifs within chromatin regions that had increased accessibility in FOXA1KD-SKBR3 cells as compared to control cells. The majority of these motifs were located in non-promoter-TSS regions (54% intronic, 40% intergenic), and the remaining 6% were distributed across exon, TTS, and promoter-TSS regions (Fig. 4E and F). TEADs are TFs that function in EMT, proliferation and transformation (18, 54). Functional analysis of these 245 genes using clusterProfiler (43) revealed significant enrichment (adjusted p-value = 0.012) in the Hallmark EMT gene set. In line with this finding, RNA-seq data confirmed upregulation of TEAD target genes involved in EMT in FOXA1KD-SKBR3 cells. (Fig. 4G and H). Spatial transcriptome analysis using Visium provided further evidence. Compared to other cell types, *FOXA1*^{high} tumor showed lower expression of EMT-related genes such as *TEAD1*, *TEAD2*, *FBN1*, *GAS1* and *FSTL1* (Supplementary Fig. 11A). Additionally, the TEAD module score was significantly lower in *FOXA1*^{high} tumors compared to the myoepithelial cells (Wilcoxon rank sum test, p-value = 6.26e-94) (Supplementary Fig. 11B).

Interactions between TEADs and YAP-TAZ signaling has been demonstrated to activate the transcription of EMT pathways (18, 55), so we investigated potential interactions between TEAD and YAP in response to FOXA1 knockdown. Immunofluorescence analysis using a pan-TEAD antibody revealed increased nuclear TEAD in FOXA1KD-SKBR3 cells, but no change in YAP immunofluorescence (Fig. 4I). RNA-seq data confirmed that YAP coding gene (*YAP1*) expression remained unchanged (Log2FC = 0.13, adjusted p-value = 0.11) upon FOXA1 knockdown. However, a proximity ligation assay (PLA), which detects protein-protein

interactions, showed a clear increase in TEAD-YAP interaction in the FOXA1KD-SKBR3 cells compared to control cells (Fig. 4J). These findings suggest that while YAP gene expression and protein levels remain stable, FOXA1 loss increases TEAD expression and TEAD-YAP interactions, potentially, contributing to the activation of EMT pathway.

While our findings suggested increased TEAD activity in FOXA1KD-SKBR3 cells, the enrichment of TRPS1 binding motifs in accessible chromatin regions raised questions. TRPS1, a GATA-type transcriptional repressor commonly found in breast cancers (56), has been shown to inhibit YAP-TEAD-dependent transcriptional activity (19). Interestingly, ATAC-seq analysis results revealed decreased chromatin accessibility within the TRPS1 gene promoter in FOXA1KD-SKBR3 cells compared to controls (Fig. 4K). This correlated with reduced TRPS1 mRNA and protein levels as shown by RNA-seq, immunoblotting and immunofluorescence analyses (Fig. 4L-N). Furthermore, PLA demonstrated diminished interaction between TRPS1 and YAP proteins in FOXA1KD-SKBR3 cells (Fig. 4O). scRNA-seq data provided further evidence, showing enriched *TRPS1* gene expression in luminal tumor subtypes (both ER-positive and HER2-positive) and significantly lower expression in the basal sub-type (Wilcoxon rank sum test, p-value = 7.94e-199 and 4.85e-161 respectively) (Fig. 4P). Spatial transcriptome analysis using Visium also revealed elevated *TRPS1* expression in *FOXA1*^{high} tumors (Supplementary Fig. 11A). Taken together, these findings suggest that FOXA1 facilitates TRPS1 expression, which in turn suppresses TEAD/YAP-TAZ-mediated gene expression and EMT-related gene expression in luminal breast tumors.

Discussion

FOXA1 is recognized for its ability to regulate the chromatin accessibility of various TFs, notably the AR and ERα (3, 57). This realization prompted extensive research into its involvement in prostate cancer and ER-positive breast cancer. Although the majority of studies of FOXA1 in breast cancer have focused on its regulation of ER activity, it has also been recognized that FOXA1 is expressed in HER2-positive breast cancers exhibiting low ERα levels (58). This study employed a range of genomic techniques to examine the role of FOXA1 in HER2-positive breast cancer more comprehensively. Analysis of single-cell sequencing data from nineteen human breast cancer patients verified co-expression of *FOXA1* with *ErbB2* in HER2-positive breast cancers, demonstrating subtype-specific correlations of FOXA1 with ERα or HER2. Notably, stronger associations were observed between ERα and FOXA1 in ERα-positive tumors, and between *ErbB2* and *FOXA1* expression in HER2-positive tumors. Epigenetic analyses utilizing ChIP-seq and ATAC-seq data indicated FOXA1's interaction with the promoter region and downstream regulatory regions of the *ErbB2* gene in HER2-positive SKBR3 breast cancer cells, interactions that were absent in ERα–positive MCF7 breast cancer cells. FOXA1 knockdown in SKBR3 cells led to diminished *ErbB2* mRNA expression and reduced HER2 and EGFR activation, demonstrating that FOXA1 binding to these sequences in the *ErbB2* gene is required for the expression of the ErbB2 gene in SKBR3 cells. Further analyses with ATAC-seq and RNA-seq data revealed that *FOXA1* depletion in HER2-positive breast cancer cells decreased *TRPS1* expression, increased GATA3 expression, augmented TEAD/YAP interactions, and enhanced the expression of EMT-related genes. Our results underscore FOXA1's crucial role in sustaining *ErbB2* expression and luminal tumor identity in HER2-positive breast cancer (Fig. 4Q).

FOXA1 has been well characterized as a pioneer factor facilitating chromatin accessibility for the sex steroid nuclear receptors ERα and AR in breast cancer and prostate cancer respectively (3). In breast cancers, it has been shown that FOXA1 plays an important role in facilitating the binding of ERα to chromatin, and it is required for the expression of the estrogen receptor as well as typical estrogen responsive genes that define the luminal A breast cancer phenotype. Our ChIPseq and ATACseq data in MCF7 cells are consistent with this model as FOXA1 binds to open chromatin regions in the *ESR1* gene promoter. However, in HER2-positive ERα-negative cells, FOXA1 does not bind to the *ESR1* promoter, but instead it binds to open chromatin in the *ErbB2* gene. This specificity goes beyond these two genes as our ChIPseq data demonstrate that there are many FOXA1 binding sites that are specific to either MCF7 or SKBR3 cells, as well as many shared sites. Although FOXA1 regulates key identifying genes (ERα vs. ErbB2) of these two sub-types of breast cancer, the mechanisms that determine which binding sites are used in the two different types of cancer remain to be identified.

Despite its known expression in HER2-positive breast cancers, the molecular mechanisms of FOXA1's interaction with HER2 have not been extensively explored (13, 59). Several studies have implicated FOXA1 in suppressing EMT across several solid tumors (60-62). Our investigation demonstrates that the FOXA1/HER2 axis plays a pivotal role in maintaining luminal lineage characteristics by inhibiting chromatin accessibility and expression of EMT-related genes in HER2-positive breast cancer cells. TRPS1, a GATA family member known to repress EMT by binding to GATA sequence (63) during development (64), exhibits tumor-suppressive roles in breast and colon cancer (56, 65). Interestingly, previous research by Huang et al. suggested that TRPS1 induces *FOXA1* expression by binding to its promoter in cancer cells

(65). In contrast, we found that FOXA1 regulates *TRPS1* expression by enhancing accessibility to its promoter site. Our data suggest that FOXA1 promotes TRPS1 expression, which, in turn suppresses EMT. TRPS1 activation is known to be associated with decreased TEAD-YAP-dependent transcriptional activity, which serve as key regulators of EMT and cancer progression (19). In line with these findings, our data demonstrate that *FOXA1* knockdown decreases *TRPS1* expression, increases the accessibility of TEAD binding motifs, and enhances TEAD-YAP protein interactions, suggesting that loss of FOXA1 disinhibits a TRPS1-suppressive pathway that normally inhbits TEAD-YAP interactions and EMT within HER2-positive breast cancers. The metastatic potential of breast cancer is closely linked to altered expression of Hippo pathway components, including TEAD and YAP (66). Therefore, our findings suggest that FOXA1's antagonistic role the Hippo pathway may regulate breast cancer progression in luminal breast cancer subtypes.

The binding motif analysis of the ATACseq data suggested a predominance of GATA family motifs, as well as TEAD family motifs, in chromatin regions exhibiting increased accessibility in response to FOXA1KD. Several studies have examined the interaction between ERα, FOXA1 and GATA3 in breast cancer and during mammary gland differentiation (67). Carroll et al. reported the crucial role of FOXA1 in ERα binding to chromatin sites and subsequent signaling in mammary epithelial cells (68). GATA3, on the other hand, emerged as a key regulator of cell-fate specification in the mammary gland, influencing both embryonic development and luminal cell differentiation (14, 69). Both FOXA1 and GATA3 are associated with Erα expression and action, while GATA3 has been reported to exhibit an inverse relationship with HER2 expression (70). Another study identified that FOXA1 regulates ERα expression, but not GATA3, in breast cancer cell lines (71). In our studies of SKBR3 cells, we

found that knocking down FOXA1 resulted in loss of ErbB2 expression and signaling, but resulted in a reciprocal increase in GATA3 expression. This occurred even though loss of FOXA1 also led to loss of luminal character and an increase in EMT-related gene expression. These results suggest that the associations between GATA3 and luminal differentiation may differ in the context of ERα expression versus ErbB2 upregulation. Although the broader role of GATA3 in breast cancer prognosis has been investigated (72), our results demonstrate that the specific impact of GATA3 expression in HER2-positive cancers remains unclear and requires further study.

In conclusion, our study demonstrates that the FOXA1/HER2 axis plays a crucial role in preserving luminal characteristics, reduced EMT and enhanced ErbB2 expression, in HER2-positive breast cancer cells. These findings are consistent with prior reports demonstrating that FOXA1 can serve to distinguish luminal and basal subtypes (73, 74). However, a deeper understanding of the molecular mechanisms governing the interactions between FOXA1 and ErbB2 in HER2-positive tumors may pave the way for developing novel therapeutic strategies for this tumor sub-type and may also contribute to a better understanding of luminal-basal plasticity during breast cancer progression. While further validation in larger cohorts and exploration of additional contributing mechanisms are needed, our findings provide a compelling rationale for further investigation of the FOXA1/HER2 axis as a potential target for improving patient outcomes in breast cancer.

Materials and Methods

Cell culture

The human cell lines, SKBR3 and MCF7, were obtained from ATCC and maintained in culture in DMEM +GlutaMAX-1 (Gibco-Life Technologies) containing 10% fetal bovine serum (FBS) and pen/strep (Gibco-Life Technologies) at 37˚C in 5% CO_2. MCF10A cells were cultured in DMEM/F12 (Gibco-Life Technologies) containing 5% horse serum, EGF (100ng/ml), hydrocortisone (1mg/ml), cholera toxin (1mg/ml), insulin (10µg/ml), and pen/strep (Gibco-Life Technologies) at 37˚C in 5% CO^2.

Knockdown Cell Line

A stable cell line expressing shRNA directed against FOXA1 was generated by transducing cells with commercially prepared lentiviruses containing three individual shRNA directed against FOXA1 mRNA (sc-37930-V) (Santa Cruz). Cells were cultured in 6-well plates and infected by adding the shRNA lentiviral particles to the culture for 48 hours per the manufacturer's instructions. Stable clones expressing the specific shRNAs were selected using 5µg/ml of puromycin (Gibco-life technologies) and pooled to generate the cells used in the experiments.

Immunofluorescence

Cells were grown on coverslips, fixed in 4% paraformaldehyde for 20 min, permeabilized with 0.2 % Triton X100 for 10 mins, washed three times with PBS, and incubated with primary antibody overnight at 4°C. The cells were washed three times with PBS and incubated with a

secondary antibody for 1 hour at room temperature. After washing, coverslips were mounted using Prolong Gold antifade reagent with DAPI (Invitrogen). Confocal microscopy images were obtained using a Zeiss LSM 880 confocal microscope. Primary antibodies included those against FOXA1 (ab170933) from Abcam (Waltham, MA); HER2 (MA1-35720) and TRPS1 (PA5-84874) from Thermo Scientific (Waltham, MA); phosphor-HER2, GATA3 (5852), and pan-TEAD (13295) from Cell Signaling (Danvers, MA); YAP (sc-101199) from Santa Cruz (Dallas, TX).

RNA Extraction and Real-Time RT-PCR

RNA was isolated using TRIzol (Invitrogen). Quantitative RT-PCR was performed with the SuperScript III Platinum One-Step qRT-PCR Kit (Invitrogen) using a Step One Plus Real-Time PCR System (Applied Biosystems) and the following TaqMan primer sets: *FOXA1*, Hs04187555_m1; *ErbB2*, Hs01001580_m1; *GATA3*, Hs00231122_m1. Human *HPRT1* (4326321E) (Invitrogen) was used as a reference gene. Relative mRNA expression was determined using Step One Software v2.2.2 (Applied Biosystems).

Immunoblotting

Protein samples were prepared from cells using standard methods, subjected to SDS-PAGE, and transferred to a nitrocellulose membrane by wet western blot transfer (Bio-Rad). The membrane was blocked in TBST buffer (TBS + 1% Tween) containing 5% milk for 1 hour at room temperature. The blocked membranes were incubated overnight at 4°C with specific primary antibodies (Odyssey blocking buffer, 927-40000). The membranes were washed three times with TBST buffer and then incubated with specific secondary antibodies provided by LI-

COR for two hours at room temperature. After three washes with TBST buffer, the membranes were analyzed using the ODYSSEY Infrared Imaging system (LI-COR). All immunoblot experiments were performed at least three times, and representative blots are shown in the figures. Primary antibodies included those against FOXA1 (ab170933) from Abcam (Waltham, MA); phospho-HER2 (thr1221/1222) (2243S), EGFR (4267S), and pan-TEAD (13295) from Cell Signaling (Danvers, MA) from Cell Signaling (Danvers, MA); HER2 (sc-33684), phospho-EGFR (sc-12351) from Santa Cruz (Dallas, Texas); TRPS1 (PA5-84874) from Thermo Scientific (Waltham, MA).

Single cell RNA sequencing data analysis

Count matrices from published single cell RNA sequencing (scRNA-seq) datasets (22) were downloaded from the NCBI Gene Expression Omnibus (GSE161529) and then analyzed using Seurat version 4.0 as described (75). 15 ER-positive and 6 HER2-positive patient-derived datasets were included for analysis. Cells were divided by a high group and low group, depending on their MAGIC imputation counts (t = 2) of *FOXA1*, *ESR1* and *ErbB2* expression (76). The high group was assigned by having an expression level greater than median, and remaining cells were designated as the low group, unless otherwise noted. To run SCENIC workflow on scRNA-seq data, each group was down-sampled to 1000 cells and applied default parameters (51). To predict transcription factor regulons, we used human v9 motif collection, as well as both hg38__refseq-r80__500bp_up and 100bp_down_tss.mc9nr.feather and hg38__refseq-r80__10kb_up_and_down_tss.mc9nr.feather databases from the cisTarget (https://resources.aertslab.org/cistarget/). The resulting AUC scores per each cell and adjacency matrix were used for downstream analysis and visualization.

Spatial transcriptomics data analysis

The Visium and Xenium dataset used in this analysis were provided by 10X Genomics. Human breast cancer sections are freely available datasets and can be downloaded at from the 10X website (https://www.10xgenomics.com/resources/datasets). Visium spots were annotated refer to the deconvolution method Robust Cell Type Decomposition (RCTD) and previous report (23). To calculate the module score for EMT and TEAD feature expression program per spot, Seurat's AddModuleScore function was applied to Visium data. For EMT module score, 136 genes were selected because they were both involved in Hallmark EMT pathway gene set and have comparable count read (77). For TEAD module score, 563 genes related to the TEAD pathway in breast cancer were used (78).

Cell segmentation boundaries and a high-resolution gene expression data at region of interest (ROI) were visualized using the Xenium Explorer software (version 1.1.0). Differential gene expression across the selected cell types were compared with the Wilcoxon test on log-normalized count data. P-value was calculated from a down-sampled dataset by 1,000 to 2,000 cells to avoid zero inflation of scRNA-seq data.

ChIP-seq and data analysis

ChIP-seq for FOXA1 was performed as described in (79, 80). (FOXA1 antibody; cat. ab170933). DNA was purified with a QIAquick PCR purification kit (Qiagen, Germantown, MD, USA) for sequencing. ChIP-seq library preparation was done using the TruSeq Nano DNA

Library Prep Kit (Illumina, San Diego, CA, USA) and sequenced on the Illumina NextSeq550 sequencing platform (performed at the Yale Sequencing Center).

The quality of raw sequencing reads was evaluated with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The raw sequencing reads from FOXA1 and matched input DNA were mapped to the human reference genome (hg38) using the Burrows-Wheeler Aligner (BWA) mem program (version 0.7.9a) (81). After alignment, PCR duplicates were removed with Picard tools (http://broadinstitute.github.io/picard/) (MarkDuplicates v.2.9.0) and FOXA1 binding regions (i.e., peaks) were identified using MACS tool (version 1.4). Peaks were annotated using annotatePeaks.pl script from HOMER. Deeptools (version v.3.1) and Integrative Genomics Viewer (IGV, version 2.12) were used for visualization.

ATAC-seq and data analysis

ATAC-seq was performed as described in ref (82). Extracted cells from SKBR3 control and FOXA1KD with two biological replicates were subsequently processed using standard DNA extraction and library preparation protocols (performed at the Yale Sequencing Center). Libraries were prepared according to the recommended Illumina NextSeq550 sequencing platform (performed at the Yale Sequencing Center).

ATAC-seq data processing was achieved as previously described (83). HOMER software was used for motif analysis based on the peak annotation files. The findMotifsGenome.pl script in HOMER was used to discover enriched motifs for each condition. To examine the chromatin accessibility changes around the peak summit, deepTools (version 3.5) was used. Heat map (tornado plots) of peaks were generated by combining scores for regions and binning the region from -1 kb to +1 kb of transcription start site. Differential accessibility of the peaks was assessed

using DESeq2 (version 3.16) using peak scores between pairwise comparisons of FOXA1KD to control.

Bulk RNA-seq and data analysis

The raw sequencing reads files in Fastq format were assessed for quality with FastQC. Pseudo-alignment was achieved using Salmon (version 1.9) with hg38 reference genome. Subsequent analyses were performed using DESeq2 (version 3.16) or IsoformSwtichAnalyzeR (version 1.18) with output quantification files. The differential expression of FOXA1KD and control samples was performed using DESeq2 with R package. Adjusted P-value of 0.05 by Benjamini-Hochberg method and log2FC value of 1 are considered as significant, unless otherwise noted. Functional pathway enrichment of differentially expressed genes were determined using clusterProfiler package (version 4.4) from the molecular signature database Hallmark collection and Oncogenic signature (version 7.4) (43, 77). Data was also analyzed with the Ingenuity Pathway Analysis (IPA, QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis). Differential isoform usage analysis was performed using the R package IsoformSwitchAnalyzeR (84). Following approaches, Salmon quantification tables were imported using tximeta function to create a swtichAnalyzeRlist object. Analysis of an alternative splicing event was evaluated using isoformSwitchAnalysisCombined function with default parameter except following options; alpha = 0.05, dIFcutoff = 0.1, switchTestMethod = DEXSeq. Different expression of all isoforms belonging to ErbB2 gene was extracted from aSwtichList object and visualized.

Genetic dependency analysis

The dependency data is available online, at https://depmap.org/portal/. RNAi and CRISPR screening data for 1,255 cell lines were used to identify cellular viability following gene perturbation. Pearson correlation between each gene's log2(TPM+1) expressions from 22Q2 Public were computed. Genes were examined to compare distinct dependency pattern according to lineage sub-subtype in breast cancer.

Proximity Ligation Assays

Proximity ligation assays (PLA) were performed using the Duolink™ assay kit (Sigma). SkBR3 and MCF10A cells were seeded onto 22 mm round collagen-coated coverslips (Corning, Cat N 354089). The experiments were performed when the cells reached 80% confluence. The cells were washed 3 times with PBS and paraformaldehyde (4% in PBS) was added to each well for 20 min, after which they were permeabilized with 0.2% Triton X100 in PBS. Permeabilized cells were incubated with combinations of the following antibodies: TRPS1 (PA5-84874) from Thermo Scientific (Waltham, MA), pan-TEAD (13295) from Cell Signaling (Danvers, MA) and YAP (sc-101199) from Santa Cruz (Dallas, TX). PLA probes were then added and the assay was performed as per the manufacturer's instructions. All images were obtained using a Zeiss LSM 880.

Quantification and statistical analysis

The number of repeats (n) refers to the number of independent experimental observations in the legend of each figure. Statistical analyses were performed with Prism 6.0 (GraphPad Software, La Jolla, CA). Statistical significance was determined using an unpaired t test for comparisons between two groups. Group data were presented as mean±SEM. The P-values less

than 0.05 (indicated by *), or less than 0.01 (**), or less than 0.001 (***), or less than 0.0001 (****) were considered as significant.

Author contributions

Study concept and design: J.Jeong. J. Wysolmerski. and J.Choi.; acquisition of data: J.Jeong., J.Lee., and J.Choi.; analysis and interpretation of the data: J.Jeong., J.Lee., K.Yoo., J.Lim and J.Choi. technical or other material support: J.Shin, J.Kim, Y.Tanaka, H.S.Tae, and I.Park.; writing and review of the manuscript: J.Jeong., J.Lee., J. Wysolmerski. and J.Choi.

Acknowledgements

This work was partly supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean government (NRF-2022R1A4A2000827, to J.Lee and J.Choi). R01 HD100468-01A1 and R01 HD076248-06A1 from the NIH to J. Wysolmerski. JC was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICAN(ICT Challenge and Advanced Network of HRD) program(IITP-2024-RS-2022-00156439) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation).

Data availability

Conflict of interest

Nothing declared.

Figure 1. Co-expression of *FOXA1* and *ErbB2* in HER2 positive breast cancer cells.

A. Uniform Manifold Approximation and Projection (UMAP) of human breast tumor cells. **B**. UMAP (upper panel) and split violin plots for *ESR1*, *ErbB2* and *FOXA1*. **C.** Pearson correlation coefficient between *ESR1* and *ErbB2* among ER-positive or HER2-positive epithelial cells. P-value was calculated from a down-sampled dataset with 1 to 10 % to avoid zero inflation of scRNA-seq data. **D.** UMAP plots and co-expression pattern of *ESR1*, *ErbB2*, and *FOXA1* in epithelial cells. **E.** Bar plot showed co-expression of two features simultaneously (Criteria that expression level more than 0.5 is used to assign positive cells). **F.** H&E staining FFPE tissue conducted pre-CytAssist is shown. **G.** Cell type specific marker genes are expressed as $\log_2$ (normalized UMI counts). **H.** Spatial distribution of cells idented by unsupervised clustering based on differential gene expression analysis. **I.** Violin plot illustrating the selected marker genes expression of each cell type. The box centerlines depict the medians, and the edges depict the first/third quartiles. **J.** Pearson correlation coefficient between FOXA1 and ERBB2 of each cell type.

Figure 2. FOXA1 regulates *ErbB2* expression.

A. Histogram showing the relative expression levels of *FOXA1* from qRT-PCR in control and FOXA1KD-SKBR3 cells. **B.** Immunofluorescence images of FOXA1 (green) in control and FOXA1KD-SKBR3 cells. Scale bars is 10μm. **C.** Histogram showing the relative expression levels of *ErbB2* from qRT-PCR in control and FOXA1KD-SKBR3 cells. **D.** Western blots showing the indicated protein levels in control and FOXA1KD-SKBR3 cells. **E.** Immunofluorescence images of HER2 (green) and pHER2 (red) in control and FOXA1KD-SKBR3 cells. DAPI was used to stain the nucleus. Scale bars is 10μm. **F.** Immunofluorescence

images of FOXA1 (green) in MCF7 and SKBR3 cells. Scale bars is 10μm. **G.** Histogram showing the relative expression levels of *FOXA1* from qRT-PCR among MCF10A, MCF7, and SKBR3 cells. **H.** Donut charts showing the distribution of FOXA1 ChIP-seq peaks in functional genomic regions according to Homer gene annotation. **I.** Venn diagram showing the overlapped functional genomic regions between FOXA1 binding regions from SKBR3 and MCF7 breast cancer cells. **J-L.** Genome tracks of *ErbB2* and *ESR1* gene showing the identified peaks over pre-defined promoter region (H3K4Me3) and regulatory region (H3K27Ac). FOXA1 binding motifs matched with ChIP-seq peak regions were marked with a red triangle. The screenshot was generated using the UCSC Genome Browser (85) or IGV. **M.** Dot plot of the Oncogenic signature gene sets significantly (adjusted p-value < 0.05, Benjamini-Hochberg correction) activated (left) or suppressed (right) in the FOXA1KD-SKBR3 cells compared to the control. **N.** Volcano plot showing the log2FC and adjusted P-values of all genes (light grey) from the differential test between control and FOXA1KD-SKBR3 cells. The red dots represent ERBB2_UP.V1_UP leading-edge subset genes and the blue dots represent ERBB2_UP.V1_DN leading-edge subset genes. Dots that have adjusted P-value 0 were marked with jittered value ± 5 at y = 300 for visualization. Blue dotted lines indicate |log2FC| = 2, adjusted P-value = 0.05. **O.** Heatmap showing the relative expression of ERBB2_UP.V1_DN leading-edge subset genes and ERBB2_UP.V1_UP leading-edge subset genes, respectively. Bar graphs represent the mean±SEM. *** denotes p<0.0005, **** denotes p<0.00005.

Figure 3. FOXA1 expression is critical for luminal tumor identity.

A. Heatmap representing the signal enrichment of control and FOXA1KD-SKBR3 cells. ATAC-seq reads in 2-kb regions centered on the TSS sites of annotated promoter-TSS sites of protein-coding genes. Down-regulated (upper panel) and up-regulated (lower panel) in FOXA1KD-SKBR3 cells than control respectively. **B.** Boxplot showing the average scores per region from ATAC-seq for control and FOXA1KD-SKBR3 cells. Wilcoxon rank sum test was performed, and the P-value was labeled at the plot. **C-D.** Dot plot showing the significantly (adjusted p-value < 0.05, Benjamini-Hochberg correction) enriched Hallmark gene sets by GSEA (C) and ORA (D) in the FOXA1KD-SKBR3 cells compared to control. Genes corresponded with DARs annotated only as Promoter-TSS were used for analysis. **E.** Volcano plot showing the log2FC and adjusted P-values of all peaks which genes annotated as promoter-TSS from the differential test between control and FOXA1KD-SKBR3 cells. The red dots represent HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION leading-edge subset genes. Blue dotted lines indicate |log2FC| = 0.5, adjusted P-value = 0.05. **F.** Scatter plot showing the correlation of the log2FC values between RNA-seq and ATAC-seq. Displayed R value determined by Pearson correlation. **G.** Scatter plot showing the correlation of the log2FC values of HALLMARK_EMT_genes between RNA-seq and ATAC-seq. Displayed R value determined by Pearson correlation. **H.** Dot plot showing the significantly enriched Hallmark gene sets by ORA performed for DEGs (|FC| > 3) in the FOXA1KD-SKBR3 cells compared to the control. Heatmap showing the expression of 41 enriched genes in Hallmark_EMT gene set. **I.** UMAP dimensional reduction of all epithelial cells. *ErbB2*/*FOXA1*-high and -low groups were divided according to the corresponding gene expression level. **J.** Feature plot of cells that each high or low group for *ErbB2* and *FOXA1*. **K.** Dot plot showing the significantly enriched Hallmark gene sets by ORA performed for DEGs (|FC| > 0.25 utilized MAST) in the *ErbB2*/*FOXA1*-high

compared to the -low group. Heatmap showing the expression of 20 enriched genes in Hallmark EMT gene set. **L.** Violin plot illustrating the EMT module score of each cell type. The box centerlines depict the medians, and the edges depict the first/third quartiles.

Figure 4. FOXA1/TRPS1 inhibits TEAD-YAP-TAZ signaling in HER2-positive breast cancer.

A-B. Top enriched TF binding motifs identified by HOMER known motif analysis associated with down-regulated (A) or up-regulated (B) peak regions from ATAC-seq. **C.** Volcano plot showing the log2FC and adjusted P-values of all expressed genes (light grey) from the differential test between control and FOXA1KD-SKBR3 cells. The red dots represent TEAD family genes. Dots which have adjusted P-value 0 were marked with jittered value ± 5 at y = 300 for visualization. Blue dotted lines indicate |log2FC| = 2, adjusted P-value = 0.05. **D.** Western blots showing the indicated protein levels in control and FOXA1KD-SKBR3 cells. **E.** Donut charts showing the TEAD-binding sites mapping to functional genomic regions according to Homer gene annotation. **F.** Heatmap representing the increased enriched motif signal in control and FOXA1KD-SKBR3 cells. ATAC-seq reads against TEAD family binding sites. **G.** Dot plot showing the Hallmark ORA performed for 245 genes that enriched in TEAD family binding motif in FOXA1KD-SKBR3 cells compared to control. **H.** Volcano plot showing the log2FC and adjusted P-values of all genes (light grey) from the differential test between control and FOXA1KD-SKBR3 cells. The red dots represent enriched genes of Hallmark_EMT ORA. Blue dotted lines indicate |log2FC| = 2, adjusted P-value = 0.05. **I.** Immunofluorescence images of pan-TEAD or YAP in breast cancer cell lines. Scale bars is 10μm. **J.** Proximity Ligation Assay (PLA) assay of TEAD/YAP in MCF10A, SKBR3, and FOXA1KD-SKBR3 cells. Nuclei stained

with DAPI (blue). Scale bars is 10µm. **K.** Genome tracks of *TRPS1* gene showing the identified peaks over pre-defined promoter region (H3K4Me3) and enhancer (H3K27Ac). The screenshot was generated using the UCSC Genome Browser (85) or IGV. **L.** Volcano plot showing the log2FC and adjusted P-values of all genes (light grey) from the differential test between control and FOXA1KDSKBR3 cells.

The blue dots represent TRPS1 gene. Blue dotted lines indicate |log2FC| = 2, adjusted P-value = 0.05. **M.** Western blots showing the TRPS1 protein levels in control and FOXA1KD-SKBR3 cells. **N.** Immunofluorescence images of TRPS1in MCF10A, SKBR3, and FOXA1KD-SKBR3 cells. Nuclei stained with DAPI (blue). Scale bars is 10µm. **O.** Proximity Ligation Assay (PLA) assay of TRPS1/YAP in MCF10A, SKBR3, and FOXA1KD-SKBR3 cells. Nuclei stained with DAPI (blue). Scale bars is 10µm. **P.** UMAP of basal and luminal subtypes breast tumor cells were represented (upper panel) and for *TRPS1* (middle panel). Boxplot showing the expression level of *TRPS1* among breast cancer subtype in scRNA-seq data (bottom panel). **Q.** Working Model explaining the role of FOXA1 in ErbB2 expression and in luminal characteristics in HER2-positive breast cancer.

References

1. Sahu B, Laakso M, Ovaska K, Mirtti T, Lundin J, Rannikko A, et al. Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. Embo j. 2011;30(19):3962-76.
2. Jin HJ, Zhao JC, Wu L, Kim J, Yu J. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. Nat Commun. 2014;5:3972.
3. Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell. 2008;132(6):958-70.
4. Bernardo GM, Bebek G, Ginther CL, Sizemore ST, Lozada KL, Miedler JD, et al. FOXA1 represses the molecular phenotype of basal breast cancer cells. Oncogene. 2013;32(5):554-63.
5. DeGraff DJ, Grabowska MM, Case TC, Yu X, Herrick MK, Hayward WJ, et al. FOXA1 deletion in luminal epithelium causes prostatic hyperplasia and alteration of differentiated phenotype. Lab Invest. 2014;94(7):726-39.
6. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS. FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nat Genet. 2011;43(1):27-33.
7. Arteaga CL, Engelman JA. ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. Cancer Cell. 2014;25(3):282-303.
8. Arteaga CL, Sliwkowski MX, Osborne CK, Perez EA, Puglisi F, Gianni L. Treatment of HER2-positive breast cancer: current status and future perspectives. Nat Rev Clin Oncol. 2011;9(1):16-32.
9. Hynes NE, MacDonald G. ErbB receptors and signaling pathways in cancer. Curr Opin Cell Biol. 2009;21(2):177-84.
10. Guy CT, Webster MA, Schaller M, Parsons TJ, Cardiff RD, Muller WJ. Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. Proc Natl Acad Sci U S A. 1992;89(22):10578-82.
11. Hanker AB, Cook RS, Arteaga CL. Mouse models and anti-HER2 therapies. Oncotarget. 2013;4(11):1866-7.
12. Angus SP, Stuhlmiller TJ, Mehta G, Bevill SM, Goulet DR, Olivares-Quintero JF, et al. FOXA1 and adaptive response determinants to HER2 targeted therapy in TBCRC 036. NPJ Breast Cancer. 2021;7(1):51.
13. Naderi A, Meyer M, Dowhan DH. Cross-regulation between FOXA1 and ErbB2 signaling in estrogen receptor-negative breast cancer. Neoplasia. 2012;14(4):283-96.
14. Kouros-Mehr H, Slorach EM, Sternlicht MD, Werb Z. GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. Cell. 2006;127(5):1041-55.
15. Nakshatri H, Badve S. FOXA1 in breast cancer. Expert Rev Mol Med. 2009;11:e8.
16. Yildirim E, Bektas S, Gundogar O, Findik D, Alcicek S, Erdogan KO, et al. The Relationship of GATA3 and Ki-67 With Histopathological Prognostic Parameters, Locoregional Recurrence and Disease-free Survival in Invasive Ductal Carcinoma of the Breast. Anticancer Res. 2020;40(10):5649-57.
17. Pan D. The hippo signaling pathway in development and cancer. Dev Cell. 2010;19(4):491-505.

18. Lamar JM, Stern P, Liu H, Schindler JW, Jiang ZG, Hynes RO. The Hippo pathway target, YAP, promotes metastasis through its TEAD-interaction domain. Proc Natl Acad Sci U S A. 2012;109(37):E2441-50.
19. Elster D, Tollot M, Schlegelmilch K, Ori A, Rosenwald A, Sahai E, et al. TRPS1 shapes YAP/TEAD-dependent transcription in breast cancer cells. Nat Commun. 2018;9(1):3115.
20. Dempster JM, Rossen J, Kazachkova M, Pan J, Kugener G, Root DE, et al. Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. bioRxiv. 2019:720243.
21. Krill-Burger JM, Dempster JM, Borah AA, Paolella BR, Root DE, Golub TR, et al. Partial gene suppression improves identification of cancer vulnerabilities when CRISPR-Cas9 knockout is pan-lethal. bioRxiv. 2022:2022.03.02.482624.
22. Pal B, Chen Y, Vaillant F, Capaldo BD, Joyce R, Song X, et al. A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. Embo j. 2021;40(11):e107333.
23. Janesick A, Shelansky R, Gottscho AD, Wagner F, Rouault M, Beliakoff G, et al. High resolution mapping of the breast cancer tumor microenvironment using integrated single cell, spatial and in situ analysis of FFPE tissue. bioRxiv. 2022:2022.10.06.510405.
24. Andersson A, Larsson L, Stenbeck L, Salmén F, Ehinger A, Wu SZ, et al. Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. Nat Commun. 2021;12(1):6012.
25. Comşa Ş, Cîmpean AM, Raica M. The Story of MCF-7 Breast Cancer Cell Line: 40 years of Experience in Research. Anticancer Res. 2015;35(6):3147-54.
26. Agrawal A, Balcı H, Hanspers K, Coort SL, Martens M, Slenter DN, et al. WikiPathways 2024: next generation pathway database. Nucleic Acids Res. 2024;52(D1):D679-d89.
27. Creighton CJ, Hilger AM, Murthy S, Rae JM, Chinnaiyan AM, El-Ashry D. Activation of mitogen-activated protein kinase in estrogen receptor alpha-positive breast cancer cells in vitro induces an in vivo molecular phenotype of estrogen receptor alpha-negative human breast tumors. Cancer Res. 2006;66(7):3903-11.
28. Carmona FJ, Montemurro F, Kannan S, Rossi V, Verma C, Baselga J, et al. AKT signaling in ERBB2-amplified breast cancer. Pharmacol Ther. 2016;158:63-70.
29. Khan IA, Yoo BH, Rak J, Rosen KV. Mek activity is required for ErbB2 expression in breast cancer cells detached from the extracellular matrix. Oncotarget. 2017;8(62):105383-96.
30. Ee PL, He X, Ross DD, Beck WT. Modulation of breast cancer resistance protein (BCRP/ABCG2) gene expression using RNA interference. Mol Cancer Ther. 2004;3(12):1577-83.
31. Malvia S, Chintamani C, Sarin R, Dubey US, Saxena S, Bagadi SAR. ABERRANT EXPRESSION OF COL14A1, CELRS3, and CTHRC1 IN BREAST CANCER CELLS. Exp Oncol. 2023;45(1):28-43.
32. Chang D, Liu XX, Liu R, Sun JW. The role and regulatory mechanism of FSCN1 in breast tumorigenesis and progression. Yi Chuan. 2023;45(2):115-27.
33. Gao H, Liang J, Duan J, Chen L, Li H, Zhen T, et al. A Prognosis Marker SLC2A3 Correlates With EMT and Immune Signature in Colorectal Cancer. Front Oncol. 2021;11:638099.
34. Itoh Y, Moriyama Y, Hasegawa T, Endo TA, Toyoda T, Gotoh Y. Scratch regulates neuronal migration onset via an epithelial-mesenchymal transition-like mechanism. Nat Neurosci. 2013;16(4):416-25.

35. Zhong J, Lin R, Wang G, Lin L, Ruan S, Liu W. KCNIP3 silence promotes proliferation and epithelial-mesenchymal transition of papillary thyroid carcinoma through activating Wnt/β-catenin pathway. Tissue Cell. 2022;75:101739.
36. Li W, Liang Y, Yang B, Sun H, Wu W. Downregulation of ARNT2 promotes tumor growth and predicts poor prognosis in human hepatocellular carcinoma. J Gastroenterol Hepatol. 2015;30(6):1085-93.
37. Guan Z, Zhang J, Wang J, Wang H, Zheng F, Peng J, et al. SOX1 down-regulates β-catenin and reverses malignant phenotype in nasopharyngeal carcinoma. Mol Cancer. 2014;13:257.
38. Álvarez C, Quiroz A, Benítez-Riquelme D, Riffo E, Castro AF, Pincheira R. SALL Proteins; Common and Antagonistic Roles in Cancer. Cancers (Basel). 2021;13(24).
39. Fang X, Cai Y, Liu J, Wang Z, Wu Q, Zhang Z, et al. Twist2 contributes to breast cancer progression by promoting an epithelial-mesenchymal transition and cancer stem-like cell self-renewal. Oncogene. 2011;30(47):4707-20.
40. Ma J, Li H, Gao Q, Zhang W, Zhu C, Chen J, et al. High MEIS3 Expression Indicates a Poor Prognosis for Patients with Stage II/III Colorectal Cancer. Front Biosci (Landmark Ed). 2023;28(12):338.
41. Du W, Liu X, Yang M, Wang W, Sun J. The Regulatory Role of PRRX1 in Cancer Epithelial-Mesenchymal Transition. Onco Targets Ther. 2021;14:4223-9.
42. Gao H, Teng C, Huang W, Peng J, Wang C. SOX2 Promotes the Epithelial to Mesenchymal Transition of Esophageal Squamous Cells by Modulating Slug Expression through the Activation of STAT3/HIF-α Signaling. Int J Mol Sci. 2015;16(9):21643-57.
43. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021;2(3):100141.
44. Starks RR, Biswas A, Jain A, Tuteja G. Combined analysis of dissimilar promoter accessibility and gene expression profiles identifies tissue-specific genes and actively repressed networks. Epigenetics Chromatin. 2019;12(1):16.
45. Jain PB, Guerreiro PS, Canato S, Janody F. The spectraplakin Dystonin antagonizes YAP activity and suppresses tumourigenesis. Sci Rep. 2019;9(1):19843.
46. Jung JW, Park SY, Seo EJ, Jang IH, Park Y, Lee D, et al. Functional expression of oxytocin receptors in pulp-dentin complex. Biomaterials. 2023;293:121977.
47. McCleskey BC, Penedo TL, Zhang K, Hameed O, Siegal GP, Wei S. GATA3 expression in advanced breast cancer: prognostic value and organ-specific relapse. Am J Clin Pathol. 2015;144(5):756-63.
48. Voduc D, Cheang M, Nielsen T. GATA-3 expression in breast cancer has a strong association with estrogen receptor but lacks independent prognostic value. Cancer Epidemiol Biomarkers Prev. 2008;17(2):365-73.
49. Querzoli P, Pedriali M, Rinaldi R, Secchiero P, Rossi PG, Kuhn E. GATA3 as an Adjunct Prognostic Factor in Breast Cancer Patients with Less Aggressive Disease: A Study with a Review of the Literature. Diagnostics (Basel). 2021;11(4).
50. Bai F, Zheng C, Liu X, Chan HL, Liu S, Ma J, et al. Loss of function of GATA3 induces basal-like mammary tumors. Theranostics. 2022;12(2):720-33.
51. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods. 2017;14(11):1083-6.

52. Cook DP, Vanderhyden BC. Context specificity of the EMT transcriptional response. Nat Commun. 2020;11(1):2142.
53. Cook DP, Vanderhyden BC. Transcriptional census of epithelial-mesenchymal plasticity in cancer. Sci Adv. 2022;8(1):eabi7640.
54. Zhao B, Ye X, Yu J, Li L, Li W, Li S, et al. TEAD mediates YAP-dependent gene induction and growth control. Genes Dev. 2008;22(14):1962-71.
55. Liu X, Li H, Rajurkar M, Li Q, Cotton JL, Ou J, et al. Tead and AP1 Coordinate Transcription and Motility. Cell Rep. 2016;14(5):1169-80.
56. Hu J, Su P, Jiao M, Bai X, Qi M, Liu H, et al. TRPS1 Suppresses Breast Cancer Epithelial-mesenchymal Transition Program as a Negative Regulator of SUZ12. Transl Oncol. 2018;11(2):416-25.
57. Jozwik KM, Carroll JS. Pioneer factors in hormone-dependent cancers. Nat Rev Cancer. 2012;12(6):381-5.
58. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003;100(14):8418-23.
59. Cruz RGB, Madden SF, Brennan K, Hopkins AM. A Transcriptional Link between HER2, JAM-A and FOXA1 in Breast Cancer. Cells. 2022;11(4).
60. Song Y, Washington MK, Crawford HC. Loss of FOXA1/2 is essential for the epithelial-to-mesenchymal transition in pancreatic cancer. Cancer Res. 2010;70(5):2115-25.
61. Lin M, Pan J, Chen Q, Xu Z, Lin X, Shi C. Overexpression of FOXA1 inhibits cell proliferation and EMT of human gastric cancer AGS cells. Gene. 2018;642:145-51.
62. Song B, Park SH, Zhao JC, Fong KW, Li S, Lee Y, et al. Targeting FOXA1-mediated repression of TGF-β signaling suppresses castration-resistant prostate cancer progression. J Clin Invest. 2019;129(2):569-82.
63. Wu L, Wang Y, Liu Y, Yu S, Xie H, Shi X, et al. A central role for TRPS1 in the control of cell cycle and cancer development. Oncotarget. 2014;5(17):7677-90.
64. Gai Z, Zhou G, Itoh S, Morimoto Y, Tanishima H, Hatamura I, et al. Trps1 functions downstream of Bmp7 in kidney development. J Am Soc Nephrol. 2009;20(11):2403-11.
65. Huang JZ, Chen M, Zeng M, Xu SH, Zou FY, Chen D, et al. Down-regulation of TRPS1 stimulates epithelial-mesenchymal transition and metastasis through repression of FOXA1. J Pathol. 2016;239(2):186-96.
66. Wei C, Wang Y, Li X. The role of Hippo signal pathway in breast cancer metastasis. Onco Targets Ther. 2018;11:2185-93.
67. Haslam SZ, Nummy KA. The ontogeny and cellular distribution of estrogen receptors in normal mouse mammary gland. J Steroid Biochem Mol Biol. 1992;42(6):589-95.
68. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell. 2005;122(1):33-43.
69. Asselin-Labat ML, Sutherland KD, Barker H, Thomas R, Shackleton M, Forrest NC, et al. Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. Nat Cell Biol. 2007;9(2):201-9.
70. Albergaria A, Paredes J, Sousa B, Milanezi F, Carneiro V, Bastos J, et al. Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours. Breast Cancer Res. 2009;11(3):R40.

71. Bernardo GM, Lozada KL, Miedler JD, Harburg G, Hewitt SC, Mosley JD, et al. FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. Development. 2010;137(12):2045-54.
72. Takaku M, Grimm SA, Wade PA. GATA3 in Breast Cancer: Tumor Suppressor or Oncogene? Gene Expr. 2015;16(4):163-8.
73. Voduc KD, Cheang MC, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. J Clin Oncol. 2010;28(10):1684-91.
74. Metovic J, Borella F, D'Alonzo M, Biglia N, Mangherini L, Tampieri C, et al. FOXA1 in Breast Cancer: A Luminal Marker with Promising Prognostic and Predictive Impact. Cancers (Basel). 2022;14(19).
75. Grinman DY, Boras-Granic K, Takyar FM, Dann P, Hens JR, Marmol C, et al. PTHrP induces STAT5 activation, secretory differentiation and accelerates mammary tumor development. Breast Cancer Res. 2022;24(1):30.
76. van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, et al. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell. 2018;174(3):716-29 e27.
77. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.
78. Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V, et al. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. Nucleic Acids Res. 2019;47(W1):W212-w24.
79. Fu X, Pereira R, De Angelis C, Veeraraghavan J, Nanda S, Qin L, et al. FOXA1 upregulation promotes enhancer and transcriptional reprogramming in endocrine-resistant breast cancer. Proc Natl Acad Sci U S A. 2019;116(52):26823-34.
80. Mahmood SR, Xie X, Hosny El Said N, Venit T, Gunsalus KC, Percipalle P. β-actin dependent chromatin remodeling mediates compartment level changes in 3D genome architecture. Nat Commun. 2021;12(1):5240.
81. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.
82. Adams EJ, Karthaus WR, Hoover E, Liu D, Gruet A, Zhang Z, et al. FOXA1 mutations alter pioneering activity, differentiation and prostate cancer phenotypes. Nature. 2019;571(7765):408-12.
83. Wang X, Yan J, Shen B, Wei G. Integrated Chromatin Accessibility and Transcriptome Landscapes of Doxorubicin-Resistant Breast Cancer Cells. Front Cell Dev Biol. 2021;9:708066.
84. Vitting-Seerup K, Sandelin A. IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. Bioinformatics. 2019;35(21):4469-71.
85. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12(6):996-1006.

Figure_1

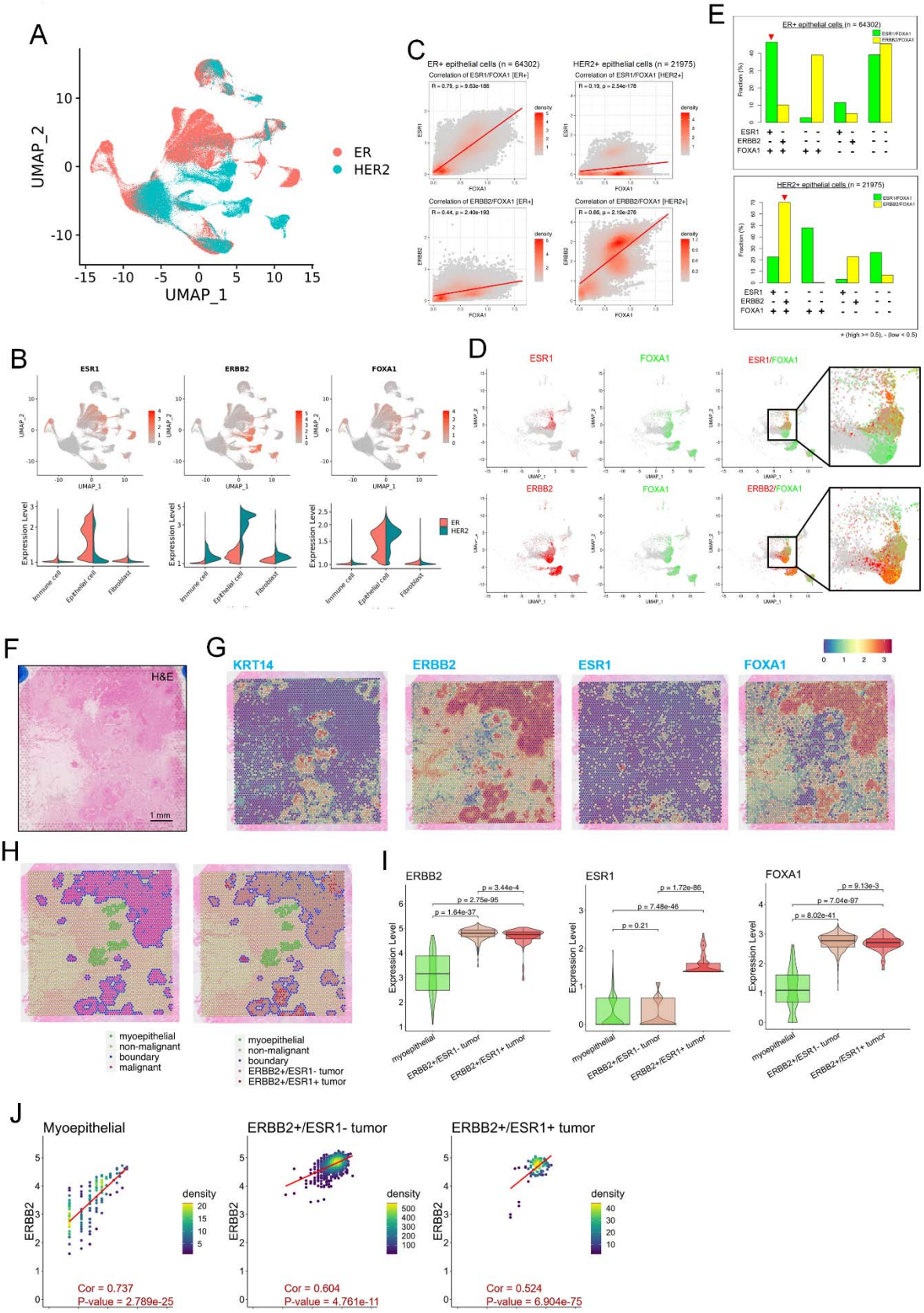

A
UMAP_2
UMAP_1
ER
HER2
B
ESR1
ERBB2
FOXA1
Expression Level
Immune cell
Epithelial cell
Fibroblast
C
ER+ epithelial cells (n = 64302)
HER2+ epithelial cells (n = 21975)
Correlation of ESR1/FOXA1 [ER+]
Correlation of ESR1/FOXA1 [HER2+]
Correlation of ERBB2/FOXA1 [ER+]
Correlation of ERBB2/FOXA1 [HER2+]
D
ESR1/FOXA1
ERBB2/FOXA1
E
Fraction (%)
F
H&E
1 mm
G
KRT14
H
myoepithelial
non-malignant
boundary
malignant
ERBB2+/ESR1- tumor
ERBB2+/ESR1+ tumor
I
p = 3.44e-4
p = 2.75e-95
p = 1.64e-37
p = 1.72e-86
p = 7.48e-46
p = 0.21
p = 9.13e-3
p = 7.04e-97
p = 8.02e-41
J
Myoepithelial
density
Cor = 0.737
P-value = 2.789e-25
Cor = 0.604
P-value = 4.761e-11
Cor = 0.524
P-value = 6.904e-75

Figure_2

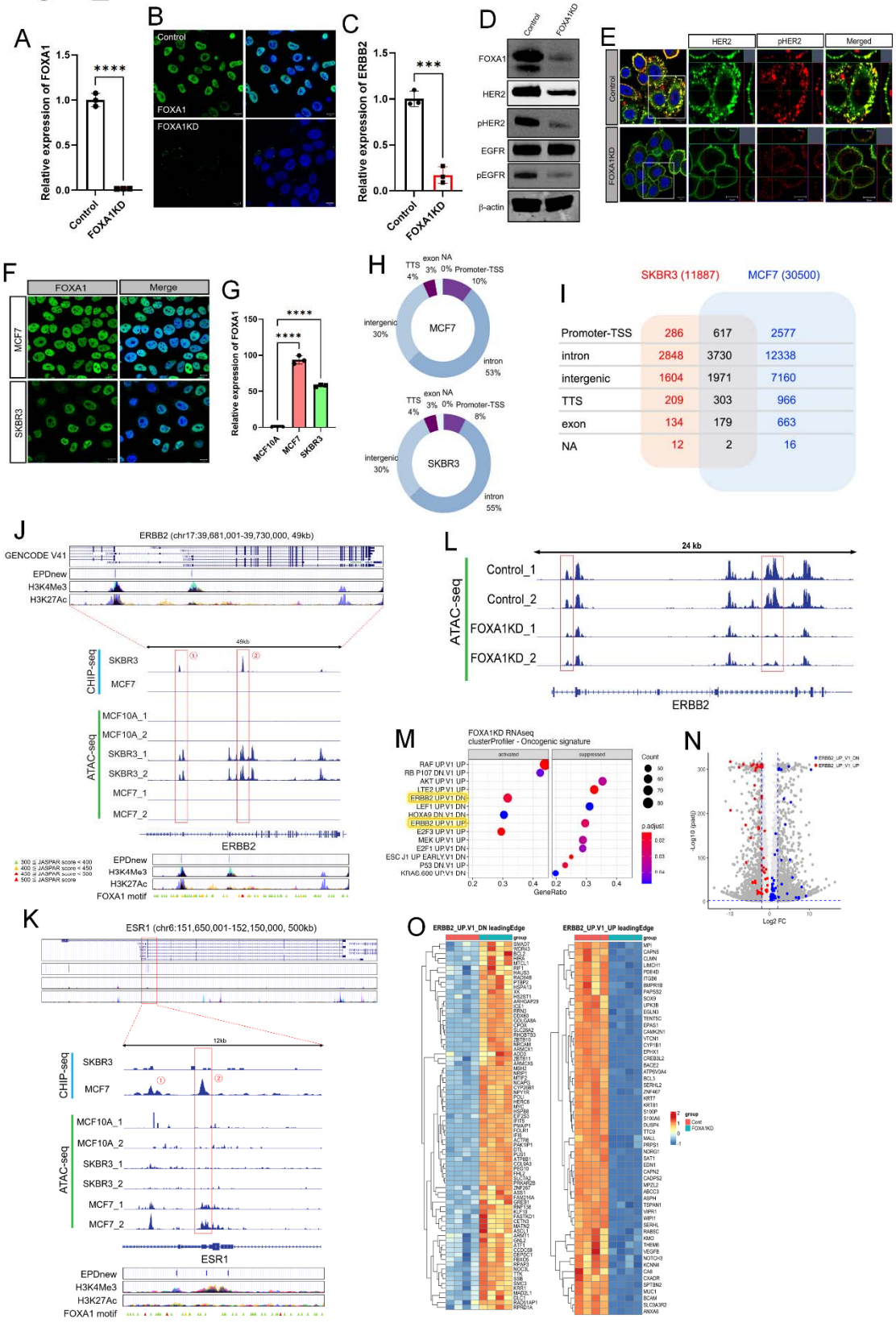

A
Relative expression of FOXA1
****
Control
FOXA1KD
B
Control
FOXA1
FOXA1KD
C
Relative expression of ERBB2
***
Control
FOXA1KD
D
Control
FOXA1KD
FOXA1
HER2
pHER2
EGFR
pEGFR
β-actin
E
HER2
pHER2
Merged
Control
FOXA1KD
F
FOXA1
Merge
MCF7
SKBR3
G
Relative expression of FOXA1
****
****
MCF10A
MCF7
SKBR3
H
TTS 4%
exon 3%
NA 0%
Promoter-TSS 10%
intergenic 30%
intron 53%
MCF7
TTS 4%
exon 3%
NA 0%
Promoter-TSS 8%
intergenic 30%
intron 55%
SKBR3
I
SKBR3 (11887)
MCF7 (30500)
Promoter-TSS 286 617 2577
intron 2848 3730 12338
intergenic 1604 1971 7160
TTS 209 303 966
exon 134 179 663
NA 12 2 16
J
ERBB2 (chr17:39,681,001-39,730,000, 49kb)
GENCODE V41
EPDnew
H3K4Me3
H3K27Ac
49kb
CHIP-seq
SKBR3
MCF7
ATAC-seq
MCF10A_1
MCF10A_2
SKBR3_1
SKBR3_2
MCF7_1
MCF7_2
ERBB2
EPDnew
H3K4Me3
H3K27Ac
FOXA1 motif
L
24 kb
ATAC-seq
Control_1
Control_2
FOXA1KD_1
FOXA1KD_2
ERBB2
M
FOXA1KD RNAseq
clusterProfiler - Oncogenic signature
activated
suppressed
RAF UP.V1 UP
RB P107 DN.V1 UP
AKT UP.V1 UP
LTE2 UP.V1 UP
ERBB2 UP.V1 DN
LEF1 UP.V1 DN
HOXA9 DN.V1 DN
ERBB2 UP.V1 UP
E2F3 UP.V1 UP
MEK UP.V1 UP
E2F1 UP.V1 DN
ESC J1 UP EARLY.V1 DN
P53 DN.V1 UP
KRAS.600 UP.V1 DN
Count
p.adjust
GeneRatio
N
ERBB2_UP.V1_DN
ERBB2_UP.V1_UP
-Log10 (padj)
Log2 FC
K
ESR1 (chr6:151,650,001-152,150,000, 500kb)
12kb
CHIP-seq
SKBR3
MCF7
ATAC-seq
MCF10A_1
MCF10A_2
SKBR3_1
SKBR3_2
MCF7_1
MCF7_2
ESR1
EPDnew
H3K4Me3
H3K27Ac
FOXA1 motif
O
ERBB2_UP.V1_DN leadingEdge
ERBB2_UP.V1_UP leadingEdge
group
Cont
FOXA1KD

Figure_3

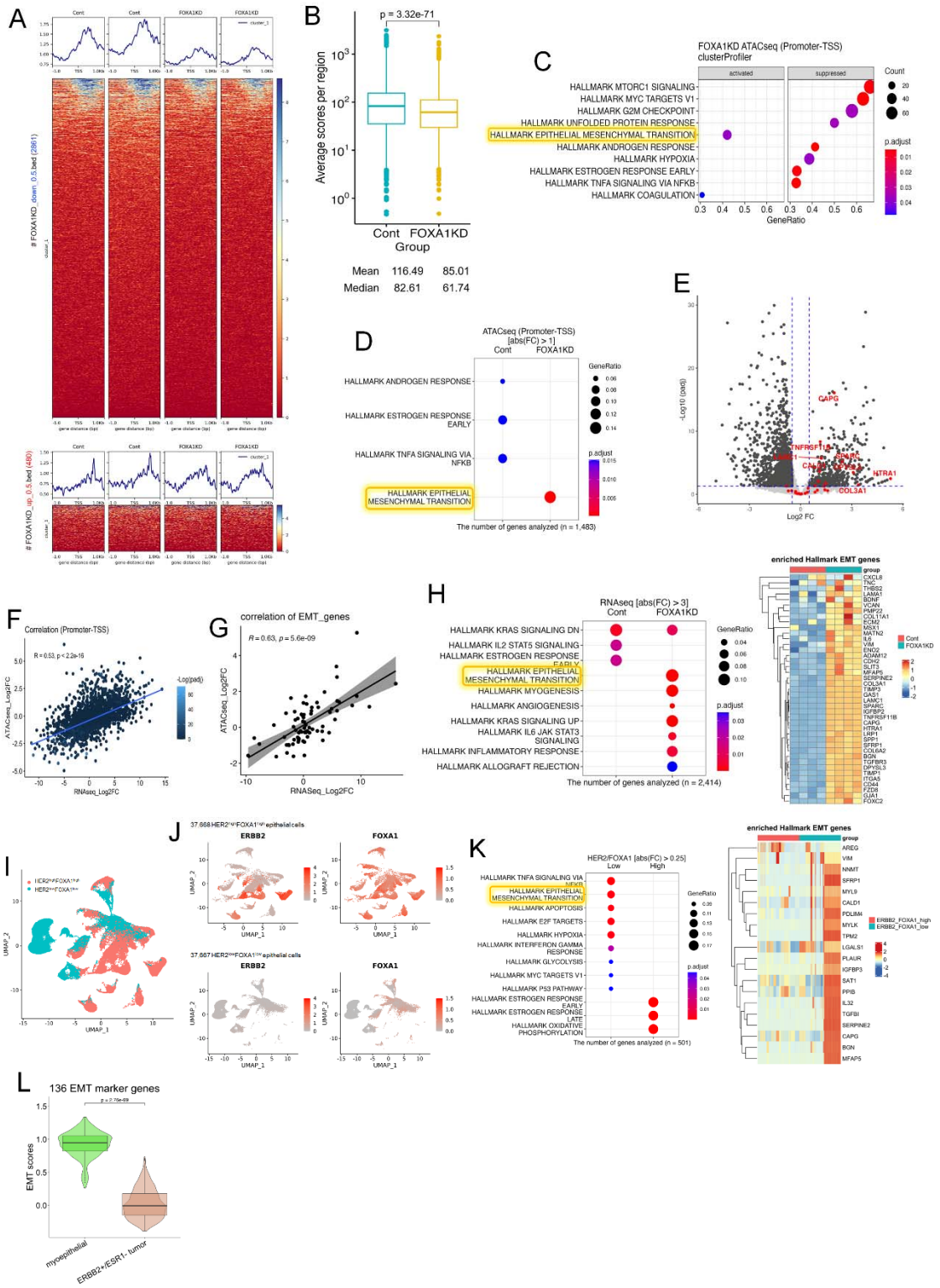
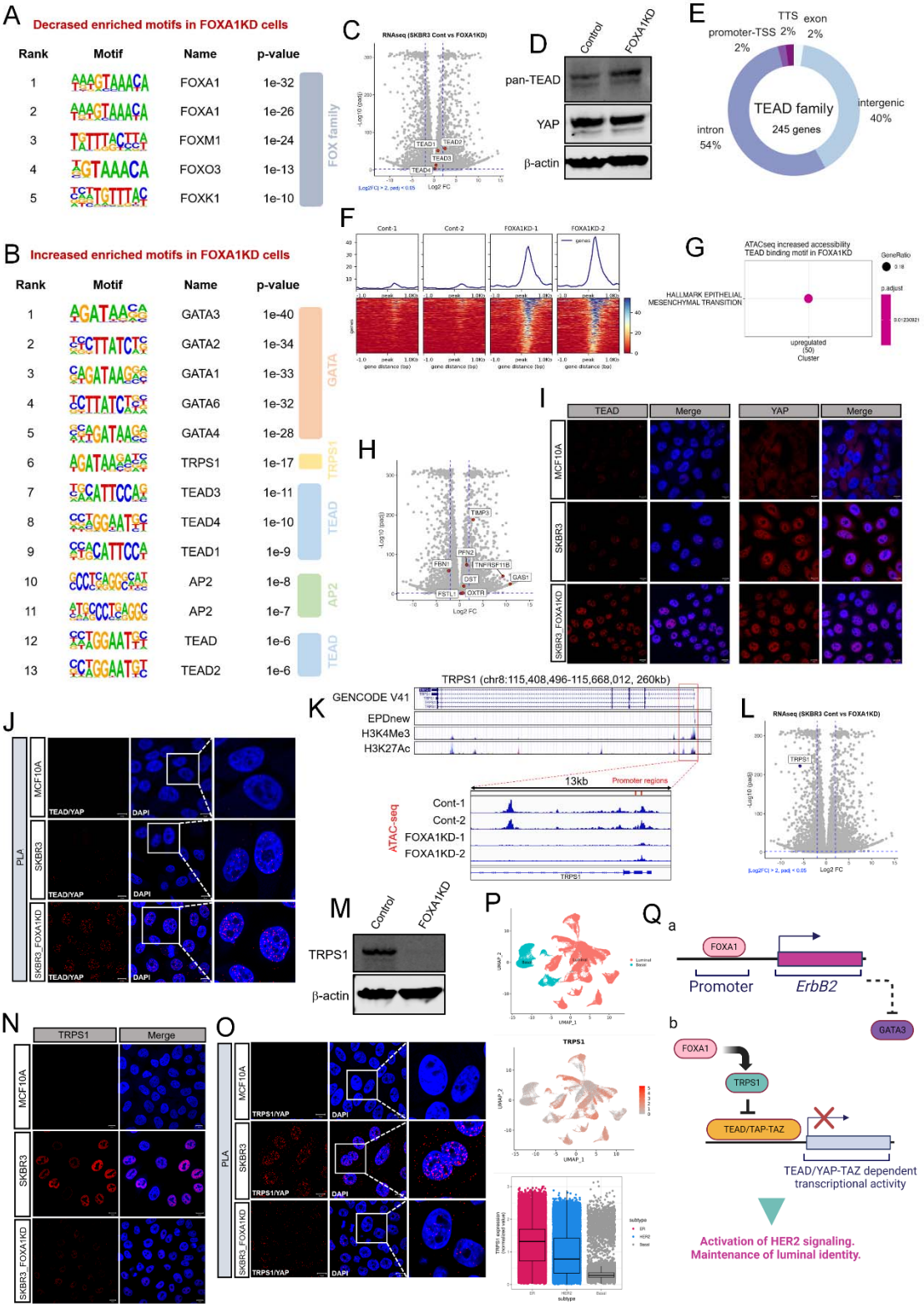

A
Cont
FOXA1KD
FOXA1KD_down_0.5.bed (2861)
FOXA1KD_up_0.5.bed (480)
B
p = 3.32e-71
Average scores per region
Cont
FOXA1KD
Group
Mean 116.49 85.01
Median 82.61 61.74
C
FOXA1KD ATACseq (Promoter-TSS)
clusterProfiler
activated
suppressed
HALLMARK MTORC1 SIGNALING
HALLMARK MYC TARGETS V1
HALLMARK G2M CHECKPOINT
HALLMARK UNFOLDED PROTEIN RESPONSE
HALLMARK EPITHELIAL MESENCHYMAL TRANSITION
HALLMARK ANDROGEN RESPONSE
HALLMARK HYPOXIA
HALLMARK ESTROGEN RESPONSE EARLY
HALLMARK TNFA SIGNALING VIA NFKB
HALLMARK COAGULATION
GeneRatio
Count
p.adjust
D
ATACseq (Promoter-TSS)
[abs(FC) > 1]
Cont
FOXA1KD
HALLMARK ANDROGEN RESPONSE
HALLMARK ESTROGEN RESPONSE EARLY
HALLMARK TNFA SIGNALING VIA NFKB
HALLMARK EPITHELIAL MESENCHYMAL TRANSITION
GeneRatio
p.adjust
The number of genes analyzed (n = 1,483)
E
-Log10 (padj)
Log2 FC
CAPG
TNFRSF11B
HTRA1
COL3A1
F
Correlation (Promoter-TSS)
R = 0.53, p < 2.2e-16
ATACseq_Log2FC
RNAseq_Log2FC
-Log(padj)
G
correlation of EMT_genes
R = 0.63, p = 5.6e-09
ATACseq_Log2FC
RNASeq_Log2FC
H
RNAseq [abs(FC) > 3]
Cont
FOXA1KD
HALLMARK KRAS SIGNALING DN
HALLMARK IL2 STAT5 SIGNALING
HALLMARK ESTROGEN RESPONSE EARLY
HALLMARK EPITHELIAL MESENCHYMAL TRANSITION
HALLMARK MYOGENESIS
HALLMARK ANGIOGENESIS
HALLMARK KRAS SIGNALING UP
HALLMARK IL6 JAK STAT3 SIGNALING
HALLMARK INFLAMMATORY RESPONSE
HALLMARK ALLOGRAFT REJECTION
GeneRatio
p.adjust
The number of genes analyzed (n = 2,414)
enriched Hallmark EMT genes
group
Cont
FOXA1KD
I
UMAP_1
UMAP_2
J
ERBB2
FOXA1
UMAP_1
UMAP_2
K
HER2/FOXA1 [abs(FC) > 0.25]
Low
High
HALLMARK TNFA SIGNALING VIA NFKB
HALLMARK EPITHELIAL MESENCHYMAL TRANSITION
HALLMARK APOPTOSIS
HALLMARK E2F TARGETS
HALLMARK HYPOXIA
HALLMARK INTERFERON GAMMA RESPONSE
HALLMARK GLYCOLYSIS
HALLMARK MYC TARGETS V1
HALLMARK P53 PATHWAY
HALLMARK ESTROGEN RESPONSE EARLY
HALLMARK ESTROGEN RESPONSE LATE
HALLMARK OXIDATIVE PHOSPHORYLATION
GeneRatio
p.adjust
The number of genes analyzed (n = 501)
enriched Hallmark EMT genes
group
ERBB2_FOXA1_high
ERBB2_FOXA1_low
L
136 EMT marker genes
EMT scores
myoepithelial
ERBB2+/ESR1- tumor

Figure_4


A
Decrased enriched motifs in FOXA1KD cells
Rank Motif Name p-value
1 FOXA1 1e-32
2 FOXA1 1e-26
3 FOXM1 1e-24
4 FOXO3 1e-13
5 FOXK1 1e-10
FOX family
B
Increased enriched motifs in FOXA1KD cells
Rank Motif Name p-value
1 GATA3 1e-40
2 GATA2 1e-34
3 GATA1 1e-33
4 GATA6 1e-32
5 GATA4 1e-28
6 TRPS1 1e-17
7 TEAD3 1e-11
8 TEAD4 1e-10
9 TEAD1 1e-9
10 AP2 1e-8
11 AP2 1e-7
12 TEAD 1e-6
13 TEAD2 1e-6
GATA
TRPS1
TEAD
AP2
TEAD
C
RNAseq (SKBR3 Cont vs FOXA1KD)
D
Control
FOXA1KD
pan-TEAD
YAP
β-actin
E
TTS 2%
exon 2%
promoter-TSS 2%
TEAD family
245 genes
intergenic 40%
intron 54%
F
Cont-1
Cont-2
FOXA1KD-1
FOXA1KD-2
G
ATACseq increased accessibility TEAD binding motif in FOXA1KD
HALLMARK EPITHELIAL MESENCHYMAL TRANSITION
upregulated (50)
Cluster
H
I
TEAD
Merge
YAP
Merge
MCF10A
SKBR3
SKBR3_FOXA1KD
J
PLA
MCF10A
SKBR3
SKBR3_FOXA1KD
TEAD/YAP
DAPI
K
TRPS1 (chr8:115,408,496-115,668,012, 260kb)
GENCODE V41
EPDnew
H3K4Me3
H3K27Ac
13kb
Promoter regions
ATAC-seq
Cont-1
Cont-2
FOXA1KD-1
FOXA1KD-2
TRPS1
L
RNAseq (SKBR3 Cont vs FOXA1KD)
M
Control
FOXA1KD
TRPS1
β-actin
P
Q
a
FOXA1
Promoter
ErbB2
GATA3
b
FOXA1
TRPS1
TEAD/TAP-TAZ
TEAD/YAP-TAZ dependent transcriptional activity
Activation of HER2 signaling.
Maintenance of luminal identity.
N
TRPS1
Merge
MCF10A
SKBR3
SKBR3_FOXA1KD
O
PLA
MCF10A
SKBR3
SKBR3_FOXA1KD
TRPS1/YAP
DAPI

Supplementary figure legends

Supplementary Figure 1

A. Mean Gene Effect scores across 1,255 cell lines for *FOXA1* gene overlapping the RNAi and CRISPR datasets. Yellow box near x-axis -1 indicates strong dependency in breast and prostate cancer cell lines. **B.** Distribution of FOXA1 Gene Effect scores among breast cancer lineage sub-subtype. **C.** Bar plots show the top enriched lineages with FOXA1 Gene Effect score.

Supplementary Figure 2

A. Uniform Manifold Approximation and Projection (UMAP) of human breast tumor cells (n = 148,945). Epithelial cells, immune cells and fibroblasts in the tumor tissue were represented. **B-D**. Feature plot of each cell types showing the representative cell type-specific markers.

Supplementary Figure 3

A. *ErbB2*, *ESR1* and *FOXA1* gene expressions are shown as Xenium transcript localization plot (decoded at Q ≥ 20) (upper panel). Invasive tumor, DCIS and Myoepithelial cells are shown individually with highlighted color (bottom left panel). Closer view to reveal detailed resolution of cell boundaries and selected transcripts (bottom right panel). **B.** Boxplot showing the average scale.data between Myoepithelial (n = 9,445), DCIS (*ErbB2*+/*ESR1*+, n = 25,461) and Invasive tumor (*ErbB2*+/*ESR1*-, n = 36,170). Wilcoxon rank sum test was performed, and the P-value was calculated from a down-sampled dataset by 1,000 (*FOXA1*) or 2,000 (*ErbB2*) cells to avoid zero inflation labeled at the plot.

Supplementary Figure 4

A-B. Summary of FOXA1 binding sites identified with JASPAR 2020 minimum score 300 within the ChIP-seq peak regions from *ErbB2* or *ESR1* gene, respectively.

Supplementary Figure 5

A. Venn diagram showing the cell line specific and common FOXA1 binding peaks and their annotated genes including Promoter-TSS, intron, exon, and TTS. **B.** Dot plot showing the WikiPathway_2023_Human ORA performed for 3,152 genes that more than twice their peak scores in SKBR3 compared to MCF7. **C.** Dot plot showing the Oncogenic_Signature ORA performed for 3,152 genes exclusive in SKBR3(top) and 4,241 genes that exclusive in MCF7(bottom), respectively. **D.** Dot plot showing the Oncogenic_Signature ORA performed for 1,446 genes that overlapped 2,246 peaks with their score ratio less than 2.

Supplementary Figure 6

A. UMAP of basal and luminal subtypes breast tumor cells were represented. **B**. UMAP (upper panel) and violin plots for *FOXA1*, *ESR1* and *ERBB2*. **C**. Feature plot showing the representative basal (upper panel) or luminal type (bottom) specific marker gene expression. **D.** Dot plot showing the significantly (adjusted p-value < 0.05, Benjamini-Hochberg correction) enriched Hallmark gene sets by GSEA using significantly up-regulated genes (Log2FC > 0, adjusted p-value < 0.05) in the FOXA1KD-SKBR3 cells compared to the control. **E**. Upstream regulators uncovered by Ingenuity pathway analysis.

Supplementary Figure 7

A. Heatmap representing the signal enrichment of control and FOXA1KD-SKBR3 cells. ATAC-seq reads in global 2-kb regions centered on the TSS sites. **B.** Boxplot showing the average scores per region from ATAC-seq for control and FOXA1KD-SKBR3 cells. Wilcoxon rank sum test was performed, and the P-value was labeled at the plot.

Supplementary Figure 8

A. Heatmap and read density representing the signal enrichment of control and FOXA1KD-SKBR3 cells. ATAC-seq reads in 2kb regions centered on the TSS sites which belonging to the Oncogenic_ERBB2_UP.V1_UP leading-edge subset genes. Boxplot showing the change in average scores per region (left) from corresponding heatmap or peak scores (right) from ATAC-seq for control and FOXA1KD SKBR3 cells which leading-edge subset genes were involved in Oncogenic_ERBB2_UP.V1_UP. Wilcoxon rank sum test was performed and P-value was labeled at plot. **B-C.** Dot plot showing the significantly (adjusted p-value < 0.05, Benjamini-Hochberg correction) enriched Hallmark gene sets by GSEA (B) and ORA (C) in the FOXA1KD-SKBR3 cells compared to control. Genes corresponded with DARs annotated as all regions (promoter-TSS, exon, TTS, intron and intergenic). **D.** Heatmap and read density representing the signal enrichment of control and FOXA1KD-SKBR3 cells. ATAC-seq reads in 2kb regions centered on the TSS sites which belonging to the Hallmark_EMT leading-edge subset genes. Boxplot showing the change in average scores per region (left) from corresponding heatmap or peak scores (right) from ATAC-seq for control and FOXA1KD-SKBR3 cells which leading-edge subset genes were involved in Hallmark_EMT. Wilcoxon rank sum test was performed and P-value was labeled at plot. **E.** Summary of pearson correlation between *FOXA1* and subset of EMT leading-edge gene's log2(TPM+1) expression from 22Q2 Public in each

breast cancer sub-subtype cell lines. **F.** Scatter plot of breast cancer cell lines categorized by lineage sub-subtype (n = 63). Pearson correlation between *FOXA1* and subset of EMT leading-edge gene's log2(TPM+1) expressions from 22Q2 Public were computed in ER-positive or HER2-positive breast cancer cell lines. Red color indicates as significant with p-value less than 0.05. **G.** UMAP dimensional reduction of all epithelial cell. *FOXA1*-high and -low groups from HER2-positive cells were divided according to the corresponding gene expression level. **H**. Dot plot showing the significantly enriched Hallmark gene sets by ORA performed for DEGs (abs|FC| > 0.25 utilized MAST) in the *FOXA1*-high compared to -low groups from HER2-positive cells. Heatmap showing the expression of nine enriched genes in Hallmark EMT gene set.

Supplementary Figure 9

A. *ErbB2*, *ESR1*, *FOXA1* and *EpCAM* gene expressions are shown as Xenium transcript localization plot (decoded at Q ≥ 20) (upper panel). Invasive tumor, DCIS and Myoepithelial cells are shown individually with highlighted color (bottom left panel). Closer view to reveal detailed resolution of cell boundaries and selected transcripts (bottom right panel). **B.** Boxplot showing the average scale.data of selected transcripts between Myoepithelial (n = 9,445), DCIS (*ErbB2*+/*ESR1*+, n = 25,461) and Invasive tumor (*ErbB2* +/*ESR1*-, n = 36,170). Wilcoxon rank sum test was performed, and the P-value was calculated from a down-sampled dataset by 2,000 cells to avoid zero inflation labeled at the plot.

Supplementary Figure 10

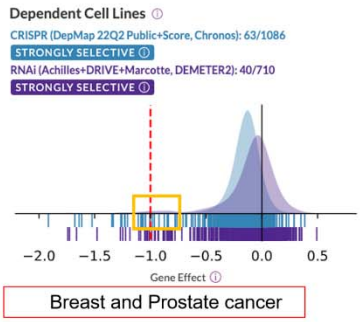
A. Volcano plot showing the log2FC and adjusted P-values of all expressed genes (light grey) from the differential test between control and FOXA1KD-SKBR3 cells. The red dots represent GATA family genes. **B.** Heatmap showing the relative expression of GATA3 and GATA4 target genes with confidence level **C.** Histogram showing the relative expression levels of *GATA3* from qRT-PCR in control and FOXA1KD-SKBR3 cells. **D.** Immunofluorescence images of GATA3 in control and FOXA1KD-SKBR3 cells. **E.** UMAP dimensional reduction of all *ESR1*^{low} luminal subtype epithelial cell. *ErbB2*-high and -low groups were divided according to the corresponding gene expression level. **F.** Pearson correlation coefficient between *FOXA1*, *ERBB2* and *GATA3* from *ESR1*^{low} luminal subtype epithelial cells. **G.** Heatmap showing the regulon activity of *ErbB2*-high and -low groups from *ESR1*^{low} epithelial cells. The number in the name of x-axis represents genes targeted by the corresponding TF. **H.** Distribution of HER2 Gene Effect scores of breast cancer lineage sub-subtype (left panel). Scatter plot of breast cancer cell lines categorized by lineage sub-subtype (n = 63) (middle panel). Pearson correlation between *ErbB2* and *GATA3* gene's log2(TPM+1) expressions from 22Q2 Public were computed according to lineage sub-subtype (right panel). Bar graphs represent the mean±SEM. *** denotes $p<0.0005$.

Supplementary Figure 11

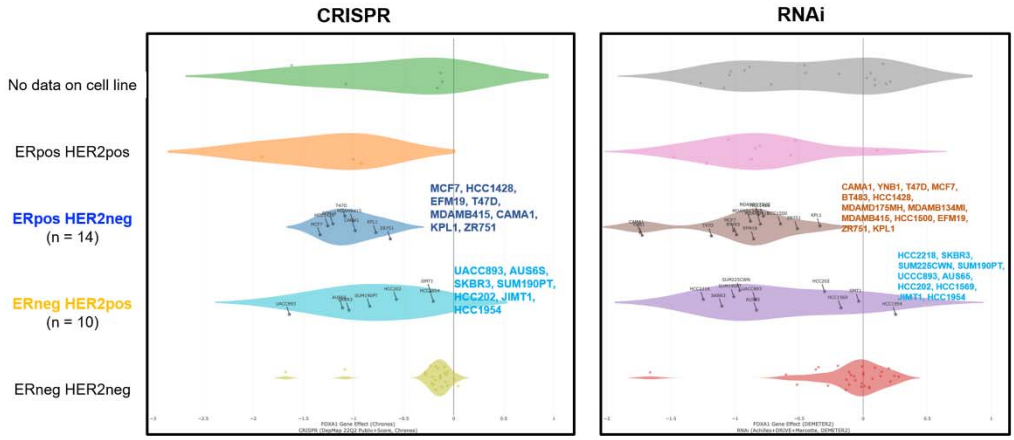
A. Visium spatial distribution of TEADs and YAP/TAZ signaling pathway related genes expressions are shown as $\log_2$ (normalized UMI counts, left panel). Violin plot illustrating the corresponding genes expression of each cell type (right panel). **B.** Violin plot illustrating the TEAD module scores between myoepithelial and *ErbB2*+/*ESR1*- tumor. The box centerlines depict the medians, and the edges depict the first/third quartiles.

Supplemental Figure_1

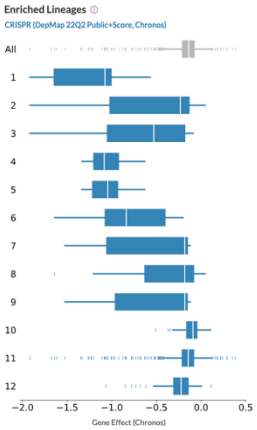
A



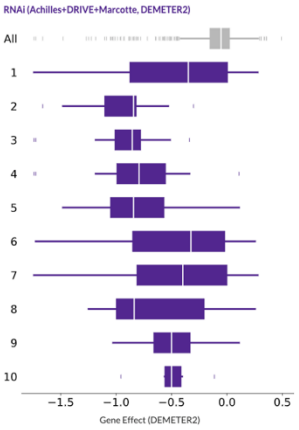
B



C



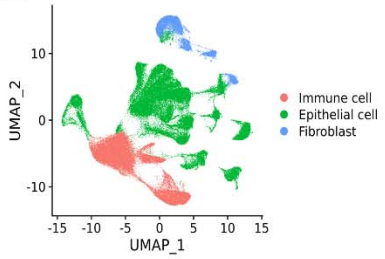
1. **HER2 Amp** (9.5e-53) n=8
2. Breast (3.3e-43) n=47
3. Breast Carcinoma (2.0e-37) n=24
4. Luminal (2.3e-37) n=9
5. **ERpos HER2neg** (2.2e-33) n=8
6. **ERneg HER2pos** (4.5e-16) n=7
7. Prostate Adenocarcinoma (3.3e-08) n=8
8. Breast Ductal Carcinoma (5.9e-08) n=21
9. Prostate (1.7e-07) n=9
10. Haematopoietic And Lymphoid (1.1e-04) n=118
11. Solid (1.1e-04) n=968
12. NSCLC Adenocarcinoma (3.7e-04) n=50



1. Breast (1.0e-38) n=79
2. **HER2 Amp** (2.5e-33) n=13
3. **ERpos HER2neg** (1.2e-32) n=13
4. Luminal (6.5e-28) n=16
5. ERpos HER2pos (9.8e-16) n=9
6. Breast Ductal Carcinoma (1.0e-14) n=34
7. Breast Carcinoma (6.9e-13) n=28
8. **ERneg HER2pos** (5.5e-10) n=10
9. Prostate (4.2e-05) n=8
10. Prostate Adenocarcinoma (2.7e-04) n=6

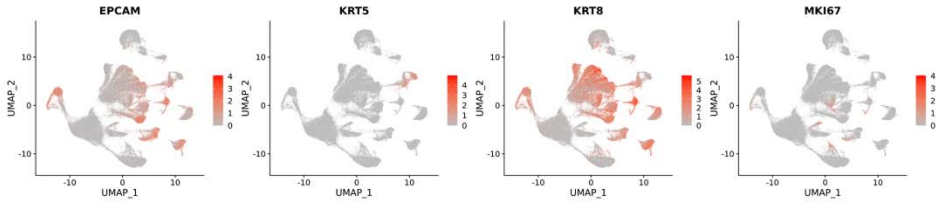
Supplemental Figure_2

A



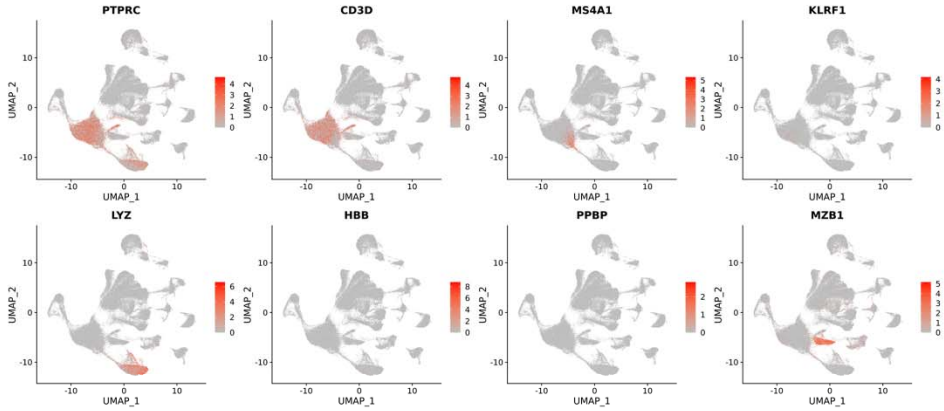
B

Epithelial cells



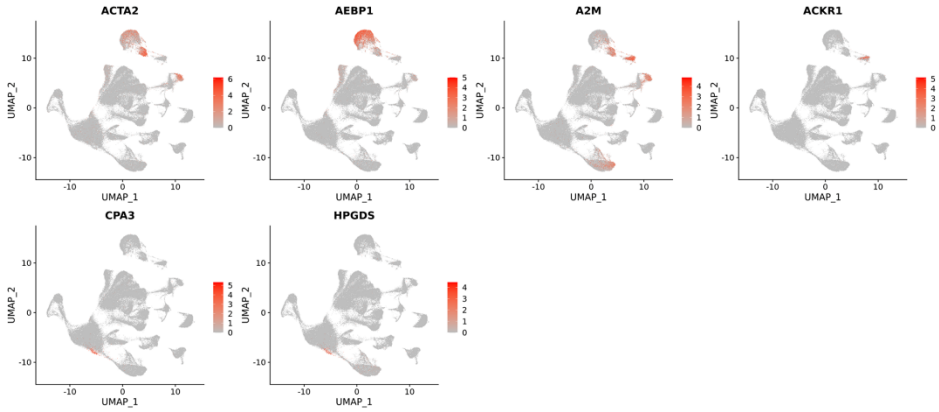
C

Immune cells



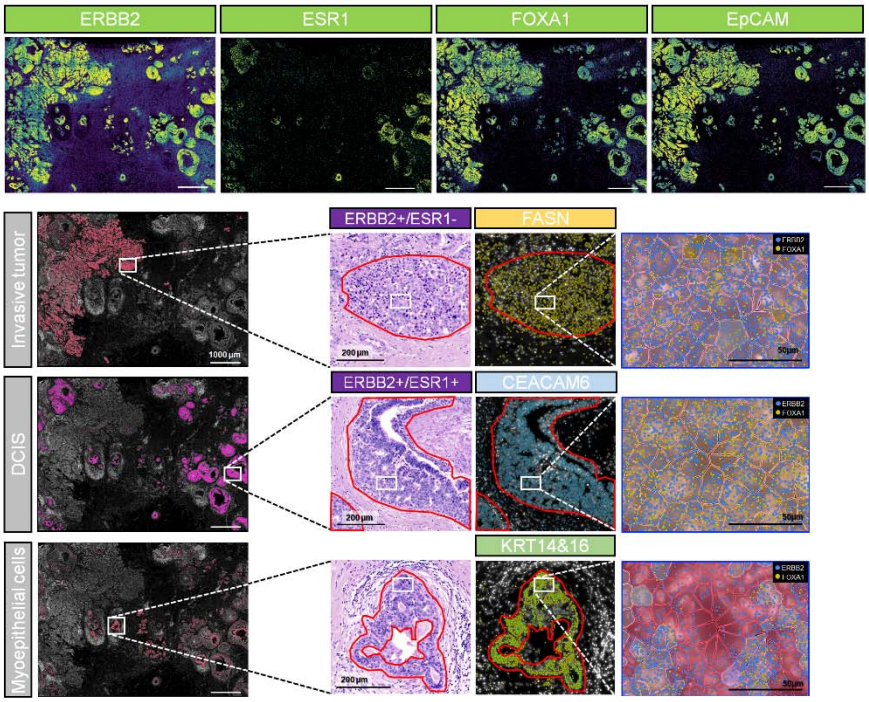
D

Fibroblast

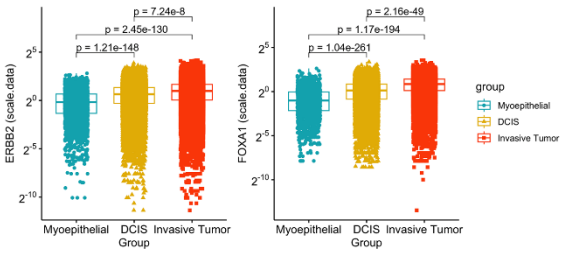


Supplemental Figure_3

A



B



Supplemental Figure_4

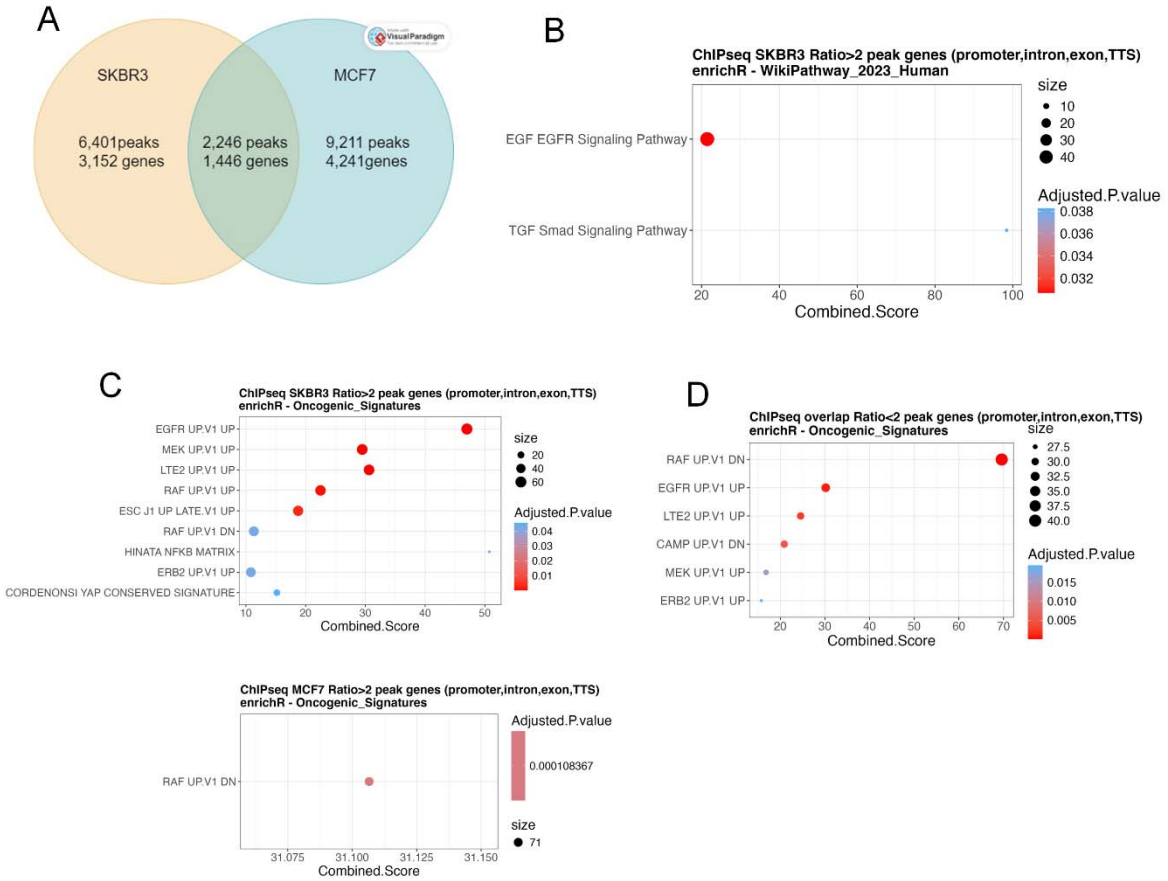
A

		Score	Position	Strand	Sequence
①	1	370	chr17:39687050-39687061	+	GAGTAAATAAAG
	2	304	chr17:39687137-39687148	+	TAGTAAAGAGGA
②	7	327	chr17:39704747-39704758	+	CTGCAAACACTT
	8	445	chr17:39704795-39704806	+	AGGTAAACAAAC
	9	324	chr17:39705208-39705219	-	TGTTGTTTTCCT

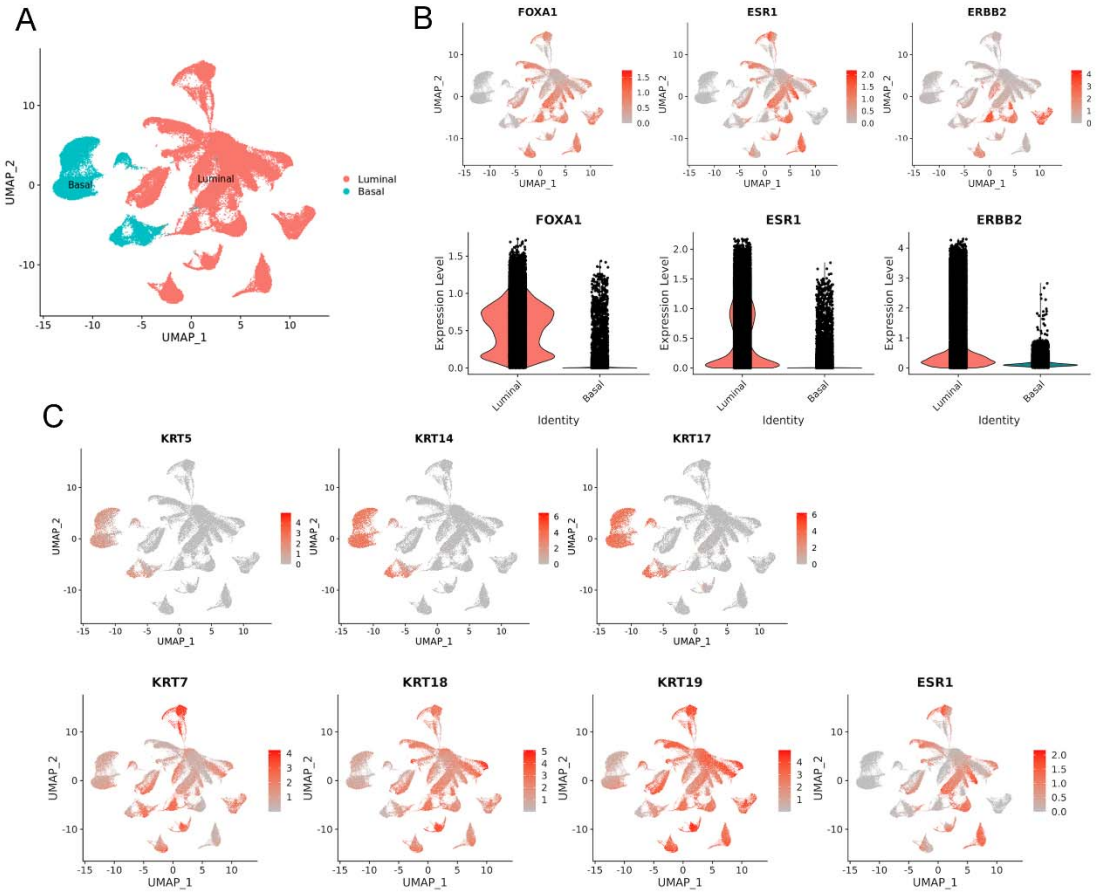
B

		Score	Position	Strand	Sequence
①	1	372	chr6:151803110-151803121	-	ATTTGTTGACTT
	2	317	chr6:151803594-151803605	+	CAGTAAAAAAAA
	3	487	chr6:151803614-151803625	+	AAGTAAACAGGT
	4	334	chr6:151803993-151804004	-	CAATCTTTACCC
	5	318	chr6:151804335-151804346	+	TAGTGAACAATT
	6	336	chr6:151804447-151804458	-	CATTGTTGACCT
	7	369	chr6:151804505-151804516	-	CAATATTTATTT
	8	329	chr6:151804509-151804520	-	ATTTATTTATAT
	9	367	chr6:151804608-151804619	-	ATATATTTGCAT
②	10	369	chr6:151807103-151807114	-	ATATGTATACCC
	11	351	chr6:151807264-151807275	+	ATGTACACACTA
	12	320	chr6:151807566-151807577	-	TGATGTTTAAGC

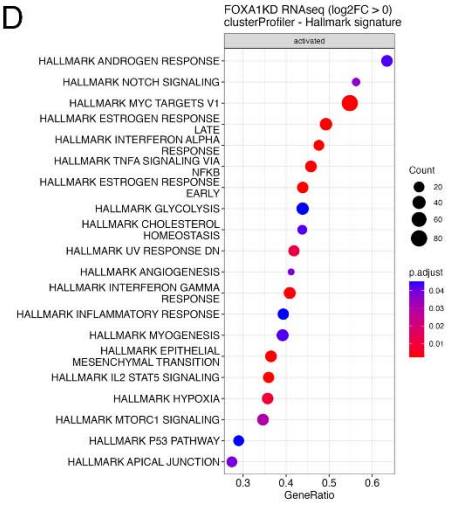
Supplemental Figure_5



Supplemental Figure_6



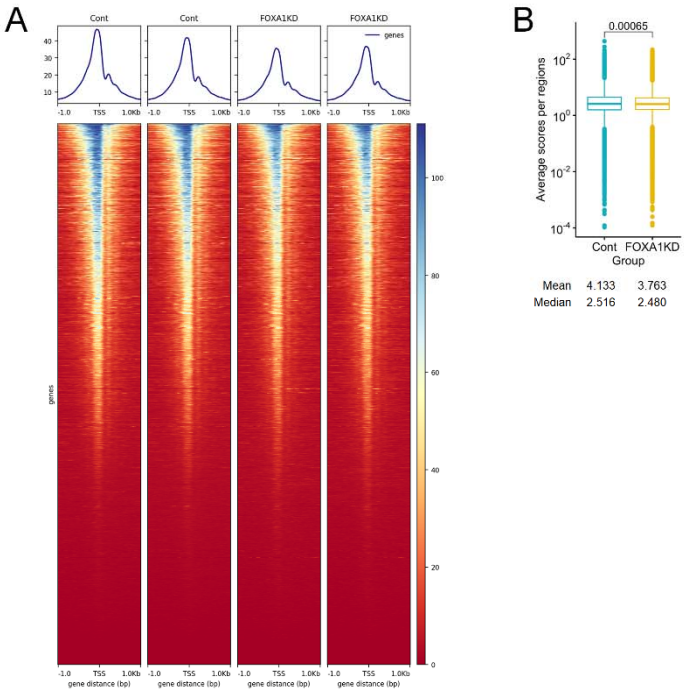
D



E

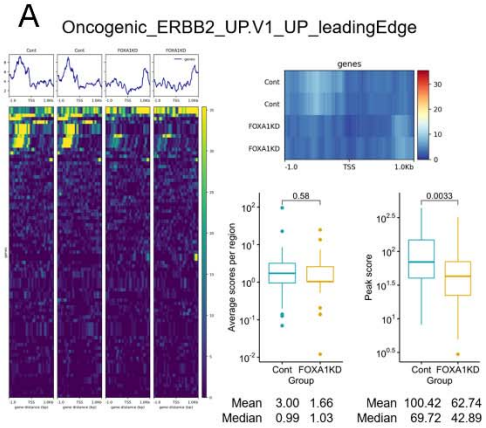
Upstream Regulator	Expr Fold Change	Predicted Activation State	Activation z-score	p-value of overlap
SP1	0.296	Activated	2.154	4.62E-12
CBX5	↑ 1.078		2.13	1.96E-10
SNAI1	↑ 1.267		2.091	1.05E-08
ERG	↑ 8.329		2.03	5.43E-08
STAT1	0.047		2.873	1.02E-05
SMAD4	↑ 1.396		2.235	3.43E-05
MED16	-0.569		2.329	1.07E-04
SOX11	↓ -2.940		2.375	1.63E-03
IRF3	0.591		2.837	2.03E-02
ZNF382	0.517		2.213	2.54E-02
FOXM1	-0.118		2.534	2.85E-02
HOXB13	↑ 4.774		2.714	3.56E-02
GATA6	↓ -1.811		2.718	4.25E-02
SPDEF	↓ -8.764	Inhibited	-2.322	7.72E-06
WT1	0		-2.109	2.56E-05
SMAD1	↓ -1.783		-2.019	2.31E-04
DNAJB6	-0.534		-2.55	6.83E-04
ZBED6	-0.197		-2.111	1.87E-02
TBX3	↑ 1.546		-2	2.54E-02

Supplemental Figure_7

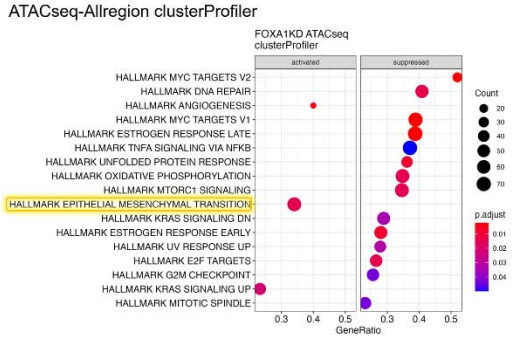


Supplemental Figure_8

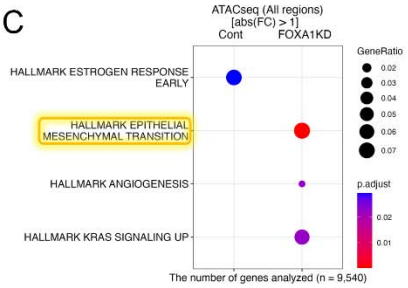
A



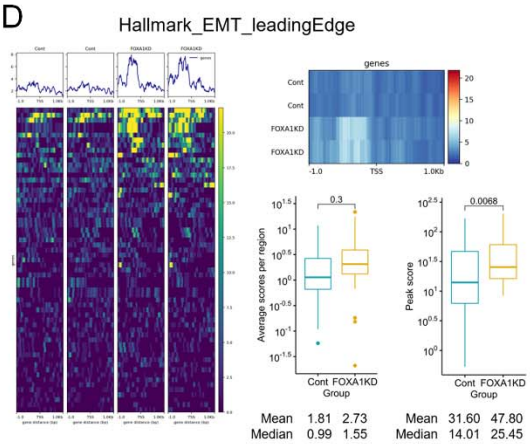
B



C



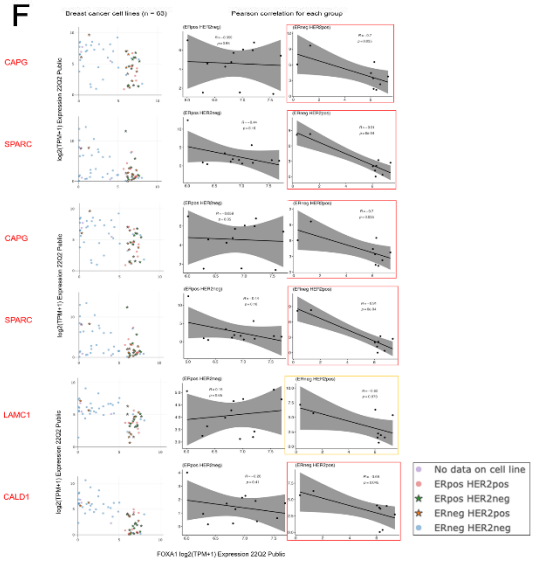
D



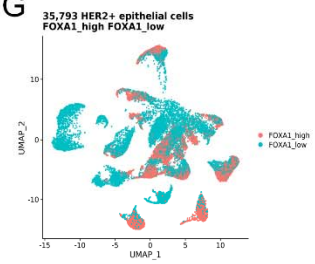
E

EMT leading edge genes	FOXA1			
	ERpos HER2neg (n=12)		ERneg HER2pos (n=9)	
	Pearson	P-value	Pearson	P-value
CAPG	-0.059	8.55E-1	-0.702	3.51E-2
TNFRSF11B	-0.198	5.36E-1	0.222	5.66E-1
SPARC	-0.436	1.56E-1	-0.912	6.00E-4
LAMC1	0.145	6.52E-1	-0.623	7.90E-2
CALD1	-0.262	4.11E-1	-0.677	4.51E-2
DPYSL3	0.247	4.39E-1	-0.951	7.92E-5
HTRA1	-0.120	7.10E-1	-0.751	1.96E-2
COL3A1	-0.082	8.00E-1	0.254	5.10E-1

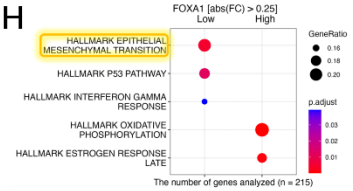
F

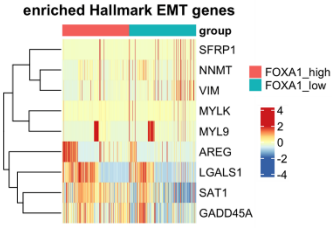


G



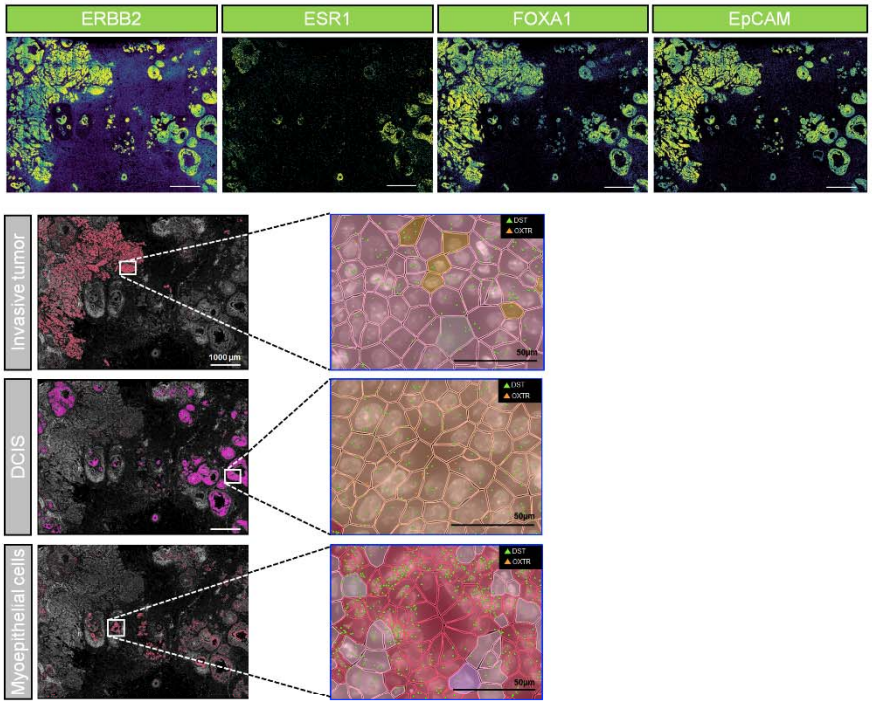
H



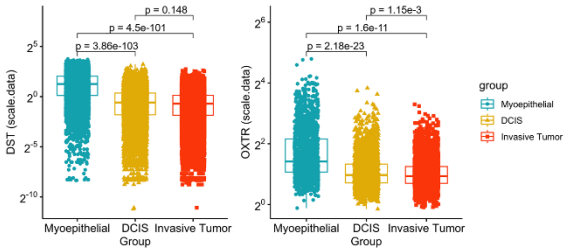


Supplemental Figure_9

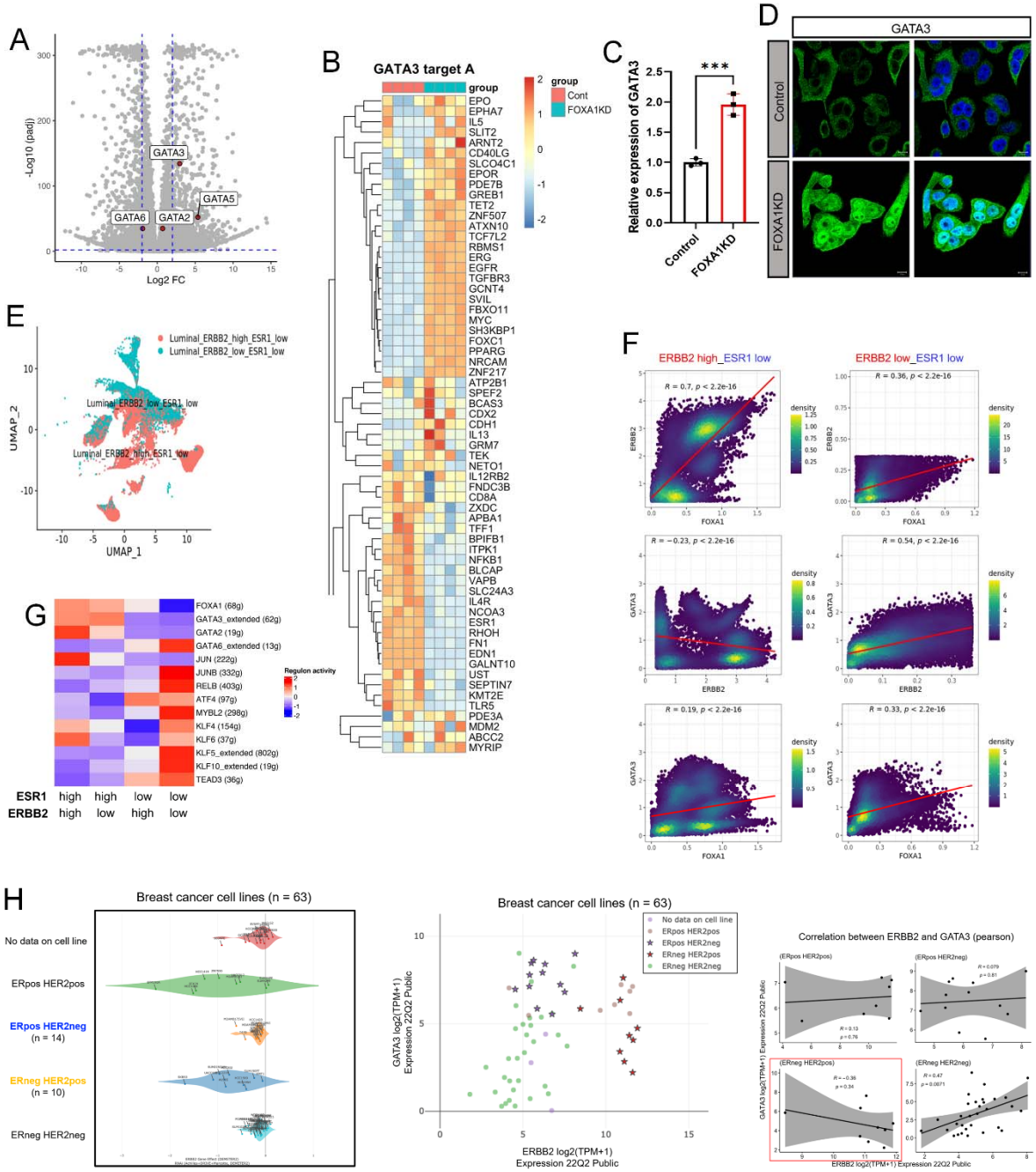
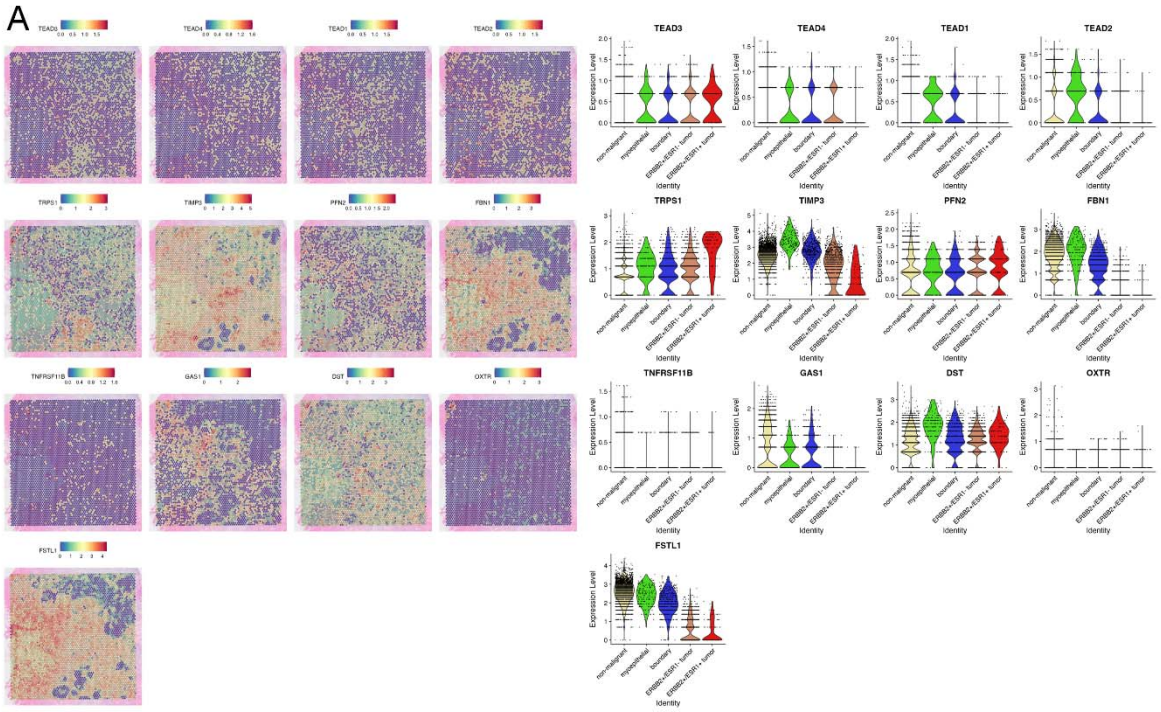
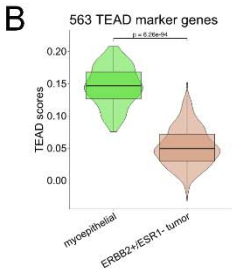
A



B



Supplemental Figure_10



Supplemental Figure_11