

Transcriptome analysis reveals the potential role of neural factor EN1 for long-terms survival in estrogen receptor-independent breast cancer

He Ren,^{1,4} Shan Liu,^{1,4} Dongchen Ji,^{1,4} Xue Li,¹ Xue Sun,¹ Wenzheng Wang,¹ Tong Liu,^{1,2} and Yingpu Li^{1,2,3}

¹Department of Oncological Surgery, Harbin Medical University Cancer Hospital, Harbin, Heilongjiang Province 150000, China; ²NHC Key Laboratory of Cell Transplantation, The First Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang Province 150001, China; ³Genomics Research Center (Key Laboratory of Gut Microbiota and Pharmacogenomics of Heilongjiang Province), College of Pharmacy, Harbin Medical University, Harbin, Heilongjiang Province 150081, China

Breast cancer patients with estrogen receptor-negative (ERneg) status, encompassing triple negative breast cancer (TNBC) and human epidermal growth factor receptor 2 positive breast cancer, are confronted with a heightened risk of drug resistance, often leading to early recurrence; the biomarkers and biological processes associated with recurrence is still unclear. In this study, we analyzed bulk RNA sequencing (RNA-seq) data from 285 cancer and paracancerous samples from 155 TNBC patients, along with transcriptome data from 11 independent public cohorts comprising 7,449 breast cancer patients and 26 single-cell RNA-seq datasets. Our results revealed differential enrichment of nerve-related pathways between TNBC patients with and without 10-year recurrence-free survival. We developed an early recurrence index (ERI) using a machine learning model and constructed a nomogram that accurately predicts the 10-year survival of ERneg patients (area under the curve [AUC]_{Training} = 0.79; AUC_{Test} = 0.796). Further analysis linked ERI to enhanced neural function and immunosuppression. Additionally, we identified EN1, the most significant ERI gene, as a potential biomarker that may regulate the tumor microenvironment and sensitize patients to immunotherapy.

INTRODUCTION

The incidence of new cases and deaths related to breast cancer continues to increase, contributing to increased mortality and mobility of the disease worldwide.^{1,2} Despite significant improvements in outcomes for breast cancer patients through various treatment regimens, including chemotherapy, targeted therapy, and immunotherapy, triple negative breast cancer (TNBC) patients still face challenges. Endocrine therapy has proven effective in significantly increasing survival rates for patients with Luminal A (LumA) or Luminal B (LumB) breast cancer, which are characterized by high expression of the estrogen receptor (ER) and are responsive to endocrine therapy. For patients with breast cancer tissues that do not express ER and progesterone receptor but exhibit high expression levels of human epidermal growth factor receptor (Her2/ErbB2), Her2-targeted therapy can be beneficial.³ However, compared with these molecular subtypes of breast cancer, TNBC does not express the aforementioned targets

and exhibits greater heterogeneity and complexity in the tumor microenvironment (TME). Consequently, there are no certified targets for the treatment of TNBC, highlighting the need for the discovery of novel therapeutic targets.^{4,5}

The majority of recurrence and death events are attributed to the acquisition of drug resistance. Currently, the treatment of TNBC primarily depends on chemotherapy and immunotherapy. While neoadjuvant immunotherapy has demonstrated significant improvements in outcomes for early-stage TNBC patients,⁶ the combination of immunotherapy and chemotherapy has not been successful in enhancing outcomes for advanced TNBC patients.⁷ Therefore, there is a pressing need to identify patients who are prone to drug resistance and early recurrence in order to facilitate precision treatment. With the rapid advancement of next-generation sequencing, several prognostic models based on transcriptome analysis have been developed and have shown promising accuracy in predicting outcomes for TNBC patients. However, a prognostic model based on genes associated with T cells,⁸ immune signature,⁹ and cell-death¹⁰ did not show an optimal efficacy in predicting long-term survival of TNBC patients, and there remains a lack of a model that can specifically recognize the TME associated with early recurrence and long-term survival.

To explore the underlying causes of recurrence and its impact on long-term survival among patients with TNBC, we collected cancer and adjacent non-cancerous tissues from 155 TNBC patients and subjected them to RNA sequencing (RNA-seq) analysis. Through

Received 11 November 2024; accepted 5 March 2025;
<https://doi.org/10.1016/j.omton.2025.200965>.

⁴These authors contributed equally

Correspondence: Yingpu Li, Department of Oncological Surgery, Harbin Medical University Cancer Hospital, Harbin, Heilongjiang Province 150000, China.

E-mail: liyongpu@hrbmu.edu.cn

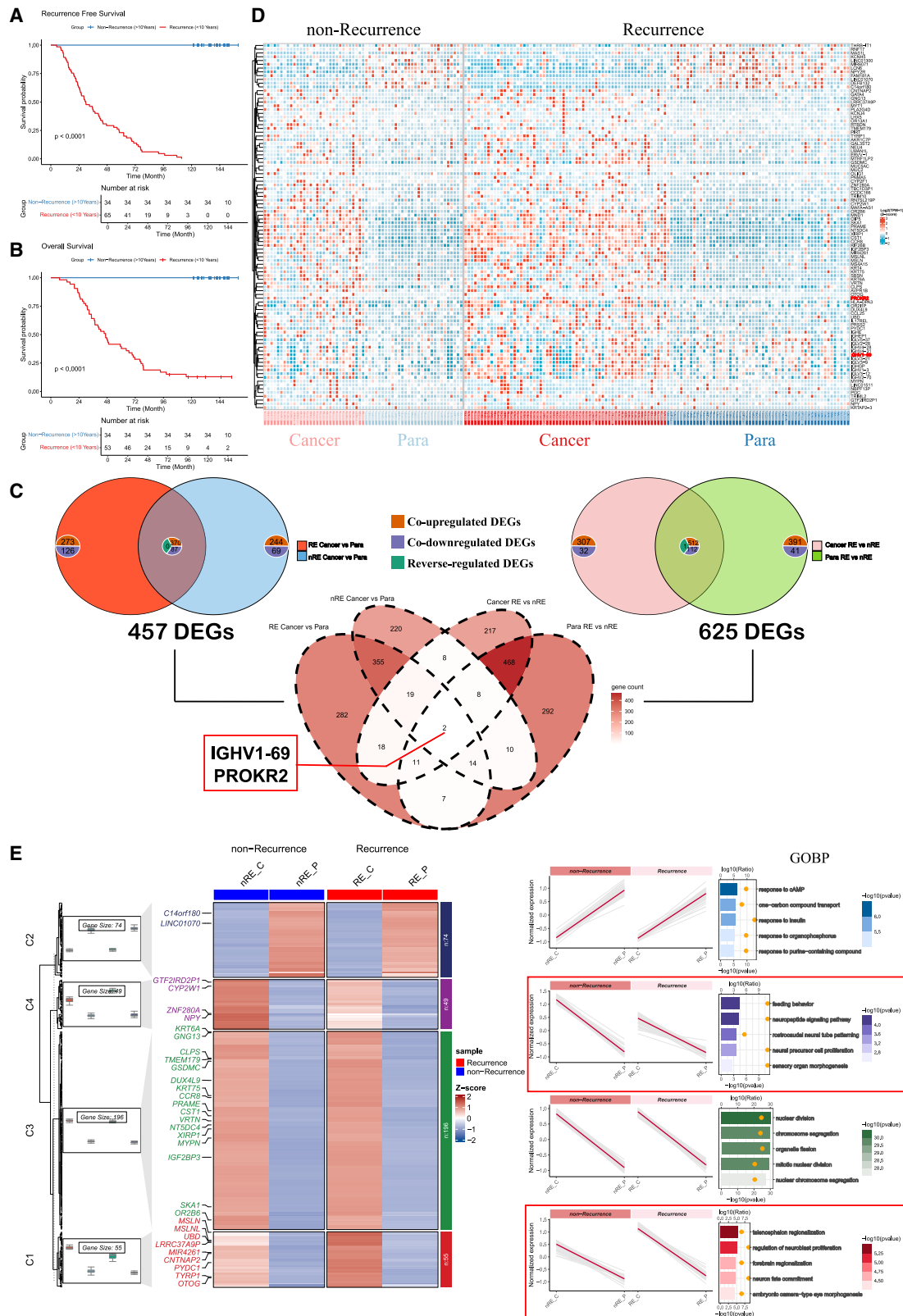
Correspondence: Tong Liu, Department of Oncological Surgery, Harbin Medical University Cancer Hospital, Harbin, Heilongjiang Province 150000, China.

E-mail: liutong@hrbmu.edu.cn

Correspondence: Wenzheng Wang, Department of Oncological Surgery, Harbin Medical University Cancer Hospital, Harbin, Heilongjiang Province 150000, China.

E-mail: wangwenzheng@hrbmu.edu.cn





(legend on next page)

this analysis, we identified gene expression patterns and core genes associated with recurrence. Leveraging these genes, we developed a prognostic model using a rigorous selection process involving 10 machine learning algorithms. The model was trained and validated on a comprehensive dataset that combined our own data with 11 additional individual cohorts. Our model demonstrated high accuracy in predicting 10-year recurrence-free survival (RFS) and overall survival (OS) for TNBC and Her2 patients, who are characterized by being ER independent. Further analysis revealed that the genes incorporated into the model are correlated with immune response and nerve-associated functions, and identified several potential biomarkers, such as Engrailed-1 (EN1), for future investigation. Thus, our study provides an effective prognostic tool for predicting long-term survival in ER-negative (ERneg) breast cancer and uncovers differential biological processes among patients with early recurrence or long-term survival.

RESULTS

Clinical characteristics and transcriptome profiling in TNBC patients with early detection of relapse or long-term relapse-free status

To identify crucial biomarkers associated with long-term RFS, we collected data from 155 TNBC patients with extensive follow-up information. Among these patients, 65 experienced local relapse or distant metastasis within 10 years after surgery and adjuvant therapy, classifying them as the recurrence group (RE). Conversely, 34 patients remained alive without experiencing relapse events, classifying them as the non-RE. The remaining 56 patients, who did not meet the aforementioned clustering condition, were not chosen for differential analysis but rather for survival analysis. Analysis of clinical information revealed no significant differences in age, T stage, N stage, P53 and KI-67 staining, or receipt of chemotherapy between RE and non-RE patients (Table S1). Survival analysis demonstrated that RE patients exhibited significantly poorer RFS and OS compared with non-RE patients (Figure 1A). Therefore, identifying TNBC patients with early relapse risk is crucial for ensuring long-term survival.

Next, we conducted transcriptomic analysis of cancer and paired paracancerous tissues from the aforementioned patients. Principal component analysis (PCA) of all tissues indicated distinct transcriptome features between cancer and paracancerous tissues (Figure S1A). Differential expression gene (DEG) analysis was performed on cancer and paracancerous tissues between RE and non-RE patients, yielding 964 and 1,057 DEGs, respectively (Figure S1B). Additionally, we identified 856 and 770 DEGs when comparing paired cancer and paracancerous tissues separately in RE or non-RE patients (Figure S1B). The intersection of these DEGs revealed several common genes, whose

expression levels were demonstrated in both RE and non-RE groups (Figure 1C). Notably, immunoglobulin IGHV1-69 and prokineticin 2 receptor (PROKR2), a sensory neuron protein, were identified as common DEGs across all comparisons (Figure 1D). This finding is consistent with the results of Gene Set Enrichment Analysis (GSEA) analysis, which showed differential enrichment of hormone- and neuron-associated pathways in cancer and paired paracancerous tissues between RE and non-RE patients (Figure S1C). Therefore, the correlation among B cells, neuron function, and hormone status may play a pivotal role in regulating TNBC progression.

Further Mfuzz analysis, based on the DEGs, uncovered four expression patterns: C1 (DEGs overexpressed in paracancerous tissues from the RE group), C2 (DEGs overexpressed in paracancerous tissues from both groups), C3 (DEGs overexpressed in cancer tissues from both groups), and C4 (DEGs overexpressed in cancer tissues from the non-RE group). These patterns showed potential correlation with early recurrence events (Figure 1E). Intriguingly, Gene Ontology Biological Process (GOBP) analysis indicated that DEGs from C1 and C4, which are potentially correlated with early recurrence, are enriched in pathways associated with neuron proliferation, differentiation, and signaling transduction. Based on the transcriptome analysis results, we hypothesize that neural function may play a role in the recurrence of TNBC.

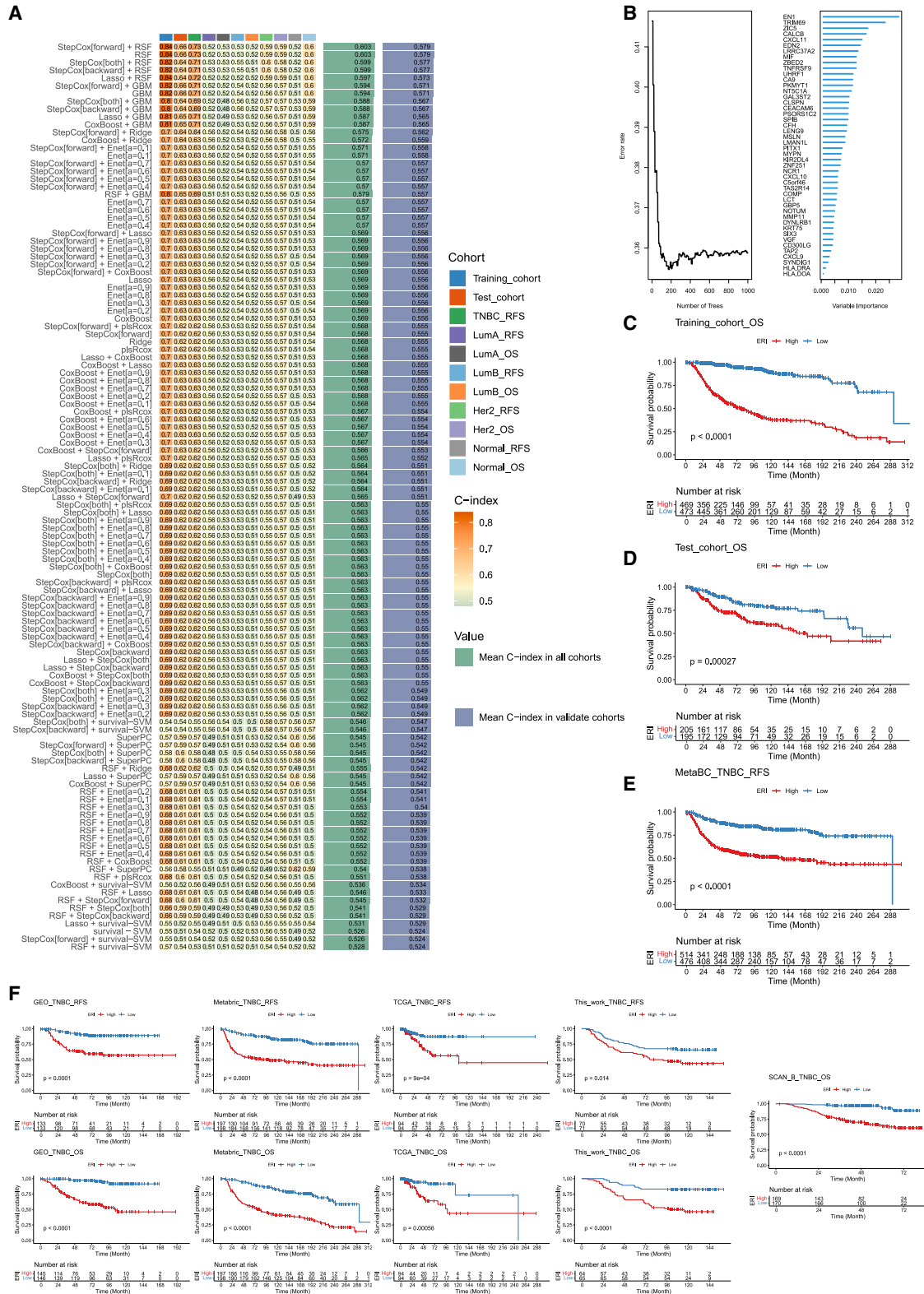
The construction of a prognostic model based on the recurrence associated DEGs

To delve deeper into the potential of abovementioned DEGs (Figure 1D) in predicting early recurrence and survival, we amassed transcriptome data from 11 independent public databases, including GSE12276, GSE16446, GSE19615, GSE20685, GSE21653, GSE58644, GSE58812, GSE88770, Metabric, The Cancer Genome Atlas (TCGA)-BRCA, and GSE96058. Among them, GSE12276, GSE16446, GSE19615, GSE20685, GSE21653, GSE58644, GSE58812, and GSE88770 were selected to construct the Gene Expression Omnibus (GEO) cohort, which was deemed as an independent database for further analysis (Figures S2A and S2B). After rigorous quality control measures and data integration, five pivotal databases—GEO, Metabric, TCGA, SCAN_B, and our own work (this work)—were utilized to generate the Meta_BC dataset, encompassing 7,231 breast cancer patients for subsequent analysis (Figures S2C and S2D). We then filtered out patients with TNBC to establish a training cohort ($n = 942$) and a test cohort ($n = 400$) from the Meta_BC dataset.

Subsequently, we employed a machine learning-based integrated model, incorporating 10 distinct algorithms, to construct a prognostic model.¹¹ The C-index of various combined algorithms in the training

Figure 1. Kaplan-Meier survival analysis and differential gene expression in TNBC patients

(A) Kaplan-Meier curve plotting RFS status of patients in RE and non-RE patients. (B) Kaplan-Meier curve plotting of OS status of patients in RE and non-RE patients. (C) Venn diagrams showing the intersection of DEGs between (1) cancer and paired paracancerous tissues in RE or non-RE group (top left); (2) RE and non-RE groups in cancer or paracancerous tissues (top right); and (3) abovementioned DEGs. (D) Heatmap illustrating intersected DEGs identified from (C) in RE and non-RE group, IGHV1-69 and PROKR2 are labeled with red highlight and in bold text. (E) GseaVis visualizing the significant discrete clusters of DEGs measured by Mfuzz and the GOBP pathway enrichment of each cluster in RE and non-RE groups.



(legend on next page)

cohort was illustrated in Figure 2A, with Cox regression followed by random forest emerging as the most effective model for predicting OS in TNBC patients. The risk score, termed the early recurrence index (ERI) in this study, was computed based on 46 significant genes derived from the prognostic model. The ERI was found to be significantly associated with the outcomes of TNBC patients (Figures 2B, S2E, and S2F). Furthermore, stratifying patients using the median ERI revealed a significant correlation between ERI^{High} and poor OS and RFS in TNBC patients compared with those with low ERI^{Low} (Figures 2C–2E). Comparable results were observed across diverse databases within the Meta-BC dataset (Figure 2F). Notably, all ERI^{High} patients from Meta-BC exhibited a poor survival outcome, suggesting a potential association of ERI with outcomes in other molecular subtypes (Figure 3A).

To further explore this association, we investigated the relationship between ERI and RFS or OS in other breast cancer subtypes, including LumA, LumB, Her2, and normal. Interestingly, no significant difference was observed between ERI^{High} and ERI^{Low} patients with the aforementioned subtypes, except for Her2 (Figures 2A and S2G). ERI^{High} Her2 patients also demonstrated a poorer outcome compared with ERI^{Low} Her2 patients in Meta-BC, and this trend was evident in individual databases (Figures S2G and S2H). Given that TNBC and Her2 patients are both characterized by ERneg status, the genes in our model may potentially link the oncogenesis of these two distinct breast cancer subtypes.

ERI^{High} ERneg patients face a greater risk of recurrence when undergoing chemotherapy

Chemotherapy serves as a foundational therapeutic approach to enhance breast cancer patient outcomes, yet the correlation between the ERI score and chemotherapy efficacy remains uncertain. To explore this relationship, we examined TNBC and Her2 patients, both those who received chemotherapy and those who did not, and assessed the link between ERI and patient outcomes. Notably, in TNBC patients, a high ERI was associated with significantly poorer outcomes, irrespective of their chemotherapy history (Figures 3C and 3D). For Her2 patients, those with high ERI who underwent chemotherapy exhibited shorter RFS and OS compared with their ERI^{Low} counterparts (Figure 3E). However, no such association was observed between ERI and outcomes in Her2 patients who did not receive chemotherapy (Figure 3F). Upon extending our investigation to other molecular subtypes of breast cancer, we observed similar trends (Figures S3A–S3F), indicating that ERI is notably associated with outcomes in breast cancer patients with ERneg status. Consequently, we combined TNBC and Her2-positive patients into an ERneg cohort and found that ERI^{High} patients, regardless of chemo-

therapy administration, showed significantly poorer outcomes (Figures 3G and 3H). Strikingly, ERI^{High} patients who received chemotherapy had a markedly shorter RFS compared with those who did not (Figure 3I). This pattern was not evident in ERI^{Low} patients, suggesting a significant role of ERI genes in the response to chemotherapy (Figure 3J).

An ERI-based nomogram effectively predicts long-term survival of ERneg breast cancer patients

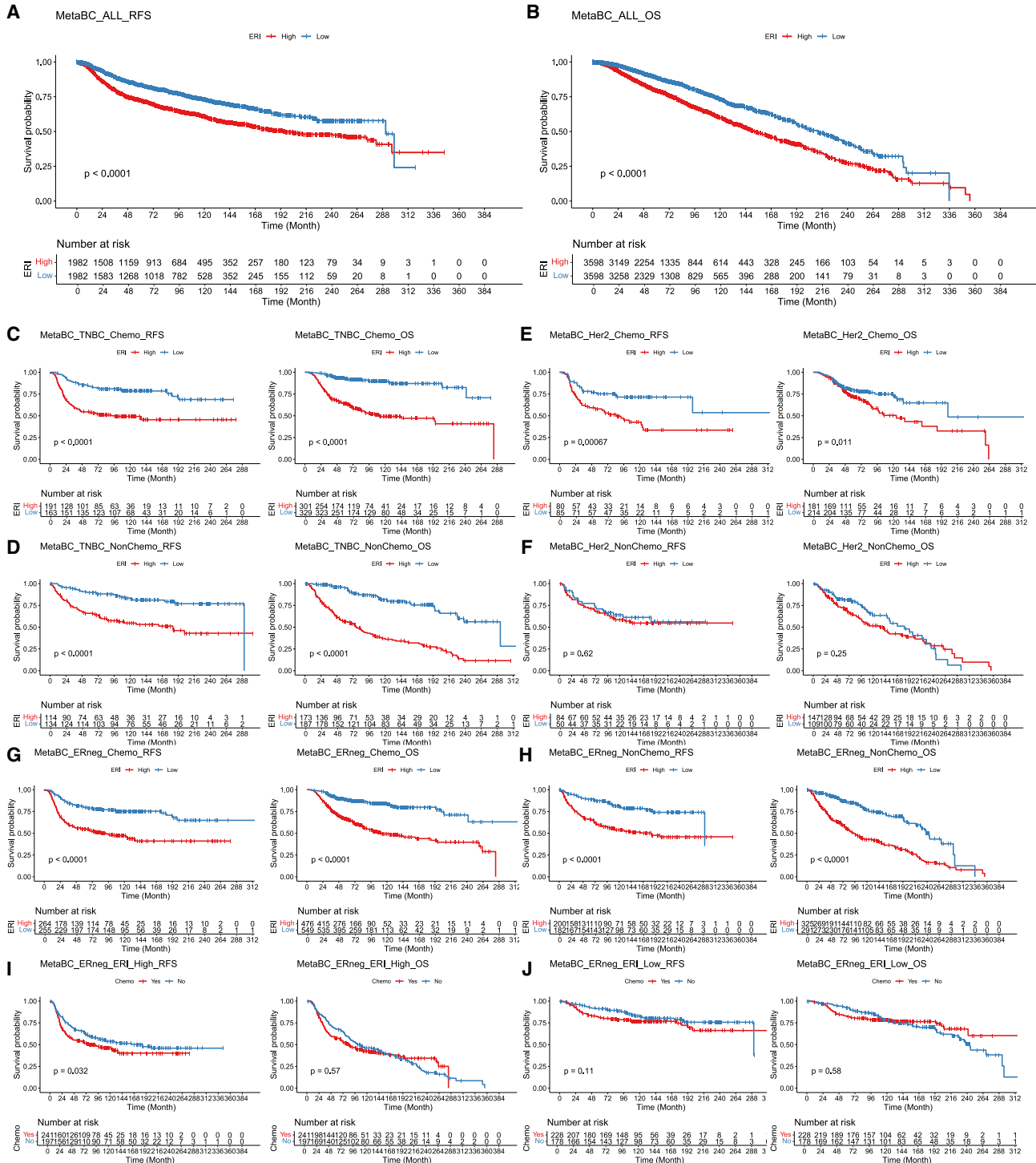
Previously, there was a scarcity of transcriptome-based models for predicting long-term survival in breast cancer patients. To bridge this gap, we conducted univariable and multivariable Cox regression analyses in a training cohort of ERneg patients. Our aim was to identify survival-associated factors among ERI and several clinical variables, including age, T stage, N stage, and tumor grade. The univariable Cox regression analysis revealed that ERI, age, T stage, and N stage were all associated with survival (Figure 4A). Subsequent multivariable Cox regression analysis further confirmed that ERI, age, and N stage were significant factors (Figure 4B). We also observed that the ERI score was higher in older patients (age ≥ 60 years), those with high-grade tumors (grade 3), advanced T stage (T3), and lymph node metastasis, particularly in TNBC patients (Figure S4).

Using these findings, we constructed a nomogram to predict the OS of patients based on ERI, age, and N stage (Figure 4C). The C-index value of the nomogram was 0.756 (95% confidence interval [CI], 0.732–0.782) in the training cohort and 0.746 (95% CI, 0.710–0.783) in the test cohort. The calibration curves demonstrated the accuracy of the nomogram in predicting 3-, 5-, and 10-year survival rates (Figure 4D). The decision curve analysis (DCA) results indicated that the nomogram model outperformed all other predictors used in this study, particularly in predicting 10-year survival (Figure 4E).

Based on the nomogram scores, we found significant differences in OS between patients with high scores (score^{High}) and those with low scores (score^{Low}) in both the training (Figure 4F) and test cohorts (Figure 4G). These differences were also observed in the ERneg cohort, regardless of whether patients received chemotherapy (Figure 4H) or not (Figure 4I). The area under the curve (AUC) values of the nomogram model in these cohorts were evaluated, and the values for predicting 10-year OS were greater than 0.79 (Figures 4F–4I). Furthermore, we assessed the AUC values in five independent cohorts, and the results showed that the nomogram had high accuracy in predicting 3-, 5-, and 10-year OS (Figure S5A) and RFS (Figure S5B) in ERneg patients. However, although there were significant differences between ERI^{High} and ERI^{Low} patients in the ER-positive (ERpos) cohort, the AUC values indicated that the nomogram model

Figure 2. Machine learning-based prognostic model for early recurrence prediction in TNBC

(A) Heatmap illustrating clustering deviation index results of 10 machine-learning algorithms in various cohorts from Meta_BC. (B) The number of trees of StepCox[forward] + RFS model (left) and the variable importance of model genes (right). (C) Kaplan-Meier curve plotting the OS status of TNBC patients with ERI^{High} and ERI^{Low} in the training cohort. (D) Kaplan-Meier curve plotting OS status of TNBC patients with ERI^{High} and ERI^{Low} in the test cohort. (E) Kaplan-Meier curve plotting RFS status of TNBC patients with ERI^{High} and ERI^{Low} in the Meta_BC cohort. (F) Kaplan-Meier curve plotting RFS and OS status of TNBC patients with ERI^{High} and ERI^{Low} in the GEO, Metabric, TCGA_TNBC, This_work, and SCAN_B cohorts.



(legend continued on next page)

could not accurately predict the RFS and OS of ERpos breast cancer patients (Figure S5C). Therefore, we have established an ERneg breast cancer-specific prognostic model that accurately predicts long-term survival.

ER^{High} patients presents enhanced nerve-regulated function and decreased immune response

In this study, we have shown that ERI serves as a crucial determinant for precisely forecasting the long-term survival of patients with ERneg breast cancer, despite the underlying biological mechanism of ERI remaining elusive. To elucidate this, we conducted a protein-protein interaction (PPI) analysis among genes implicated in ERI and discovered multiple interacting proteins (Figure 5A). The functions of these proteins are linked to immune responses, extracellular matrix formation, and neural tube development. Subsequently, we examined the DEGs (Figure S6A) between ER^{High} and ER^{Low} patients across four cohorts: GEO, Metabric, TCGA, and our own dataset. A Venn diagram identified 136 and 576 common DEGs in ER^{High} and ER^{Low} patients, respectively, within these four cohorts (Figure 5B). Strikingly, GOBP pathway analysis revealed that the common DEGs expressed in the ER^{High} group were enriched in several nerve-related pathways, including positive regulation of axonogenesis, tube formation, and *in utero* embryonic development (Figure 5C). This finding aligns with our earlier observation that nerve-associated pathways exhibited differential enrichment between RE and non-RE patients (Figure 1E).

Furthermore, we performed GSEA analysis of GOBP enrichment and identified 411 common differential enriched pathways (DEPs) in the ER^{Low} group (Figure 5D). These pathways were associated with the activation of immune responses, such as cell killing (Figure S6B). Notably, only the skin development pathway was identified as a DEP in the ER^{High} group (Figure S6C). Upon examining the results of the GSEA analysis in each cohort, we found that multiple nerve-associated pathways were significantly enriched in the ER^{High} group (Figure 5E). To validate the significance of these nerve-associated pathways, we applied Gene Set Variation Analysis (GSVA) based on the C2 collection, which encompasses 7,411 gene sets. Our findings indicated that presynaptic depolarization and calcium channel opening, as well as the S1P pathway, were separately enriched in ER^{High} TNBC and Her2 patients, whereas immune response pathways were enriched in the ER^{Low} group (Figures S6D and S6E). Additionally, several nerve-associated pathways were differentially enriched between ER^{High} and ER^{Low} patients with ERneg, TNBC, and Her-2 breast cancers (Figure S6F). These results underscore that tumor tissues from ER^{High} patients, who are at a higher risk of early recurrence and mortality, exhibit enhanced nerve-regulated function and reduced immune responses.

ERI and core genes are correlated with immune suppression and neuron infiltration

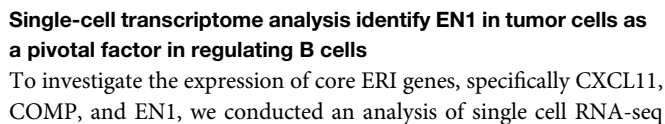
Several studies have reported that the enhancement of the epithelial-mesenchymal transition (EMT) process and cancer stemness can lead to recurrence and metastasis in cancer patients. However, no significant difference was observed in the EMT and stemness scores between patients with ER^{Low} and ER^{High}, who had ERneg, TNBC, or Her2 tumors (Figure S7A). Consequently, we shifted our focus to examining the correlation between the TME and ERI. To delineate the components of the TME in ERneg patients, we used seven deconvolution algorithms, which revealed an increased infiltration of immune cells associated with higher ERI scores (Figure S7B). Cibersort analysis further demonstrated that ER^{Low} patients exhibited higher levels of infiltration by anti-tumor immune cells, such as CD8⁺ T cells, activated CD4⁺ T cells, and M1 macrophages, whereas T regulatory cells and neutrophils showed increased infiltration in ER^{High} patients (Figure S7C).

A correlation analysis was conducted between ERI and infiltration scores calculated by various deconvolution algorithms, and significant correlations were identified across five cohorts (GEO, Metabric, TCGA, SCAN_B, and This_work) (Figure 5F). Notably, anti-tumor immune cells, including T cells, B cells, natural killer (NK) cells, and dendritic cells, showed a significant negative correlation with ERI (Figure 5F). Intriguingly, we also observed a positive correlation between ERI and the infiltration score of neurons, as calculated by Xcell (Figure S7D). The Mantel test analysis further suggested that the infiltration of neurons was negatively correlated with the aforementioned immune cells (Figures 5G, 5H, and S7E).

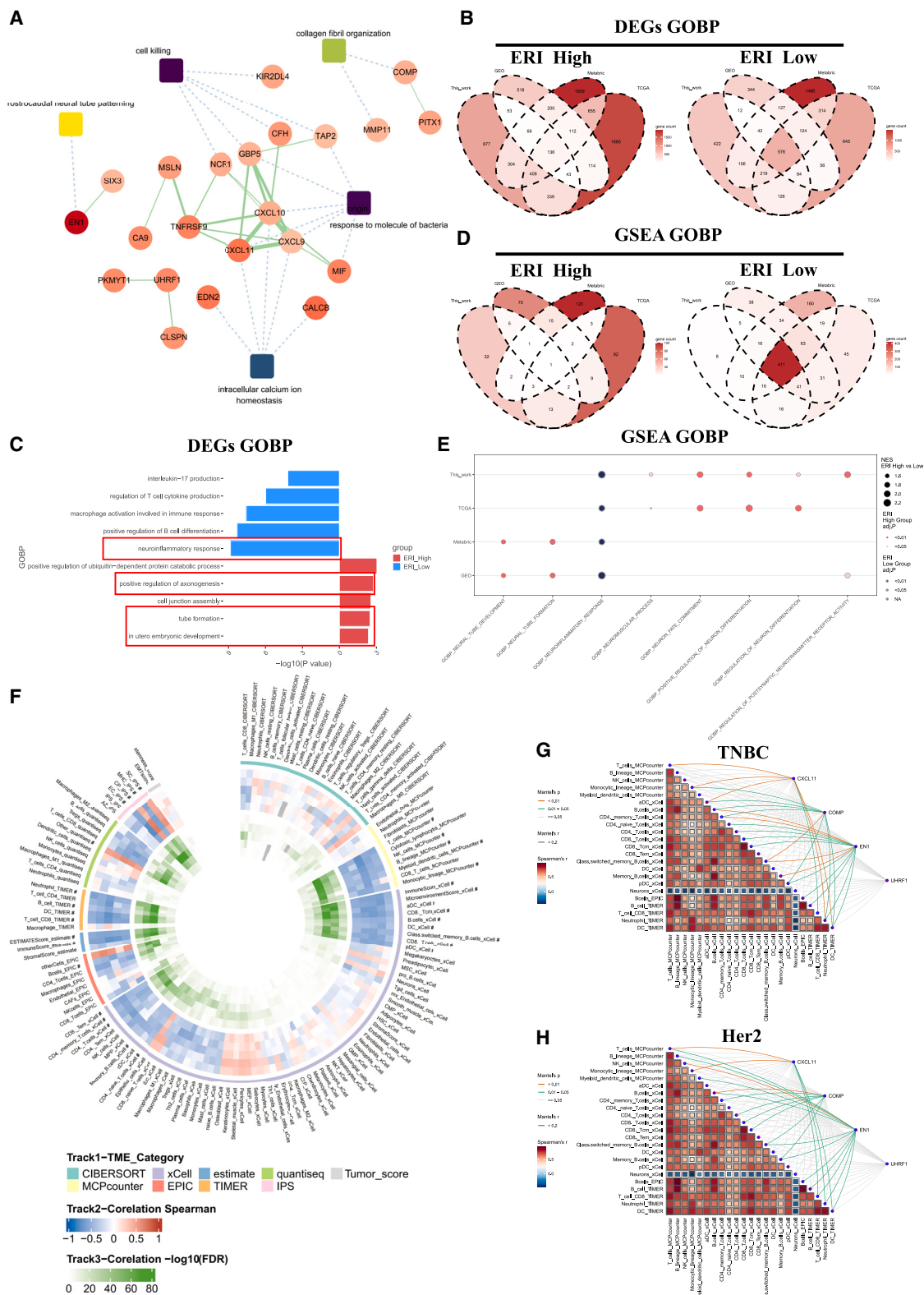
Furthermore, our results indicated that the core genes of ERI (Figure 5A), including C-X-C motif chemokine ligand 11 (CXCL11), cartilage oligomeric matrix protein (COMP), and EN1, had a significant impact on the infiltration of these immune cells (Figures 5G and 5H). EN1 was found to be correlated with poor outcomes in TNBC (Figure S8A) and Her2 patients (Figure S8B). Additionally, EN1 showed a positive correlation with the infiltration of T helper type 1 cells, T helper type 2 cells, and M1 macrophages, while it was negatively correlated with the infiltration of CD4⁺ T memory cells, mast cells, and M2 macrophages (Figure S8C). According to the Mantel test analysis, EN1 was significantly correlated with the infiltration of several immune cells (Figure 5H).

In conclusion, our bulk transcriptome analysis highlights the potential clinical application of our model in predicting the long-term survival of ERneg patients and suggests a potential regulatory role of ERI genes on immune cells and neurons.

plotting RFS and OS status of chemotherapy-treated Her2 patients with ER^{High} and ER^{Low} in the Meta_BC cohort. (F) Kaplan-Meier curve plotting RFS and OS status of non-chemotherapy-treated Her2 patients with ER^{High} and ER^{Low} in the Meta_BC cohort. (G) Kaplan-Meier curve plotting RFS and OS status of chemotherapy-treated ERneg patients with ER^{High} and ER^{Low} in the Meta_BC cohort. (H) Kaplan-Meier curve plotting RFS and OS status of non-chemotherapy-treated ERneg patients with ER^{High} and ER^{Low} in the Meta_BC cohort. (I) Kaplan-Meier curve plotting RFS and OS status of ER^{High} ERneg patients treated with or without chemotherapy in the Meta_BC cohort. (J) Kaplan-Meier curve plotting RFS and OS status of ER^{Low} ERneg patients treated with or without chemotherapy in the Meta_BC cohort.



(scRNA-seq) data derived from ERneg breast cancer patients sourced from the GSE176078 and GSE161529 databases. After rigorous quality control and data integration, a total of 137,032 cells were retained for subsequent analysis, as depicted in [Figure S9A](#). Our analysis



(legend on next page)

revealed 38 distinct cell clusters within the TNBC and Her2 patient samples, which were further annotated into seven cell types: epithelial cells, T/NK cells, B cells, plasma cells, myeloid cells, endothelial cells, and fibroblasts, based on CellMarker data (Figure S9B). InferCNV analysis confirmed that the epithelial cells in our dataset were indeed tumor cells (Figure S9C). Upon examining the gene expression profiles, we found that CXCL11 and COMP were primarily expressed in myeloid cells and fibroblasts, while EN1 was predominantly expressed in tumor cells (Figure 6C). Given our previous finding that EN1 is associated with the infiltration of immune cells, particularly B cells (Figures 5G and 5H), we delved deeper into the relationship between EN1-expressing tumor cells and B cells.

We classified the tumor cells into EN1^{High} and EN1^{Low} groups based on the expression level of EN1 (Figure 6D). Subsequently, we extracted tumor cells, B cells, and plasma cells from the dataset and analyzed the ligand-receptor interactions among these cell types. Our analysis revealed multiple interactions between tumor cells, B cells, and various plasma cells expressing IGKC, IGKV1-9, IGLV1-40, or MZB1 (Figure 6E). Notably, B cells exhibited a bidirectional interaction with EN1^{Low} tumor cells, but only received signals from EN1^{High} tumor cells (Figure S9D). Among these interactions, the MIF (CD74-CXCR4) signaling pathway emerged as the most significant communication pattern (Figure 6F). Importantly, MIF is a survival-related gene that is included in the ERI score (Figure 2B). Further analysis revealed that tumor cells, particularly EN1^{High} tumor cells, were the primary sources of MIF signaling to B cells (Figures 6G and S9E) and that the MIF (CD74-CXCR4) pathway was significantly more prominent in interactions between EN1^{High} tumor cells and B cells compared with other signaling pathways (Figure 6H).

To explore the potential role of EN1 in the immune microenvironment and its response to immunotherapy, we utilized the Kaplan-Meier Plotter website. Our analysis revealed that in non-epithelial-derived tumors such as melanoma, EN1 is associated with a poor response to immunotherapy (Figure S9F). However, in urothelial cancer, high EN1 expression in males was associated with poor outcomes, while high EN1 expression in females was associated with an optimal response to immunotherapy (Figure 6I). These findings suggest a potential role for EN1 in immunotherapy response.

DISCUSSION

Breast cancer subtypes such as TNBC and HER2 breast cancer continue to pose challenges due to drug resistance, early recurrence, and distant metastasis. Identifying these patients is crucial for implementing precision medicine. In line with previous studies that did not

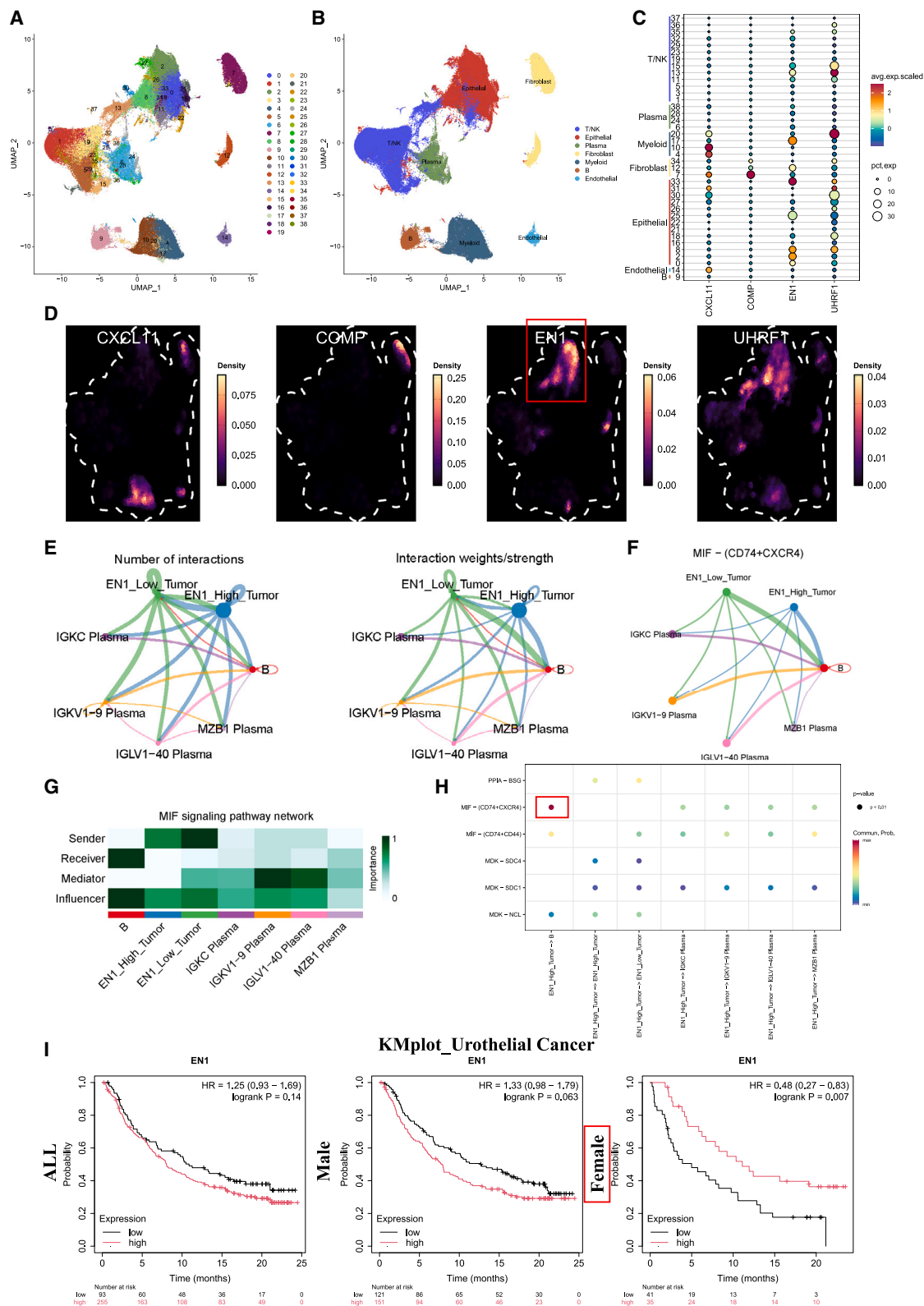
focus on specific gene clusters with similar functions,¹² we identified genes related to recurrence by comparing DEGs in TNBC patients with short- or long-term RFS (Figure 1C). This approach ensured that the selected candidate genes could enhance the effectiveness of the subsequent prognostic model. Strikingly, although our model was trained on TNBC patients, it also demonstrated efficacy in predicting outcomes for HER2 patients, another molecular subtype characterized by ERneg (Figures 3 and 4). To our knowledge, no previous work has generated a prognostic model for ERneg breast cancer patients using transcriptome data. Notably, our prognostic model exhibits optimal accuracy in predicting long-term survival for both TNBC and HER2-positive patients (Figures 3 and 4). Furthermore, our model's accuracy in predicting 5- or 10-year survival outperforms models using genes associated with T cells,⁸ immune signature,⁹ and cell death¹⁰ in TNBC. Intriguingly, DEG and Mfuzz analysis of our data suggested that nerve-associated pathways may be linked to TNBC recurrence, identifying the sensitive neuron receptor PROKR2 as a core DEG.¹³ Although further analysis did not include PROKR2 as a prognostic gene in our model, we identified another nerve-related gene: EN1.

EN1 has been reported to play crucial roles in various biological development processes. For instance, EN1 is essential for the development and survival of dopaminergic neurons.^{14,15} The absence of the EN1 gene leads to abnormal bone formation and Mendelian diseases.^{16,17} In contrast with these findings, the roles of EN1 in tumorigenesis and progression are gaining attention. A pan-cancer epigenetic and transcriptomic atlas has revealed that EN1 is one of the tumor-specific epigenetic drivers associated with cancer transitions.¹⁸ Findings from other studies also support the notion that EN1 promotes lung metastasis of salivary adenoid cystic carcinoma through the PI3K-AKT pathway¹⁹ and enhances pancreatic cancer metastasis via the MAPK pathway.²⁰ More evidence of EN1's role in promoting cancer metastasis has been reported in breast cancer, particularly in TNBC. Reduced levels of bromodomain-containing protein 4, along with EN1, co-regulate the proliferation and migration of TNBC.²¹ Moreover, brain metastasis of TNBC depends on the transcriptional function of EN1.²²

These results align with our finding that EN1 is the most important gene included in our prognostic model (Figure 2B). We also found that EN1 is significantly correlated with the infiltration of various immune cells, especially B cells (Figure 5G). Nevertheless, a higher ERI score inversely correlates with anti-tumor immune infiltration (Figure 5F), which makes the contradicting observation with EN1. The ERI is computed based on the scoring of multiple genes that are linked

Figure 5. ERI-associated molecular features and tumor microenvironment alterations

(A) Cytoscape illustrating the relative importance score, PPI network, and enriched GOBP pathway of DEGs obtained from the ERI model. (B) Venn diagrams showing the overlaps of GOBP analysis on DEGs expressed in EN1^{High} or EN1^{Low} ERneg patients. (C) Boxplot illustrating results of GOBP in EN1^{High} or EN1^{Low} ERneg patients. (D) Venn diagrams showing the overlaps of GOBP analysis on GSEA in EN1^{High} or EN1^{Low} ERneg patients. (E) Dotplot showing GSEA analysis of nerve-associated pathways in ERneg patients with EN1^{High} or EN1^{Low}. (F) Heatmap illustrating correlation analysis of ERI with various TME score in individual cohort. #Significant correlation, Icor value of greater than 0.3 and a false discovery rate of less than 0.05. *p* values are adjusted by Benjamini-Hochberg method. (G) Mantel test results of core ERI genes with significant correlated TME score in TNBC patients. (H) Mantel test results of core ERI genes with significant correlated TME score in Her2 patients.



(legend on next page)

to patient survival. Consequently, the relationship between ERI and the TME can be viewed as a complex, overarching association, where the specific correlation of each gene may vary. Clearly, targeting every gene contributing to ERI for the treatment of ERneg breast cancer patients is impractical. Hence, the contrasting result prompted us to delve deeper into whether EN1 could serve as a potential biomarker for immunotherapy in ERneg breast cancer. As illustrated in Figure 6I, EN1 displayed an opposing association with the response to immunotherapy between male and female cancer patients. Considering that EN1 is correlated with the poor outcome of ERneg patients, we hypothesized that EN1 is a biomarker for the clinical application of immunotherapy in these patients (Figures S8A and S8B).

Genomic analysis suggests that EN1 is a basal-like specific oncogene.^{23,24} However, our data indicate that EN1 may play a crucial role in regulating the TME in HER2-positive patients (Figures 5A and 5H). Thus, EN1 demonstrates potential for clinical application in both TNBC and HER2-positive patients, although its role in HER2 breast cancer has been overlooked. While some studies have focused on attenuating TNBC by using N-terminal cell-penetrating peptide/nuclear localization sequence targeting EN1,²⁵ designing an EN1-targeted therapy remains challenging due to its role as a transcriptional factor. Perhaps exploring the different roles of EN1 in immunotherapy could facilitate further clinical applications (Figure 6I).

MATERIALS AND METHODS

Patient information and sample collection

A total of 285 samples, comprising 147 cancer tissues and 138 adjacent non-cancerous tissues, were collected from 155 female patients who had been histologically confirmed to have TNBC at the Harbin Medical University Cancer Hospital from 2011 to 2015. These patients had previously undergone surgery and received adjuvant therapy as their first-line treatment. Patients who experienced a recurrence event within 10 years were classified as the RE group, while those with more than 10 years of RFS were classified as the non-RE group. The tissue samples were obtained during the surgical procedure and stored at -80°C until further analysis. This study was approved by the Ethical Committee of Harbin Medical University Cancer Hospital and adhered to all relevant ethical regulations. All patients provided written informed consent for the collection of their tissue samples for research and analysis.

Library construction, quality control, and sequencing of transcriptome analysis

Total RNA was used as the starting material for RNA sample preparations. Sequencing libraries were generated using the NEBNext Ultra

RNA Library Prep Kit for Illumina (NEB) according to the manufacturer's instructions. Briefly, mRNA was purified using poly-T oligo-attached magnetic beads, followed by fragmentation and synthesis of first and second strand cDNAs. The cDNA fragments were then adenylated, ligated to NEB Next Adaptors, and purified to select fragments of 370–420 bp. After treatment with USER Enzyme, PCR was performed using Phusion High-Fidelity DNA polymerase and Universal PCR primers. The PCR products were purified and the library quality was assessed on the Agilent 5400 system and quantified by QPCR. The qualified libraries were pooled and sequenced on Illumina platforms with the PE150 strategy at Novogene Bioinformatics Technology Co., Ltd, based on the effective library concentration and the required data amount.

RNA-seq raw data quality control

The original fluorescence image files obtained from Illumina platform are transformed to short reads (raw data) by base calling and these short reads are recorded in FASTQ format, which contains sequence information and corresponding sequencing quality information.²⁶ Sequence artifacts, including reads containing adapter contamination, low-quality nucleotides and unrecognizable nucleotide, undoubtedly set the barrier for the subsequent reliable bioinformatics analysis. Hence, quality control is an essential step and applied to guarantee the meaningful downstream analysis. We used Fastp (version 0.19.7)²⁷ to perform basic statistics on the quality of the raw reads. The steps of data processing were as follows. (1) Discard a paired reads if either one read contains adapter contamination. (2) Discard a paired reads if more than 10% of bases are uncertain in either one read. (3) Discard a paired reads if the proportion of low quality (Phred quality <5) bases is greater than 50% in either one read.

Public bulk transcriptome data collection and integration

In this study, microarray data were collected from multiple sources, including GSE12276,²⁸ GSE16446,²⁹ GSE19615,³⁰ GSE20685,³¹ GSE21653,³² GSE58644,³³ GSE58812,³⁴ and GSE88770,³⁵ to construct the GEO cohort. Additionally, we downloaded Metabric microarray data, RNA-seq count data from TCGA-BRCA, and SCAN_B.³⁶ Based on the five cohorts (GEO cohort, Metabric, TCGA-BRCA, and SCAN_B), we generated the Meta_BC cohort. R (version 4.3) was utilized for bioinformatic analysis of the transcriptome data. The combat function of the R package “sva” (version 3.50.0) was used to eliminate batch effects.³⁷ The R package “IOBR” (version 0.99.8) was used to perform transcript per million normalization and PCA.³⁸ Ultimately, the Meta_BC cohort comprised 7,604 breast cancer patients, including 3,203 LumA,

Figure 6. Single-cell transcriptome analysis identifies EN1 as a key regulator in ER-negative breast cancer

(A) Uniform Manifold Approximation and Projection (UMAP) plot showing 38 clusters of 137,032 cells from 26 ERneg patients. (B) UMAP plot showing 7 cell types from 26 ERneg patients. (C) Dotplot of CXCL11, COMP, EN1, and UHRF1 expression in each cluster and cell type. (D) Density plot of CXCL11, COMP, EN1, and UHRF1 expression in each cell. (E) Ligand-receptor interaction analysis among tumor cells, B cells, and plasma cells. (F) Ligand-receptor interaction analysis of MIF (CD74+CXCR4) pathway among tumor cells, B cells, and plasma cells. (G) Heatmap illustrating cell interaction of MIF signaling pathway network among tumor cells, B cells, and plasma cells. (H) Dotplot of different MIF signaling pathways among tumor cells, B cells, and plasma cells. (I) Kaplan-Meier curve plotting OS status of male or female urothelial cancer patients treated with immunotherapy.

1,614 LumB, 811 Her2, 1,468 TNBC, and 508 Normal-like (normal) patients. Both the Meta_BC cohort and the individual cohorts were used for further analysis.

DEG and pathway enrichment analysis

For the analysis of DEGs in RNA-seq data, we used the “DESeq2” package (version 1.42.1) in R.³⁹ For microarray data, we used the “limma” package (version 3.58.1) for DEG analysis.⁴⁰ Genes with an adjusted *p* value of less than 0.05 and an absolute log2 fold change of greater than 2.5 were considered as DEGs. After the identification of DEGs, we conducted a global pathway analysis of plasma proteins using the Molecular Signatures Database (“msigdb”, version 7.5.1), which includes GOBP, GO Molecular Function, GO Cellular Component, and Kyoto Encyclopedia of Genes and Genomes Pathways.⁴¹ Reactome pathway analysis was performed using the R package “ReactomePA” (version 1.46.0).⁴² Ranked GSEA and GSVA were conducted using the “clusterProfiler” package (version 4.12.3).⁴³

Pattern expression analysis of recurrence-related genes

The expression patterns of DEGs recognized in RE and non-RE patients was analyzed by “Mfuzz”.⁴⁴ Visualization of the results was carried out with “GseaVis” (github.com/junjunlab/GseaVis, version 0.1.0). The Mfuzz analysis was conducted using default parameters to classify the DEGs from RE and non-RE patients into distinct clusters. Subsequently, DEGs from each cluster were selected for GOBP pathway enrichment analysis. Furthermore, we calculated the membership score of each protein within the clusters.

Prognostic model construction and survival analysis

A total of 1,342 TNBC patients with outcome information were randomly divided into a training cohort (*n* = 942) and a test cohort (*n* = 400) at a ratio of 7:3. To select the most accurate prognostic model in the training cohort, we utilized the R package “Mime1” (version 0.0.0.9000), which includes 10 individual machine learning algorithms.¹¹ The risk score calculated by the model for predicting OS of each TNBC patient was termed as the ERI. Subsequently, risk scores for all patients in Meta_BC were computed, and patients were categorized into ERI^{High} and ERI^{Low} groups based on the mean value of ERI in each cohort. Kaplan-Meier plots were generated to visualize survival curves between different groups using the “survival” (version 3.7-0) and “survminer” (version 0.4.9) packages.

Next, TNBC and Her2 patients were combined to form the ERneg cohort, which was randomly divided into a training cohort (*n* = 1,302) and a test cohort (*n* = 550) at a ratio of 7:3. Univariate and multivariate Cox regression analyses were conducted to examine the relationship between ERI and various clinical factors, such as age, T stage, N stage, and tumor grade, using the “forestmodel” package (version 0.6.2). Based on the analysis results, independent factors predicting OS were identified. A nomogram predicting OS of ERneg patients and a plot of the calibration curve were generated using the “rms” package (version 6.7-0) based on these factors. DCA was calculated and plotted using the “ggDCA” package (version 1.2). The AUC

of the receiver operating characteristic curve and diagnostic accuracy were used to evaluate the performance of the nomogram. Survival analysis for immunotherapy was performed using the Kaplan-Meier Plotter website (kmplot.com).⁴⁵

PPI and TME analysis

The PPI analysis of ERI genes was conducted using the STRING dataset (version 12.0, accessible at <https://string-db.org/>). Interacting proteins from each cluster were further subjected to GOBP analysis, and the top five enriched pathways were selected for visualization. Cytoscape (version 3.9.1) was used to illustrate the PPI, importance, and pathway enrichment of ERI genes. To calculate the infiltration score of TME components, we used the R package “IOBR” (version 0.99.8), which incorporates various algorithms such as Cibersort, EPIC, Estimate, IPS, MCP_counter, TIMER, and xCell.³⁸

Correlation analysis and Mantel test analysis

The correlation between individual TME components and ERI was assessed using Spearman’s correlation, and the results were visualized with the R packages “corrplot” (version 0.92) and “ggcorrplot” (version 0.1.4.1). A Spearman correlation coefficient $|R| > 0.4$ and a *p* value of less than 0.05 were considered indicative of a significant correlation. The Mantel test results were investigated using the “linkET” package (version 0.0.7.4).

Single-cell RNA-seq data integration and analysis

For this study, scRNA-seq data from ncb and GSE161529 were downloaded from the GEO database. These datasets provided scRNA-seq data from 17 TNBC and 9 Her2 primary tumor samples. The “Seurat” package (version 4.4) was utilized to integrate the scRNA-seq data.⁴⁶ Cells with an expressing gene count of less than 200 or greater than 6,000 were excluded. Additionally, cells with a mitochondrial content exceeding 15% were filtered out, resulting in a total of 137,032 cells with single-cell transcriptome data for further analysis. The “harmony” package (version 1.2.1) was used to mitigate batch effects between GSE176078 and GSE161529.⁴⁷ Uniform Manifold Approximation and Projection dimensionality reduction was used to project cells in two dimensions. Cell annotation was based on CellMarker 2.0.⁴⁸ The genomic variation of epithelial cells was examined using “infercnv” (version 1.16.0).⁴⁹ The ligand-receptor interactions among different cell clusters were measured using the R package “CellChat” (version 2.1.2).⁵⁰

DATA AVAILABILITY

The raw data files will be provided for scientific research upon request complying with the law due to human patient privacy concerns.

ACKNOWLEDGMENTS

This study was approved by the Ethical Committee of Harbin Medical University Cancer Hospital and complied with all relevant ethical regulations (KY2023-66). All patients provided written informed consent for the collection of samples for research and genomic profiling. This study was supported by the following grants: National Natural Science Foundation of China (82322048; 82272872); Natural Science Foundation of Heilongjiang Province of China (JQ2023H004); and Basic scientific research business expenses of colleges and universities in Heilongjiang Province (31031240062).

AUTHOR CONTRIBUTIONS

D.J., X.S., and X.L. collected the samples and clinical information of patients. H.R. and S.L. processed the analysis and finished the writing of this manuscript. Y.L., T.L., and W.W. supervised the work.

DECLARATION OF INTERESTS

The authors have no conflicts of interest to declare in relation to this work. All of the authors read and approved the final manuscript for publication.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.omton.2025.200965>.

REFERENCES

- Siegel, R.L., Giaquinto, A.N., and Jemal, A. (2024). Cancer statistics. *CA Cancer J. Clin.* 74, 12–49. <https://doi.org/10.3322/caac.21820>.
- Giaquinto, A.N., Sung, H., Miller, K.D., Kramer, J.L., Newman, L.A., Miniham, A., Jemal, A., and Siegel, R.L. (2022). Breast Cancer Statistics. *CA Cancer J. Clin.* 72, 524–541. <https://doi.org/10.3322/caac.21754>.
- Harbeck, N., and Gnant, M. (2017). Breast cancer. *Lancet* 389, 1134–1150. [https://doi.org/10.1016/S0140-6736\(16\)31891-8](https://doi.org/10.1016/S0140-6736(16)31891-8).
- Onkar, S.S., Carleton, N.M., Lucas, P.C., Bruno, T.C., Lee, A.V., Vignali, D.A.A., and Oesterreich, S. (2023). The Great Immune Escape: Understanding the Divergent Immune Response in Breast Cancer Subtypes. *Cancer Discov.* 13, 23–40. <https://doi.org/10.1158/2159-8290.CD-22-0475>.
- Bianchini, G., De Angelis, C., Licata, L., and Gianni, L. (2022). Treatment landscape of triple-negative breast cancer - expanded options, evolving needs. *Nat. Rev. Clin. Oncol.* 19, 91–113. <https://doi.org/10.1038/s41571-021-00565-2>.
- Schmid, P., Cortes, J., Dent, R., McArthur, H., Pusztai, L., Kümmel, S., Denkert, C., Park, Y.H., Hui, R., Harbeck, N., et al. (2024). Overall Survival with Pembrolizumab in Early-Stage Triple-Negative Breast Cancer. *N. Engl. J. Med.* 391, 1981–1991. <https://doi.org/10.1056/NEJMoa2409932>.
- Dent, R., André, F., Gonçalves, A., Martin, M., Schmid, P., Schütz, F., Kümmel, S., Swain, S.M., Bilici, A., Loirat, D., et al. (2024). IMpassion132 double-blind randomised phase III trial of chemotherapy with or without atezolizumab for early relapsing unresectable locally advanced or metastatic triple-negative breast cancer. *Ann. Oncol.* 35, 630–642. <https://doi.org/10.1016/j.annonc.2024.04.001>.
- Guo, S., Liu, X., Zhang, J., Huang, Z., Ye, P., Shi, J., Stalin, A., Wu, C., Lu, S., Zhang, F., et al. (2023). Integrated analysis of single-cell RNA-seq and bulk RNA-seq unravels T cell-related prognostic risk model and tumor immune microenvironment modulation in triple-negative breast cancer. *Comput. Biol. Med.* 161, 107066. <https://doi.org/10.1016/j.combiomed.2023.107066>.
- He, Y., Jiang, Z., Chen, C., and Wang, X. (2018). Classification of triple-negative breast cancers based on Immunogenomic profiling. *J. Exp. Clin. Cancer Res.* 37, 327. <https://doi.org/10.1186/s13046-018-1002-1>.
- Zou, Y., Xie, J., Zheng, S., Liu, W., Tang, Y., Tian, W., Deng, X., Wu, L., Zhang, Y., Wong, C.W., et al. (2022). Leveraging diverse cell-death patterns to predict the prognosis and drug sensitivity of triple-negative breast cancer patients after surgery. *Int. J. Surg.* 107, 106936. <https://doi.org/10.1016/j.ijsu.2022.106936>.
- Liu, H., Zhang, W., Zhang, Y., Adegboro, A.A., Fasoranti, D.O., Dai, L., Pan, Z., Liu, H., Xiong, Y., Li, W., et al. (2024). Mime: A flexible machine-learning framework to construct and visualize models for clinical characteristics prediction and feature selection. *Comput. Struct. Biotechnol. J.* 23, 2798–2810. <https://doi.org/10.1016/j.csbj.2024.06.035>.
- Verschoor, N., Smid, M., Jager, A., Sleijfer, S., Wilting, S.M., and Martens, J.W.M. (2023). Integrative whole-genome and transcriptome analysis of HER2-amplified metastatic breast cancer. *Breast Cancer Res.* 25, 145. <https://doi.org/10.1186/s13058-023-01743-z>.
- Liu, S., Wang, Z., Su, Y., Qi, L., Yang, W., Fu, M., Jing, X., Wang, Y., and Ma, Q. (2021). A neuroanatomical basis for electroacupuncture to drive the vagal-adrenal axis. *Nature* 598, 641–645. <https://doi.org/10.1038/s41586-021-04001-4>.
- Alves dos Santos, M.T.M., and Smidt, M.P. (2011). En1 and Wnt signaling in midbrain dopaminergic neuronal development. *Neural Dev.* 6, 23. <https://doi.org/10.1186/1749-8104-6-23>.
- Rekaik, H., Blaudin de Thé, F.X., Fuchs, J., Massiani-Beaudoin, O., Prochiantz, A., and Joshi, R.L. (2015). Engrailed Homeoprotein Protects Mesencephalic Dopaminergic Neurons from Oxidative Stress. *Cell Rep.* 13, 242–250. <https://doi.org/10.1016/j.celrep.2015.08.076>.
- Zheng, H.F., Forgetta, V., Hsu, Y.H., Estrada, K., Rosello-Diez, A., Leo, P.J., Dahia, C.L., Park-Min, K.H., Tobias, J.H., Kooperberg, C., et al. (2015). Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature* 526, 112–117. <https://doi.org/10.1038/nature14878>.
- Allou, L., Balzano, S., Magg, A., Quinodoz, M., Royer-Bertrand, B., Schöpflin, R., Chan, W.L., Speck-Martins, C.E., Carvalho, D.R., Farage, L., et al. (2021). Non-coding deletions identify Maenli lncRNA as a limb-specific En1 regulator. *Nature* 592, 93–98. <https://doi.org/10.1038/s41586-021-03208-9>.
- Terekhanova, N.V., Karpova, A., Liang, W.W., Strzalkowski, A., Chen, S., Li, Y., Southard-Smith, A.N., Iglesia, M.D., Wendt, M.C., Jayasinghe, R.G., et al. (2023). Epigenetic regulation during cancer transitions across 11 tumour types. *Nature* 623, 432–441. <https://doi.org/10.1038/s41586-023-06682-5>.
- Cui, Y., Zhang, Y., Liu, Y., Zhou, Z., Zhu, L., and Zhou, C.X. (2024). EN1 promotes lung metastasis of salivary adenoid cystic carcinoma by regulating the PI3K-AKT pathway and epithelial-mesenchymal transition. *Cancer Cell Int.* 24, 51. <https://doi.org/10.1186/s12935-024-03230-7>.
- Xu, J., Roe, J.S., Lee, E., Tonelli, C., Ji, K.Y., Younis, O.W., Somerville, T.D.D., Yao, M., Milazzo, J.P., Tiriak, H., et al. (2024). Engrailed-1 Promotes Pancreatic Cancer Metastasis. *Adv. Sci.* 11, e2308537. <https://doi.org/10.1002/adv.202308537>.
- Wu, S.Y., Lee, C.F., Lai, H.T., Yu, C.T., Lee, J.E., Zuo, H., Tsai, S.Y., Tsai, M.J., Ge, K., Wan, Y., and Chiang, C.M. (2020). Opposing Functions of BRD4 Isoforms in Breast Cancer. *Mol. Cell* 78, 1114–1132.e10. <https://doi.org/10.1016/j.molcel.2020.04.034>.
- Peluffo, G., Subedee, A., Harper, N.W., Kingston, N., Jovanović, B., Flores, F., Stevens, L.E., Beca, F., Trinh, A., Chilamakuri, C.S.R., et al. (2019). EN1 Is a Transcriptional Dependency in Triple-Negative Breast Cancer Associated with Brain Metastasis. *Cancer Res.* 79, 4173–4183. <https://doi.org/10.1158/0008-5472.CAN-18-3264>.
- Ping, J., Jia, G., Cai, Q., Guo, X., Tao, R., Ambrosone, C., Huo, D., Ambs, S., Barnard, M.E., Chen, Y., et al. (2024). Using genome and transcriptome data from African-ancestry female participants to identify putative breast cancer susceptibility genes. *Nat. Commun.* 15, 3718. <https://doi.org/10.1038/s41467-024-47650-5>.
- Su, Y., Subedee, A., Bloustain-Qimron, N., Savova, V., Krzystanek, M., Li, L., Marusyk, A., Tabassum, D.P., Zak, A., Flacker, M.J., et al. (2015). Somatic Cell Fusions Reveal Extensive Heterogeneity in Basal-like Breast Cancer. *Cell Rep.* 11, 1549–1563. <https://doi.org/10.1016/j.celrep.2015.05.011>.
- Beltran, A.S., Graves, L.M., and Blancafort, P. (2014). Novel role of Engrailed 1 as a prosurvival transcription factor in basal-like breast cancer and engineering of interference peptides block its oncogenic function. *Oncogene* 33, 4767–4777. <https://doi.org/10.1038/onc.2013.422>.
- Cock, P.J.A., Fields, C.J., Goto, N., Heuer, M.L., and Rice, P.M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* 38, 1767–1771. <https://doi.org/10.1093/nar/gkp1137>.
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- Bos, P.D., Zhang, X.H.F., Nadal, C., Shu, W., Gomis, R.R., Nguyen, D.X., Minn, A.J., van de Vijver, M.J., Gerald, W.L., Foekens, J.A., and Massagué, J. (2009). Genes that mediate breast cancer metastasis to the brain. *Nature* 459, 1005–1009. <https://doi.org/10.1038/nature08021>.
- Juul, N., Szallasi, Z., Eklund, A.C., Li, Q., Burrell, R.A., Gerlinger, M., Valero, V., Andreopoulou, E., Esteve, F.J., Symmans, W.F., et al. (2010). Assessment of an RNA interference screen-derived mitotic and ceramide pathway metagene as a predictor of response to neoadjuvant paclitaxel for primary triple-negative breast cancer: a retrospective analysis of five clinical trials. *Lancet Oncol.* 11, 358–365. [https://doi.org/10.1016/S1470-2045\(10\)70018-8](https://doi.org/10.1016/S1470-2045(10)70018-8).

30. Li, Y., Zou, L., Li, Q., Haibe-Kains, B., Tian, R., Li, Y., Desmedt, C., Sotiriou, C., Szallasi, Z., Iglehart, J.D., et al. (2010). Amplification of LAPTM4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. *Nat. Med.* 16, 214–218. <https://doi.org/10.1038/nm.2090>.
31. Kao, K.J., Chang, K.M., Hsu, H.C., and Huang, A.T. (2011). Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. *BMC. Cancer.* 11, 143. <https://doi.org/10.1186/1471-2407-11-143>.
32. Sabatier, R., Finetti, P., Cervera, N., Lambaudie, E., Esterni, B., Mamessier, E., Tallet, A., Chabannon, C., Extra, J.M., Jacquemier, J., et al. (2011). A gene expression signature identifies two prognostic subgroups of basal breast cancer. *Breast Cancer Res. Treat.* 126, 407–420. <https://doi.org/10.1007/s10549-010-0897-9>.
33. Tofigh, A., Suderman, M., Paquet, E.R., Livingstone, J., Bertos, N., Saleh, S.M., Zhao, H., Souleimanova, M., Cory, S., Lesurf, R., et al. (2014). The prognostic ease and difficulty of invasive breast carcinoma. *Cell Rep.* 9, 129–142. <https://doi.org/10.1016/j.celrep.2014.08.073>.
34. Jezequel, P., Loussouarn, D., Guerin-Charbonnel, C., Campion, L., Vanier, A., Gouraud, W., Lasla, H., Guette, C., Valo, I., Verrielle, V., and Campone, M. (2015). Gene-expression molecular subtyping of triple-negative breast cancer tumours: importance of immune response. *Breast. Cancer. Res.* 17, 43. <https://doi.org/10.1186/s13058-015-0550-y>.
35. Metzger-Filho, O., Michiels, S., Bertucci, F., Catteau, A., Salgado, R., Galant, C., Fumagalli, D., Singhal, S.K., Desmedt, C., Ignatiadis, M., et al. (2013). Genomic grade adds prognostic value in invasive lobular carcinoma. *Ann. Oncol.* 24, 377–384. <https://doi.org/10.1093/annonc/mds280>.
36. Brueffer, C., Vallon-Christersson, J., Grabau, D., Ehinger, A., Häkkinen, J., Hegardt, C., Malina, J., Chen, Y., Bendahl, P.O., Manjer, J., et al. (2018). Clinical Value of RNA Sequencing-Based Classifiers for Prediction of the Five Conventional Breast Cancer Biomarkers: A Report From the Population-Based Multicenter Sweden Cancerome Analysis Network-Breast Initiative. *JCO Precis. Oncol.* 2, 1–18. <https://doi.org/10.1200/PO.17.00135>.
37. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28, 882–883. <https://doi.org/10.1093/bioinformatics/bts034>.
38. Zeng, D., Ye, Z., Shen, R., Yu, G., Wu, J., Xiong, Y., Zhou, R., Qiu, W., Huang, N., Sun, L., et al. (2021). IOBR: Multi-Omics Immuno-Oncology Biological Research to Decode Tumor Microenvironment and Signatures. *Front. Immunol.* 12, 687975. <https://doi.org/10.3389/fimmu.2021.687975>.
39. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
40. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47. <https://doi.org/10.1093/nar/gkv007>.
41. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1, 417–425. <https://doi.org/10.1016/j.cels.2015.12.004>.
42. Yu, G., and He, Q.Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* 12, 477–479. <https://doi.org/10.1039/c5mb00663e>.
43. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* 2, 100141. <https://doi.org/10.1016/j.xinn.2021.100141>.
44. Kumar, L., and E Futschik, M. (2007). Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* 2, 5–7. <https://doi.org/10.6026/97320630002005>.
45. Kovacs, S.A., Fekete, J.T., and Gyorffy, B. (2023). Predictive biomarkers of immunotherapy response with pharmacological applications in solid tumors. *Acta. Pharmacol. Sin.* 44, 1879–1889. <https://doi.org/10.1038/s41401-023-01079-6>.
46. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502. <https://doi.org/10.1038/nbt.3192>.
47. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* 16, 1289–1296. <https://doi.org/10.1038/s41592-019-0619-0>.
48. Hu, C., Li, T., Xu, Y., Zhang, X., Li, F., Bai, J., Chen, J., Jiang, W., Yang, K., Ou, Q., et al. (2023). CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res.* 51, D870–D876. <https://doi.org/10.1093/nar/gkac947>.
49. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401. <https://doi.org/10.1126/science.1254257>.
50. Jin, S., Plikus, M.V., and Nie, Q. (2025). CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics. *Nat. Protoc.* 20, 180–219. <https://doi.org/10.1038/s41596-024-01045-4>.