

ARTICLE **OPEN**

# The novel endogenous micropeptide XLH-36 binds Gemin4 to promote triple-negative breast cancer metastasis

Mengwei Li<sup>1,5</sup>, Wenwei Xie<sup>1,5</sup>, Kexin Hu<sup>1</sup>, Tiantian Li<sup>1</sup>, Xin Li<sup>1</sup>, Li Shun<sup>1</sup> , Xiaowei Xu<sup>1</sup>, Junhui Chen<sup>3</sup> and Hanmei Xu<sup>1,4</sup>

© The Author(s) 2025

Triple-negative breast cancer (TNBC) is the most malignant breast cancer subtype. The 5-year overall survival rate of TNBC patients is 77%, but once cancer metastasis occurs, this rate drops to 12%. To date, the discovery of biomarkers associated with TNBC metastasis remains a major challenge. In this study, we identified a novel translated product, XLH-36 micropeptide, encoded by lncRNA (long non-coding RNA) C5orf66-AS1, which acts as an “oncogenic driver” in TNBC. Through conservation analyses across 101 species, we found that XLH-36 is highly conserved in humans and primates. Analysis of RNA-seq data from 1295 breast cancer patients, including 165 TNBC patients from different cohorts, revealed significantly elevated XLH-36 expression levels in breast cancer and TNBC. Specifically, TNBC patients with low C5orf66-AS1 (encoding XLH-36) expression demonstrated a 20% higher overall survival rate compared to those with high C5orf66-AS1 expression over a 50-month follow-up. XLH-36 knockout inhibited the growth and metastasis of TNBC xenografts in mice. Furthermore, we found that XLH-36 directly binds Gemin4, causing it to remain in the cytoplasm and preventing it from carrying out its role in promoting S100A4 mRNA splicing in the nucleus. This leads to a compensatory increase in ICAM1 levels, ultimately resulting in the promotion of epithelial-to-mesenchymal transition (EMT) in TNBC cells and an increase in tumor metastasis. In summary, our findings highlight the crucial role of XLH-36 in TNBC metastasis, which could be exploited in the development of therapeutic and diagnostic strategies for TNBC patients.

*Oncogene* (2026) 45:264–277; <https://doi.org/10.1038/s41388-025-03621-8>

## INTRODUCTION

Breast cancer is the most prevalent cancer type globally and accounts for 23.8% of all cancer cases in women [1]. Triple-negative breast cancer (TNBC) accounts for 15–20% of all breast cancer cases. Due to the lack of HER2, estrogen, or progesterone receptors, the standard of care for non-surgical TNBC remains non-specific chemotherapy [2, 3]. The 5-year overall survival rate of TNBC patients is 77%, but once metastasis occurs, this rate drops to 12% [4]. Although several biomarkers (including EGFR, VEGF, PARP1, and C-kit) [5, 6] are associated with TNBC metastasis, clinical treatment faces challenges as no substantial improvements in outcomes have been made. Recently, three micropeptides (CIP2A-BP, ASRPS, and XBP1SBM) [7–9] were found to be involved in TNBC development, and the functions and mechanisms of more novel micropeptides are largely unknown. These new discoveries and other studies derived from the detection of unannotated small open reading frames (sORFs) that previously hidden in non-coding regions (such as lncRNAs, circRNAs, pri-miRNAs, 5'-UTR, and 3'-UTR) actually have the ability to encode micropeptides [10–12]. Emerging evidence suggests that these micropeptides can act as proto-oncogenes or oncogenes in different tumors to regulate the processes of cell proliferation, metastasis, angiogenesis, energy metabolism, and drug resistance [13–16].

By combining Ribo-seq, mass spectrometry (MS), machine learning, and gene editing technology, our team established a new micropeptide discovery platform and found that MIAC, encoded by the lncRNA AC025154.2, has an important role in head and neck squamous cell carcinoma and kidney cancer [17, 18]. In the present study, using the same platform, we discovered the novel 36-amino acid translated product, and we named it micropeptide XLH-36, which is encoded by the lncRNA C5orf66-AS1. Furthermore, combining analysis of clinical samples with cell biological and molecular studies, we explored the binding proteins and molecular mechanisms of XLH-36, and investigated the relationship between XLH-36 and TNBC tumorigenesis, hoping to provide potential biomarkers and targets for TNBC diagnosis and treatment.

## MATERIALS AND METHODS

### Mice and animal housing

BALB/C female nude mice were purchased from Jiangsu Gempharmatech. All mice were housed in the SPF barrier environment of the Pharmaceutical Animal Center, China Pharmaceutical University.

### Xenograft studies

Five-week-old female nude mice obtained from Jiangsu Gempharmatech were maintained under specific pathogen-free conditions. Subcutaneous

<sup>1</sup>Jiangsu Province Engineering Research Center of Synthetic Peptide Drug Discovery and Evaluation, China Pharmaceutical University, Nanjing, China. <sup>2</sup>Key Laboratory of Quantitative Synthetic Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China.

<sup>3</sup>Department of Minimally Invasive Intervention, Peking University Shenzhen Hospital, Shenzhen, Guangdong, China. <sup>4</sup>State Key Laboratory of Natural Medicines, Ministry of Education, China Pharmaceutical University, Nanjing, China. <sup>5</sup>These authors contributed equally: Mengwei Li, Wenwei Xie. email: xuhanmei6688@126.com

Received: 10 April 2024 Revised: 23 September 2024 Accepted: 24 October 2025  
Published online: 29 November 2025

injections were performed at a density of 100  $\mu$ l containing different groups of  $6.5 \times 10^6$  MDA-MB-231 cells in the left para-mammary gland of nude mice. Tumor growth was observed every 3 days, and the volume was calculated by the following formula: volume = (length  $\times$  width $^2$ )/2. The nude mice were euthanized according to their status, and tumors were dissected and weighed for analysis.

For the *in vivo* metastasis model, MDA-MB-231 cells ( $3 \times 10^6$ ) with different genetic manipulations (XLH-36 overexpression or knockout) were injected into the tail vein of mice, tumor metastasis was monitored twice a week through the IVIS Spectrum living imaging system (PerkinElmer) for 4 weeks. At the endpoint of the experiment, mice were humanely euthanized. All experiments using animals were conducted under the Institutional Animal Care and Use Committee (IACUC)-approved protocols at China Pharmaceutical University (IACUC-202312016) by NIH and institutional guidelines.

### Cell lines and cell culture

MCF10A, BT474, MCF7, MDA-MB-231, MDA-MB-468, and 293T cells were cultured in DMEM (GIBCO) supplemented with 10% FBS (Bioind). T47D cells were cultured in RPMI 1640 (GIBCO) supplemented with 10% FBS (Bioind). Cells were kept at 37 °C and 5% CO<sub>2</sub>.

### RNA isolation and qRT-qPCR

Total RNA was extracted via TRIzol (TIANGEN Biotech #G1012). Two milligrams of the RNA (measured by Nanodrop spectrophotometer) was reverse transcribed to cDNA by 5X All-In-One MasterMix (with AccuRT Genomic DNA Removal Kit, Applied Biological Materials #G492) in 20  $\mu$ l system. qRT-PCR analysis was performed using Blastaq 2X qPCR MasterMix (Applied Biological Materials #G891). The mRNA expression was normalized by the expression of GAPDH and relative expression levels were calculated using the  $2^{-\Delta\Delta C_t}$  method in the indicated cell lysates. The primer sequences were listed in Supplementary Table 2.

### Cell transfection and lentivirus infection

The siRNA sequences targeting ICAM1, S100A4, and XLH-36 gene and non-targeting siRNA control were listed in Supplementary Table 3. Using Lipofectamine 8000 (Beyotime Biotechnology) according to the instructions. Lentiviruses were produced in HEK293T cells with package vectors pMD2.G and psPAX2. The virus was collected to transduce target cells, followed by selection with puromycin (2  $\mu$ g/mL).

### CRISPR/Cas9-mediated knock-in and knockout

For XLH-36 knock-in (KI), the single-guide RNA (sgRNA) was designed to target the stop codon region of XLH-36 ORF to insert a 1\* FLAG- epitope at its C-terminal. The designed sgRNA was synthesized (GenScript) and cloned into a pSpCas9(BB)-2A-GFP vector (PX458, Addgene). As a DNA donor, a single-stranded oligodeoxynucleotide (ssODN) with homology arms specific to the XLH-36 locus was used: CGGACCCCTAGCTACGCCCG AGGGAGGCCAGGACCCGACTACAAGGACGACGATGACAAGTAGCCGGCGG-GACTGCGCCGCCCTCTCCCGCAGGTCCC. PX458 was co-transfected with ssODN into cells using Lipofectamine 3000 (Thermo) and sorted using the FITC channel on a BD FACSaria II SORP. The editing for KI was validated by PCR, DNA sequencing, immunofluorescence, and IP followed by MS identification.

For XLH-36 knock-out (KO), two different single-guide RNAs (sgRNA) were designed to target the start and stop codon region of XLH-36 ORF, respectively. The designed sgRNAs were synthesized and cloned into pYSY-U6-sgRNA-EFla-mCherry and pYSY-U6-sgRNA-EFla-eGFP (YSY Biotech, Nanjing, China). The lentivirus particle expressing Cas9 protein was generated (Hanbio, Shanghai, China) and transfected with MDA-MB-231 cells (MDA-MB-231-Cas9). The pYSY-U6-sgRNA-EFla-mCherry and pYSY-U6-sgRNA-EFla-eGFP plasmids were co-transfected into MDA-MB-231-Cas9 cells and sorted using the FITC and PE channels on a BD FACSaria II SORP. The developed monoclonal XLH-36-KO populations were examined for editing by DNA sequencing and Western blot. The sgRNAs for XLH-36 KI and KO were listed in Supplementary Table 3.

### Preparation of XLH-36 monoclonal antibody

XLH-36 peptide was used as the immunogen, which was synthesized by Abmart. Each set of antigens will be used to immunize 12 Balb/c mice (8–12 weeks old) and their serum potency will be monitored to determine the optimal number of immunizations. The optimized adjuvants and

immunization methods produce high-affinity antibodies (IgG subtype) against most antigenic peptides. Initial immunization is followed by three to four boosts, after which mouse sera will be taken for titer testing (recombinant protein encapsulated as an antigen) by ELISA assay. Serum potency needs to be greater than 10 K, otherwise continue to strengthen the immunization. Whole spleen and 1/2 of the lymph nodes are taken and fused with myeloma SP2/0 cell line. The fused cells were spread onto four 384-well plates ( $10^2$ – $10^4$  cells per well) and cultured. Supernatants of all wells were collected and positive cell lines were screened by ELISA method to detect the reaction of cell supernatants against XLH-36 peptide, and positive wells with cells microscopically detected were transferred to 96-well plates for further culture. After a few days of growth, the supernatants of all wells were collected to detect the reactivity of cell supernatants with XLH-36 peptide by ELISA. Monoclonal hybridoma cells were spread in a 96-well plate and cultured until the bottom 1/6 plate was covered. ELISA was performed to detect the reaction of each well supernatant against the XLH-36 peptide, and the two wells with high OD values and good cell status were selected for the next round of subcloning. Repeat the above steps until the wells are 100% positive for the cell line. After the last round of subcloning, all the positive cells were immediately expanded and cultured, and the produced ascites was purified with Protein A/G and used for subsequent assays.

### Protein extraction and Western blot analysis

Cells were collected and total protein was extracted using RIPA lysis buffer (Beyotime Biotechnology) supplemented with Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific). Supernatants were obtained by centrifugation at 12,000  $\times g$  for 10 min at 4 °C. Cell lysate was added to 12% SDS-PAGE and transferred to PVDF membranes (Millipore Sigma). Membranes were incubated with primary antibodies overnight. HRP-linked goat anti-rabbit/mouse IgG was used as the secondary antibody. GAPDH was used as the loading control. The blots were visualized by ECL. Blot images were obtained using Tanon software (TANON-5200 system, Shanghai, PR China).

### Immunoprecipitation (IP)

Cells were collected and lysed. Two microliter primary antibody was added into 1.5 mg protein solution. Indicated 0.4  $\mu$ l control IgG was added to the 400  $\mu$ g lysates. After incubation at 4 °C overnight, 35  $\mu$ l protein A/G magnetic beads were added to the lysates and then incubated for 3 h. Then the magnetic beads were washed 3 times with 500  $\mu$ l protease cracking solution containing RIPA at 4 °C and 6000  $\times g$ , each time for 1 min. The bead-enriched proteins were detected by Western blot analysis.

### Subcellular fractionation

Using Subcellular Fractionation Isolation Kit (Thermo Fisher Scientific) according to the instructions. Cells were collected and lysed. Cytoplasmic solution, cell membrane solution, nuclear solution, chromatin solution and cytoskeletal solution were separated by CEB, MEB, NEB and PEB reagents.

### Endoplasmic reticulum mitochondrial sub-fractionation

Using Endoplasmic Reticulum Isolation Kit (Sigma-Aldrich) according to the instructions. Lysed and collected  $2 \times 10^8$  cells. Hypotonic solution with three times the volume of precipitation was added and the cells were incubated at 4 °C for 20 min. After centrifugation at 600  $\times g$  for 5 min, two times the volume of 1× Isotonic solution was added, ground in a homogenizer for 20 min, centrifuged at 1000  $\times g$  for 25 min, slowly add 7.5 times the volume of CaCl<sub>2</sub> solution to the supernatant solution, stirred at 4 °C for 15 min, and centrifuged at 8000  $\times g$  for 15 min to obtain the precipitate. 300  $\mu$ l 1× Isotonic solution was added to the precipitate and ground in a homogenizer for 15 min to obtain an ER solution.

Using Mitochondrial Isolation Kit (Thermo Fisher Scientific) according to the instructions. Lysed and collected  $2 \times 10^7$  cells. 800  $\mu$ l of reagent A was added, vortexed for 5 s at medium speed, and incubated on ice for 2 min. Then 10  $\mu$ l of reagent B was added, vortexed for 5 s at high speed, incubated on ice for 5 min, vortexed once per minute, and 800  $\mu$ l of reagent C was added and gently mixed. After centrifugation at 700  $\times g$  for 10 min at 4 °C, the supernatant solution was further centrifuged at 1200  $\times g$  for 15 min at 4 °C, 500  $\mu$ l of reagent C was added to the precipitate and centrifuged at 12,000  $\times g$  for 5 min at 4 °C to obtain the mitochondrial precipitate.

### Immunofluorescence and confocal microscopy

Cells were fixed with  $\beta$ -galactosidase staining fixative (Beyotime Biotechnology #C0602) for 15 min at RT. After two washes with PBS, the membranes were permeabilized with Triton X-100 (Beyotime Biotechnology #P0096) for 15 min and washed twice with PBS. Block with 5% BSA at RT for 30 min, wash with PBS, add primary antibody, and incubate at 4 °C overnight. Cells were washed with PBS twice and added secondary antibodies for 2 h at RT in the dark. After washing twice with PBS, added 10  $\mu$ l DAPI (Servicebio Biotechnology # GDP1024), and the cells were incubated in the dark for 10 min and observed under a fluorescence microscope.

### Cell proliferation assay (CCK-8)

Cell proliferation was evaluated using cell-counting kit-8 (CCK-8 Kit, Yeasen Biotechnology, #40203ES76). Cells were seeded into 96-well plates at the density of 2000 cells/well. At the indicated time points, 10  $\mu$ l CCK-8 was added into the medium of the cells, and OD was measured by spectrophotometer at a wavelength of 450 nm (Thermo Scientific). All experiments were performed in triplicate.

### Flow cytometric (FC) analysis-cell cycle analysis

Using Cell Cycle Assay Kit-PI/RNase Staining (CCS012, MULTI SCIENCES (LIanke) BIOTECH CO., LTD) according to the instructions. Lysed and collected  $5 \times 10^5$  cells. Precooled 70% ethanol was added to the precipitate, and cells were resuspended and incubated overnight at -20 °C. After centrifugation at 2000 rpm for 5 min, 100  $\mu$ l RNase solution was added to the precipitate and incubated at 37 °C for 30 min, followed by 900  $\mu$ l PI solution and incubation at 37 °C for 30 min. The samples were analyzed using a CytoFLEX (Beckman Coulter) and FlowJo software.

### Cell migration and cell invasion assay

A total of  $1.5 \times 10^4$  cells were cultured in 200  $\mu$ l DMEM without FBS in the upper chamber, and 500  $\mu$ l DMEM containing 10% FBS was added to the lower chamber. After incubation of these chambers for 48 h, non-migrated cells on the upper membrane layer were removed, cells that had migrated through the membrane were fixed in precooled formaldehyde for 30 min, stained with 0.1% crystal violet for 15 min, and five randomly selected fields ( $\times 10$ ) of each chamber were photographed and counted.

After removing the OrganoGel Matrigel (CELLada #OM-1) from -20 °C, it should be thawed at 4 °C to a liquid state. All products touching Matrigel should be pre-cooled by ice bath before use, and the whole process is carried out on ice. Matrigel was diluted with pre-cooled DMEM without FBS (1:7), and 10  $\mu$ l of diluted Matrigel was evenly spread in the transwell chamber, which was placed at 37 °C for 30 min. The cell suspension was digested and prepared.  $3 \times 10^4$  cells were added to each chamber, and subsequent manipulations were consistent with cell migration. At 0, 12, 24, and 48 h, the cell migration distance was photographed under an inverted microscope at the same position and calculated. The distance of cell migration was measured and calculated using ImageJ-win64 software.

### Cell adhesion assay

Recombinant fibrin was dissolved in sterile PBS to 10  $\mu$ g/ml, mixed, and added to a 96-well plate, 50  $\mu$ l per well, and coated at 37 °C for 1 h. The well plates were washed twice by dipping with serum-free medium. Lysed and collected  $1 \times 10^4$  cells. Cells were resuspended in serum-free medium, cultured at 37 °C for 1 h, washed three times with PBS, and CCK8 was added to detect the number of cells.

### Cell wound scratch assay

Cells were inoculated in 24-well culture plates until the confluence reached 90%. Scratches were created by scraping the monolayer with a sterile pipette tip. The medium was then changed to DMEM with 2% FBS.

### Bioinformatic analysis

Homologous sequences of C5ORF66-AS1 ORF1 were identified using NCBI-BLAST or BLAT search of the UCSC Genome Browser (<https://genome.ucsc.edu/>). Further evaluations of XLH-36 structure and domains were performed using the following online servers: C-I-TASSER (<https://zhanglab.ccmb.med.umich.edu/C-I-TASSER/>), TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) and HDOCK (<http://hdock.phys.hust.edu.cn/>).

### Molecular docking

The 3D structure of the XLH-36 peptide was generated in HDOCK, and the 3D structure of the protein Gemin4 were downloaded from AlphaFold structure protein Database (AF-P57678-F1). The ClusPro server [19] was used for molecular docking of XLH-36 and Gemin4. The docked structures were analyzed using MOE 2019.1.

### MicroScale thermophoresis (MST)

Human Gemin4 protein was expressed by HEK293 and labeled by CY5 fluorescence (Xi'an ruixi Biological Technology Co., Ltd) via the sulphydryl group on the cysteine. The labeled Gemin4 protein and chemosynthetic XLH-36 peptide (Shanghai GL Biochem Ltd) were incubated for 5 min at room temperature, and reaction mixtures were enclosed in premium-coated glass capillaries and loaded into the instrument (Monolith NT.115, NanoTemper, Germany).  $K_d$  values were determined using the NanoTemper MO. Affinity analysis tool as described.

### Hematoxylin-eosin staining (H&E)

Mouse tissues were fixed in 4% formalin for 48 h, and cryo-sectioned after embedding using Tissue-TEK O.C.T. Compound (SAKURA #4583) to obtain 9  $\mu$ m thickness sections. The plates were fixed in 4% formalin for 5 min at RT. Sections were immersed in 70%, 85%, and 100% ethanol for 1 min. The plates were stained with hematoxylin staining solution (BBI #E607317-0100) for 3 min, rinsed with small water flow for 5 min, and then dried and stained with eosin staining solution (BBI #E607321-0100) for 3 min and rinsed with small water flow for 4 min. Sections were transferred from the following solutions with indicated times. 70% Ethanol (5 s)-85%Ethanol (5 s)-100% Ethanol (2 min)- Xylene (4 min)- Xylene (4 min). After staining, the slides were air-dried and covered with neutral balsam (Solarbio #G8590).

### LC-MS/MS analysis

The LC-MS/MS analysis of the immunoprecipitation complex with XLH-36 antibody in MDA-MB-231 cells was performed as previously described [17]. Total proteins were collected and separated by SDS-PAGE on 10–17% Bis-Tris acrylamide gels and stained with silver. Samples were dissolved in 0.1% TFA and 2% acetonitrile, and the resulting peptides were analyzed with a QExactive mass spectrometer (Thermo Fisher Scientific) coupled with an HPLC instrument (Dionex NCS3500). For the identification of XLH-36 interactors, the resulting spectra were searched against the UniProt database. A maximum of two missed cleavage sites was allowed. The top 10 candidate XLH-36 interactors are listed in the Supplementary Table 4.

### Sample and library preparation for RNA-seq

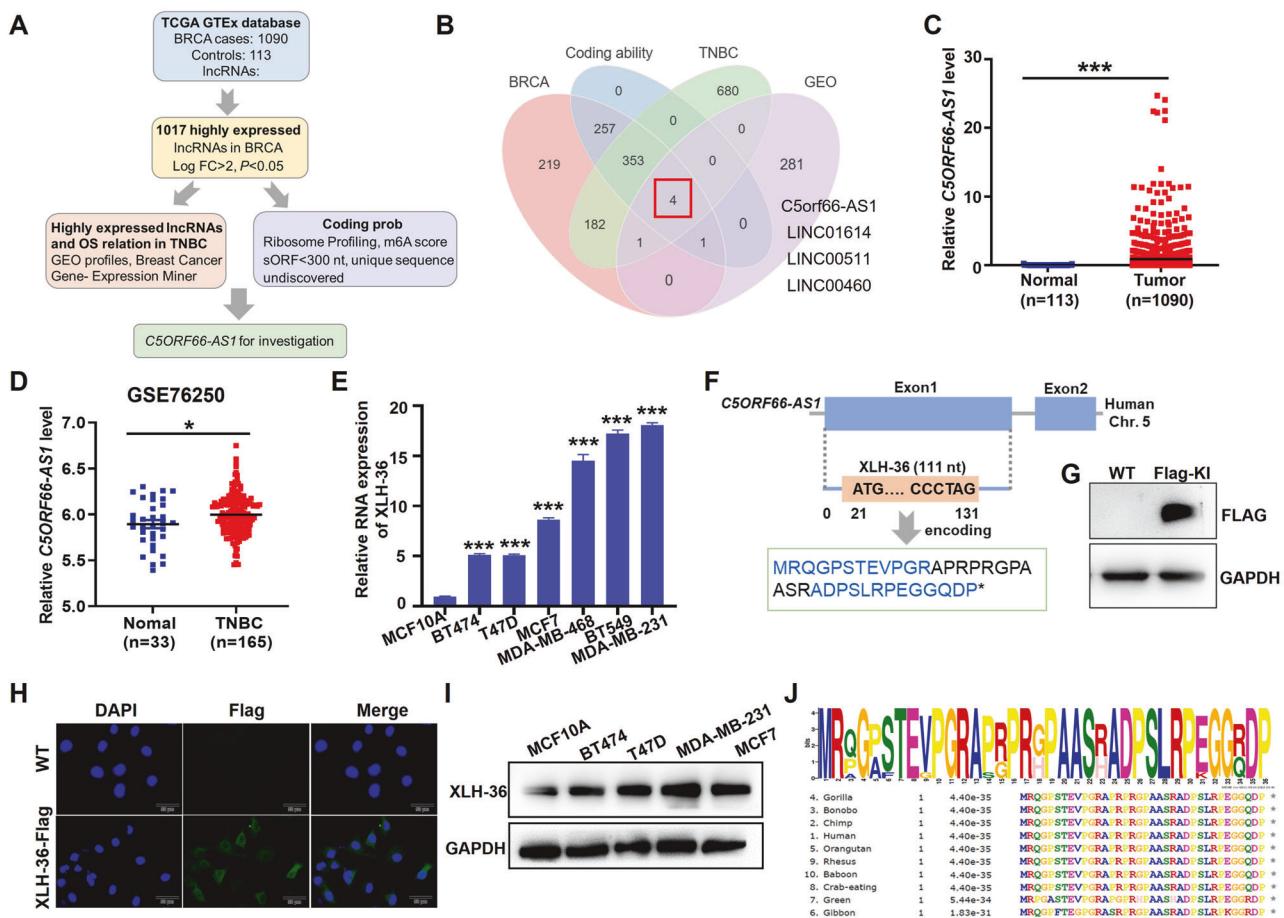
One million MDA-MB-231/WT and MDA-MB-231/XLH-36-KO cells were harvested and washed with cold PBS to remove trypsin by spinning at 500  $\times$  g for 5 min at 4 °C. RNA was purified using the Invitrogen TRIzol Reagent and quality control was assessed by Agilent 2100 BioAnalyzer (RIN  $\geq$  7.0, and rRNA ratio  $\geq$  1.5). Libraries for sequencing were prepared using the Illumina NEBNext Ultra RNA Library Prep Kit and sequenced on the Illumina NextSeq 500 with 1100-bp single-end sequencing. Quality control and library preparation for RNA from MDA-MB-231/WT and MDA-MB-231/XLH-36-KO cell lines were performed at CapitalBio Technology following their standard protocols. Replicates for specific RNA-seq experiments are indicated in Fig. 7B.

### Statistical methods

The data are presented as the means  $\pm$  standard deviation (SD) of a minimum of three biological replicates and were compared using the student's *t* test. Multiple comparisons were made by unpaired, two-tailed Student's *t* tests or one-way ANOVA were used. OS curves were assessed with the Kaplan-Meier method and compared by the Log-rank test. Statistical tests were conducted using GraphPad Prism (Version 8; La Jolla, CA, USA) software unless otherwise indicated. NS non-significant,  $^*P < 0.05$ ,  $^{**}P < 0.01$ ,  $^{***}P < 0.001$ .

### Analysis of RNA-seq data

Gene expression was evaluated as read count at the gene level with Hisat2. Gene expression data were normalized to counts per million (CPM), and differential expression between experimental groups was evaluated using LIMMA [20] and DESeq with  $P < 0.05$  and  $|\log_2(\text{fold change})| \geq 1$  as cut-offs. GO enrichment analysis was performed using PANTHER to determine the molecular and biological functions that were enriched among DEGs in



**Fig. 1** Large-scale screening identified the novel micropeptide XLH-36 in triple-negative breast cancer. **A** Schematic screening of differentially expressed of potential coding lncRNAs from TCGA and the bioinformatic database. **B** Venn diagrams showing differentially expressed lncRNAs overlaps between breast cancer, TNBC samples, and coding ability analysis. **C** C5orf66-AS1, encoding XLH-36, was significantly upregulated in BRCA patients ( $n = 1090$ ) compared to adjacent normal tissues ( $n = 113$ ) from TCGA. **D** The GSE76250 dataset indicated that C5orf66-AS1 was significantly upregulated in TNBC patients ( $n = 165$ ) compared to adjacent normal tissues ( $n = 33$ ). **E** XLH-36-encoding C5orf66-AS1 expression in breast cancer cells and normal cells. Expression levels were normalized to GAPDH. **F** Diagram of the location XLH-36-encoding region in C5orf66-AS1. The XLH-36-Flag fusion protein was expressed in MDA-MB-231 cells and detected by immunoblot (**G**) and immunofluorescence analysis (**H**). Nuclei were stained with DAPI (blue). **I** XLH-36 expression was detected in breast cancer cells and normal cells were detected by immunoblot with prepared anti-XLH-36 antibody. **J** Amino acid alignment of XLH-36 from different vertebrates. Data are presented as mean  $\pm$  SD from  $n = 3$  biologically independent experiments.  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ .

XLH-36-KO cells. The input gene lists included differentially upregulated and downregulated genes in two comparisons based on adjusted  $P < 0.05$ ,  $|\log_2(\text{fold change})| \geq 1$  (MDA-MB-231/WT compared to MDA-MB-231/XLH-36-KO). In total, 640 DEGs were identified. A cut-off value of FDR  $< 0.05$  was used to select the top enriched GO terms. RNA-seq data have been deposited in the GSA database (GSA accession number: HRA008378).

## RESULTS

### Differentially expressed lncRNA and potential coding sORF were selected by combining 1060 BRCA samples and 165 TNBC samples

Emerging evidence shows that functional peptides could be encoded by sORFs within lncRNAs [21], which have a structure and transcription process similar to those of mRNAs [22]. To identify lncRNAs with coding potential that are highly expressed in TNBC, we analyzed lncRNA data from 1090 BRCA and 113 non-neoplastic breast tissues from the TCGA database (Supplementary Table 1), as well as RNA-seq data from 165 TNBC samples from GEO database (GSE76250). Overall, 1017 highly expressed lncRNAs candidates ( $\log_2(\text{Fold Change}) \geq 2$ ,  $P \leq 0.01$ , Supplementary Fig. 1A) in BRCA tissues were identified based on the following inclusion criteria: (1) highly expressed in breast cancer and TNBC samples, with a

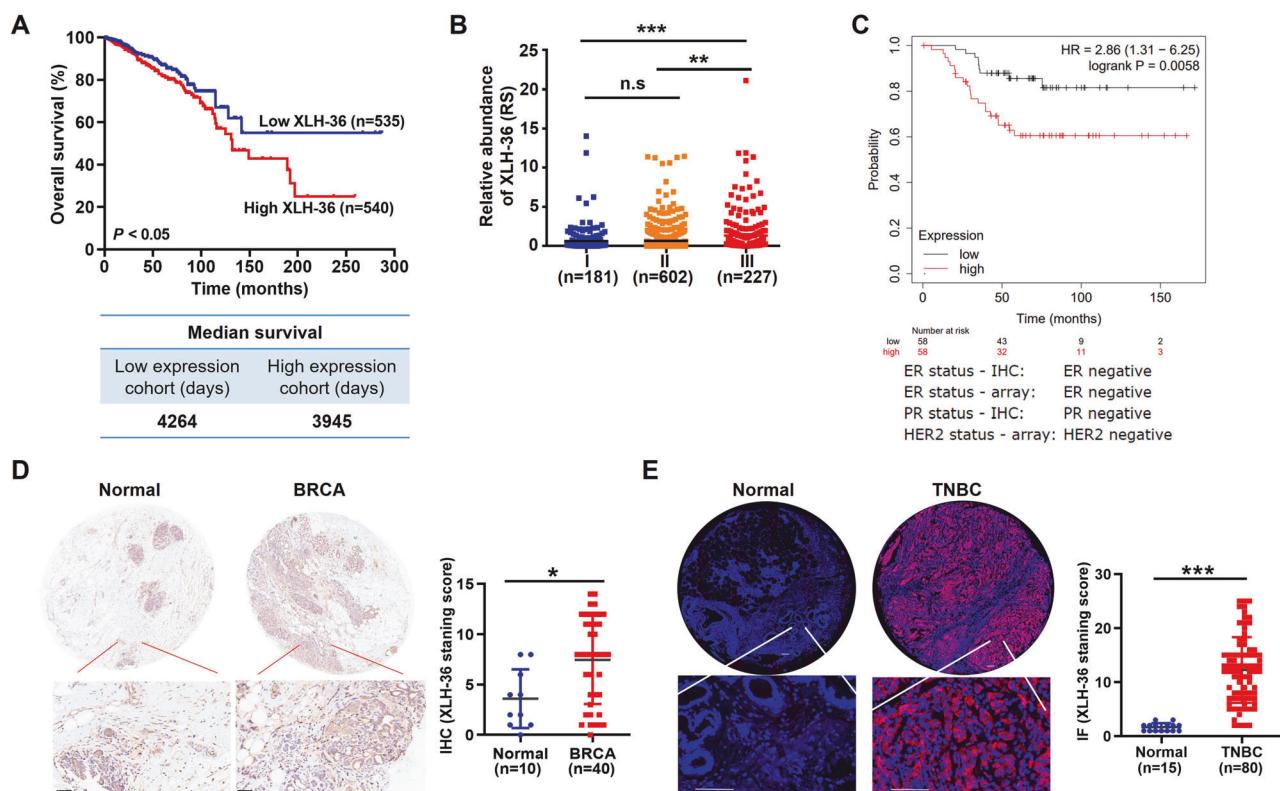
significant correlation with clinical prognosis; (2) good coding probability based on ribosome profiling and the m6A score according to the Translnc database; and (3) potential sORF length  $< 300$  nt, with a unique and undiscovered peptide sequence (Fig. 1A). There are four lncRNAs (C5orf66-AS1, LINC01614, LINC00511, LINC00460) that meet all three of the aforementioned screening conditions simultaneously (Fig. 1B), and C5orf66-AS1 has the highest overall score among all of them (Table 1 and Fig. 1C). Furthermore, the C5orf66-AS1 expression was significantly higher in TNBC clinical samples (Fig. 1D) and three TNBC cells (MDA-MB-231, MDA-MB-468, and BT549, Fig. 1E). Thus, in the present study, we focused on the role of this candidate micropeptide-encoding lncRNA in TNBC progression.

### Discovery of the highly conserved novel endogenous micropeptide XLH-36 in primates

To analyze the coding potential of C5orf66-AS1, we used PhyloCSF, the Coding Potential Calculator (CPC), and ORF finder analysis, and we found that a sORF spanning the 21–131 bp in exon 1 of C5orf66-AS1 could encode a 36-amino acid micropeptide (Fig. 1F and Supplementary Fig. 1B), which we termed XLH-36. To further confirm that XLH-36 is endogenously expressed, we knocked in a Flag-tag (Flag KI) upstream of the XLH-36 stop codon

**Table 1.** Summary of the high-potential coding lncRNAs differently overexpressed in breast cancer and TNBC patients.

| No | Gene name   | Log Fc | P value     | OS correlation | Upregulated in TNBC | sORF length | Coding ability | Unique |
|----|-------------|--------|-------------|----------------|---------------------|-------------|----------------|--------|
| 1  | C5ORF66-AS1 | 6.04   | $P < 0.001$ | $P < 0.05$     | $P < 0.001$         | 36 aa       | 0.4982         | Yes    |
| 2  | LINC01614   | 5.56   | $P < 0.001$ | $P < 0.05$     | $P < 0.001$         | 27 aa       | 0.2824         | Yes    |
| 3  | LINC00511   | 2.96   | $P < 0.001$ | $P < 0.05$     | $P < 0.001$         | 73 aa       | 0.6387         | Yes    |
| 4  | LINC00460   | 2.60   | $P < 0.001$ | $P < 0.001$    | $P < 0.001$         | 18 aa       | 0.567          | Yes    |



**Fig. 2** XLH-36 is highly expressed in TNBC samples and is a significant predictor of poor outcomes. **A** Overall survival curves of BRCA patients grouped according to XLH-36 expression. **B** XLH-36-encoding C5orf66-AS1 was significantly associated with the clinical stage of BRCA patients. **C** Low XLH-36 expression was associated with a high overall survival rate of TNBC patients. **D** XLH-36 levels were significantly increased in BRCA tissues by IHC assay with XLH-36 antibodies. Scale bar, 200  $\mu$ m. **E** XLH-36 levels were significantly increased in TNBC tissues by IF assay with XLH-36 antibodies. Scale bar, 50  $\mu$ m. The overall survival curve was generated with Kaplan–Meier analysis and the log-rank test. Lines show the means  $\pm$  SD. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

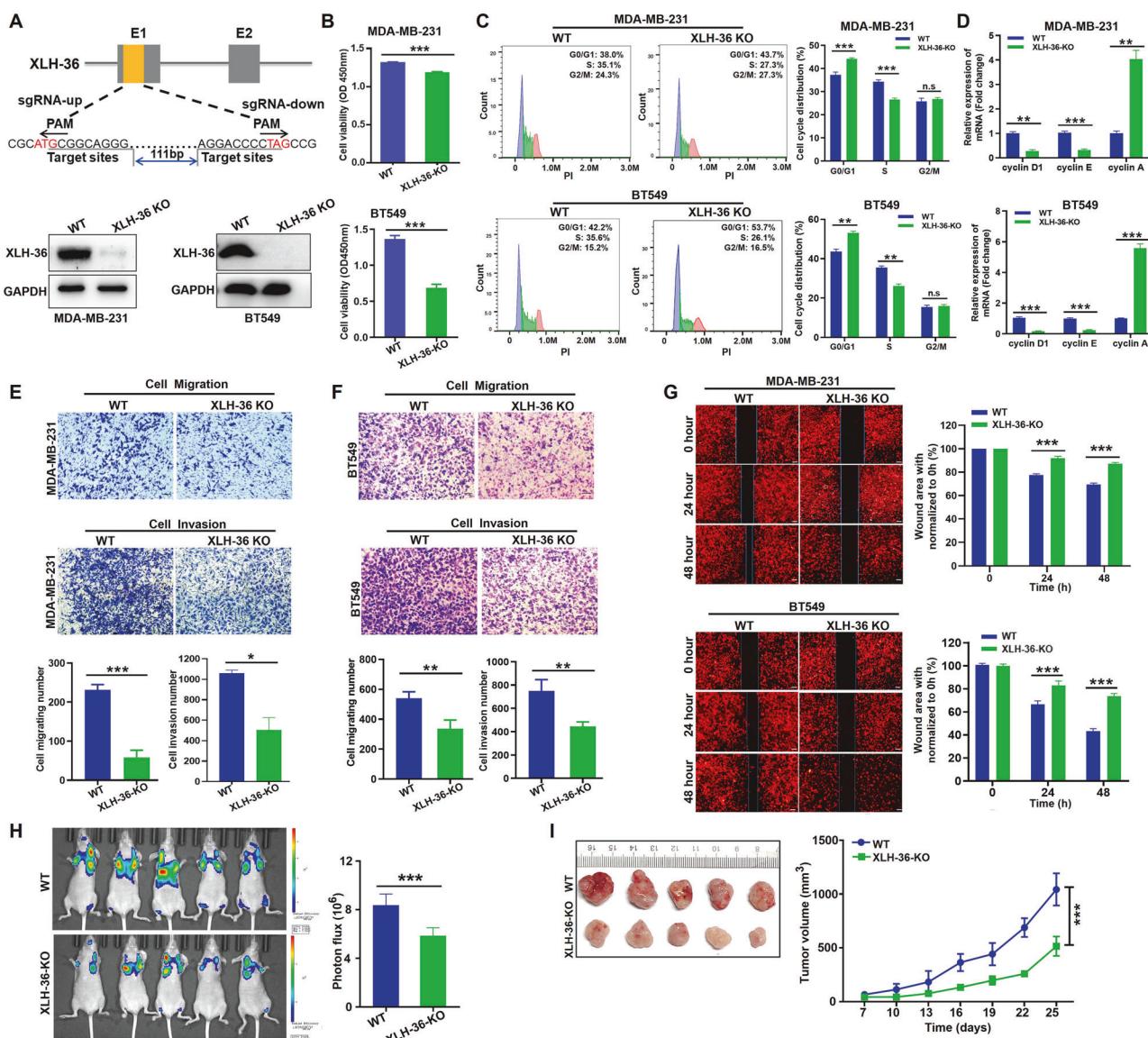
(Supplementary Fig. 1C). Immunoblotting (IB) and immunofluorescence (IF) analysis using anti-Flag suggested that XLH-36 was indeed endogenously translated (Fig. 1F, G). Moreover, we generated a monoclonal anti-XLH-36 antibody (Supplementary Fig. 1D, E) and performed IB in different BRCA cell lines (Fig. 1H). XLH-36-specific peptide sequences were also identified by immunoprecipitation (IP) with Flag-KI-enriched cell lysates, followed by MS analysis (Supplementary Fig. 1F, G).

As evolutionary conservation is considered the hallmark of functional ORFs, we investigated the codon conservation of XLH-36. Amino acid analysis also showed that XLH-36 is relatively conserved in primates (Fig. 1I), indicating that XLH-36 might be a functional micropeptide. Collectively, these data indicate that XLH-36 is naturally and endogenously translated from the sORF located in lncRNA C5orf66-AS1 (Chr.5 135,040,047–135,039,596).

#### XLH-36 is an oncogenic micropeptide that serves as a prognostic biomarker for TNBC

To uncover the clinical relevance of XLH-36, we divided the participants into two groups (XLH-36-low and XLH-36-high)

based on the median expression level of C5orf66-AS1. Kaplan–Meier survival analysis showed that patients with higher XLH-36 expression levels had a poor prognosis (Fig. 2A), and XLH-36 was upregulated in patients with advanced stage BRCA (Fig. 2B). In an independent TNBC patient cohort, Kaplan–Meier survival analysis revealed that high XLH-36 expression was associated with poor prognosis (Fig. 2C), while the expression level of XLH-36 does not strongly correlate with the prognosis of patients who are ER, PR, or HER2 positive (Supplementary Fig. 2A–C). Moreover, XLH-36 expression levels were determined by immunohistochemical analysis using anti-XLH-36 antibody. Consistent with the RNA expression, XLH-36 protein expression was significantly higher in the BRCA samples than in the normal breast tissue samples (Fig. 2D). Notably, XLH-36 was found to be significantly upregulated in TNBC patients (Fig. 2E) compared to control subjects and the high XLH-36 expression group exhibited enhanced clinical stage (Supplementary Fig. 2D). Taken together, these results suggest that high XLH-36 expression is correlated with poor clinical outcomes in TNBC patients.



**Fig. 3** XLH-36 functions as an oncomicropeptide to promote TNBC metastasis in vitro and in vivo. **A** Upper: Schematic diagram of the CRISPR/Cas9 based strategy to produce XLH-36-KO MDA-MB-231 and BT549 cells. Lower: Immunoblot of XLH-36 in wild-type and XLH-36-KO MDA-MB-231 and BT549 cells. **B** Cell proliferation was evaluated by the CCK-8 assay in the indicated cells. **C** Flow cytometry analysis of the cell cycle distribution in the indicated cells. **D** qPCR analysis of the cell cycle marker in the WT and XLH-36-KO MDA-MB-231 and BT549 cells. Cell metastatic properties were assessed by the Transwell migration assay (**E**) invasion assay (**F**) and wound scratch assay (**G**). **H** Representative images and statistical analysis of lung metastases of tumor-bearing mice after i.v. injection of the indicated MDA-MB-231 cells. In vivo-generated tumors and analyses of tumor volume (**I**) are shown. Data are presented as mean  $\pm$  SD from  $n = 5$  mice per group. Two-way ANOVA analysis. Lines show the means  $\pm$  SD. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

#### XLH-36 promotes TNBC metastasis in vitro and in vivo

To investigate the potential biological function of XLH-36 in TNBC progression, a set of silencing/overexpression experiments were performed in two TNBC cell lines (MDA-MB-231 and BT549, Fig. 3A and Supplementary Fig. 3A). Results showed that silencing XLH-36 (XLH-36 KO) in MDA-MB-231 and BT549 cells dramatically inhibited (Fig. 3B–G) while overexpressing XLH-36 (XLH-36 OE) significantly promoted cell proliferation (Supplementary Fig. 3B), cell cycle progression (Supplementary Fig. 3C–G), migration, invasion and wound healing (Supplementary Fig. 4A–C). These results suggest that XLH-36 functions as an oncomicropeptide to promote TNBC progression.

We next evaluated the tumor-promotive role of XLH-36 in vivo. Injection of XLH-36-KO cells substantially reduced the incidence of pulmonary metastasis (Fig. 3H), and xenograft tumor growth (Fig.

3I and Supplementary Fig. 5A) with decreased cell proliferation (indicated by Ki-67 staining, Supplementary Fig. 5B). Furthermore, hematoxylin and eosin (H&E) staining showed that XLH-36 knockout led to a decrease in the number and size of metastatic foci in the lung (Supplementary Fig. 5C, D). In contrast, XLH-36 overexpression resulted in the opposite phenotypes (Supplementary Fig. 5E–I). These data collectively corroborate the oncogenic role of XLH-36 in TNBC metastasis and show that targeting XLH-36 may elicit anti-tumor effects.

#### XLH-36, but not lncRNA C5orf66-AS1, exerts the specific pro-cancer effect

Moreover, to distinguish the potential role of C5orf66-AS1 from the potential role of XLH-36, XLH-36-ORF (XLH-36-OE), C5orf66-AS1 derived sORF with mutated ATG start codon (XLH-36-OE-ATG-

Mut), and XLH-36-ORF with synonymous nucleotide substitutions of besides the start codon and stop codon (XLH-36-OE-SNS) that generated the same peptide product were separately overexpressed in MDA-MB-231 and BT549 cells (Fig. 4A). Importantly, overexpression of XLH-36-OE or XLH-36-OE-SNS but not XLH-36-OE-ATG-Mut resulted in elevated XLH-36 expression (Fig. 4B), cell proliferation, and cell migration (Fig. 4C–E). Moreover, the capability of cell proliferation and migration suppressed by XLH-36 KO could be restored by overexpression of XLH-36 or XLH-36-OE-SNS, while XLH-36-OE-ATG-Mut failed to rescue these phenotypes (Fig. 4F–I), implying that XLH-36 but not C5orf66-AS1 is the major executor of the pro-cancer functions in TNBC.

### XLH-36 is localized in the endoplasmic reticulum

The subcellular localization of a protein is closely related to its function, to study how XLH-36 contributes to TNBC progression, we analyzed its intracellular distribution. We conducted subcellular fractionation and structure prediction (using the TMHMM and HDOCK servers), and the result showed a prominent distribution in the cytoplasm (Fig. 5A and Supplementary Fig. 6A). Notably, XLH-36 was co-localized with the endoplasmic reticulum but not with mitochondria (Fig. 5B, C and Supplementary Fig. 6B).

### XLH-36 directly binds to Gemin4

To identify potential interactors of XLH-36, we performed a co-IP assay, followed by MS analysis (Fig. 5D). In total, 140 proteins that were pulled down with XLH-36 were identified (Supplementary Fig. 6C). Remarkably, among the three candidate interactors with the top score (Gemin4, EEF2, and ACLY) that were validated by Co-IP, only Gemin4 was confirmed to interact with XLH-36 (Fig. 5E and Supplementary Fig. 6D). IF analysis of in MDA-MB-231 cells showed that XLH-36 indeed interacted with Gemin4 (Fig. 5F).

To further demonstrate the interaction between XLH-36 and Gemin4, we conducted a yeast two-hybrid assay, which supported the direct interaction between XLH-36 and Gemin4 (Fig. 5G). Subsequently, we performed molecular docking simulations (Fig. 5H) using the predicted XLH-36 structure and Gemin4 structure (Supplementary Fig. 7A, D). We found that residues R12, P16, and R17 in XLH-36 were involved in the binding to residues K608, V897, H925, L931, H994, S1037, E1040, I894 and R1033 in Gemin4 by establishing salt bridges and hydrogen bonds (Fig. 5I). We next performed microscale thermophoresis (MST) analysis using chemically synthesized XLH-36 (Supplementary Fig. 7B, C) and purified Gemin4 protein (Supplementary Fig. 7E, F). The  $K_D$  value of XLH-36 bound to Gemin4 was 4.76 E-06 (Fig. 5J), suggesting that XLH-36 has a strong binding affinity to Gemin4. To better illustrate the key binding sites between XLH-36 and Gemin4, we mutated R12, P16, and R17 to Ala, respectively, according to the molecular docking analysis and performed MST assay. Our results showed that the mutant peptides had decreased affinity for Gemin4 compared to the original XLH-36 peptide. The most significant decrease in affinity was observed after mutation at the P16 position (Supplementary Fig. 7G), indicating that this residue plays a crucial role in the interaction between XLH-36 and Gemin4.

### XLH-36 promotes TNBC progression by binding to Gemin4

Existing studies have shown that Gemin4, an essential component of the survival motor neuron (SMN) complex [23], binds to the small nucleus by working together with other proteins in the complex, including motor neuron protein (SMN) and GEMIN2, 3, 5, 6, 7, and 8. This binding is crucial for the structural integrity of small ribonucleoproteins (snRNPs), which are responsible for the assembly of the spliceosome and the activation of downstream pre-mRNA [24]. No relevant reports have been found on the relationship between Gemin4 and the TNBC pathogenesis. Continuing to investigate the role of Gemin4 in triple-negative breast cancer is crucial for gaining a comprehensive understanding of its biological functions.

Analysis of TCGA data revealed that Gemin4 expression was significantly decreased in BRCA tissues compared to normal tissues (Fig. 6A). Furthermore, there was a strong correlation between low Gemin4 levels and advanced clinical stage in BRCA patients (Fig. 6B). Kaplan–Meier survival analysis revealed that patients with higher Gemin4 expression had a lower risk of BRCA-related death (Fig. 6C). Furthermore, we evaluated the function of Gemin4 in TNBC progression, (Fig. 6D) knocking down of Gemin4 significantly promoted cell proliferation, and migration in MDA-MB-231 cells (Fig. 6E, F). Meanwhile, knockdown of Gemin4 can reverse the inhibitory effect of XLH-36 on TNBC cell proliferation and migration (Fig. 6G, H). Based on our above-mentioned finding that XLH-36 could directly interact with Gemin4, we investigated whether XLH-36 regulates Gemin4 or vice versa. Notably, overexpression of XLH-36 did not affect Gemin4 expression (Supplementary Fig. 8A, B), and Gemin4 knockdown similarly had no effect on XLH-36 expression (Supplementary Fig. 8C, D), implying there is no regulatory relationship between XLH-36 and Gemin4. Overall, these data indicate that XLH-36 exerts its oncogenic function by binding to Gemin4, which acts as a new tumor suppressor in TNBC.

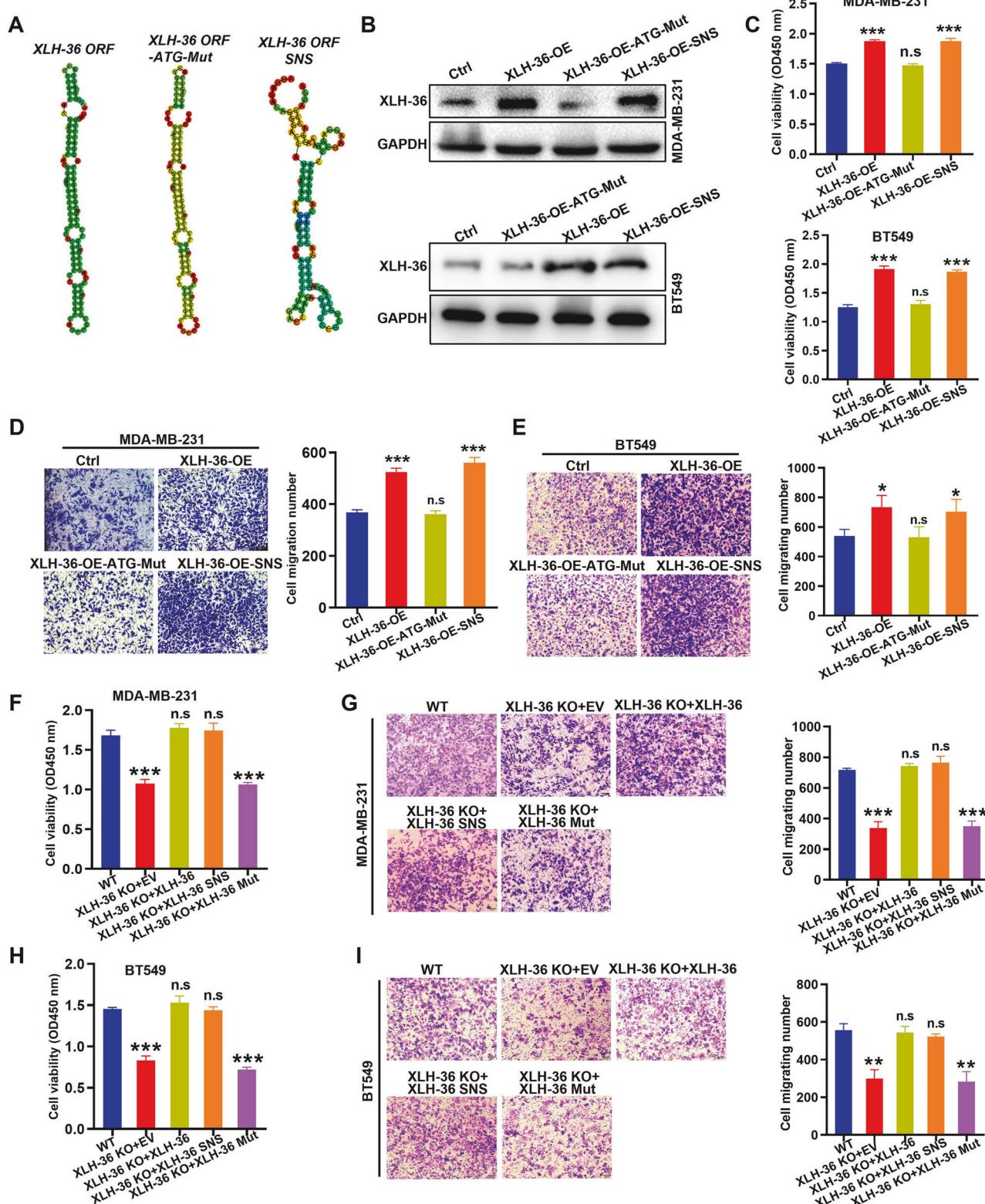
### XLH-36 regulates TNBC metastasis primarily through ICAM1 and S100A4

We next performed RNA-seq analysis in XLH-36-KO cells to identify the downstream pathway through which XLH-36 promotes TNBC metastasis. A total of 640 differentially expressed genes (DEGs, fold change  $\geq 2$  and  $P < 0.05$ ) were identified, including 310 upregulated genes and 330 downregulated genes (Supplementary Fig. 9A). Furthermore, KEGG pathway enrichment and GESA analysis revealed that these DEGs were enriched in ECM receptor interaction, focal adhesion, and cell adhesion molecules (CAMs), which are functionally involved in cell migration and invasion (Fig. 7A and Supplementary Fig. 9B). The top 10 upregulated genes and the top 10 downregulated genes in the RNA-seq (Fig. 7B) results were selected for qRT-PCR validation (Fig. 7C). Among them, the mRNA or protein expression levels of ICAM1 and S100A4 were most significantly regulated by XLH-36 (Fig. 7D and Supplementary Fig. 9C, D).

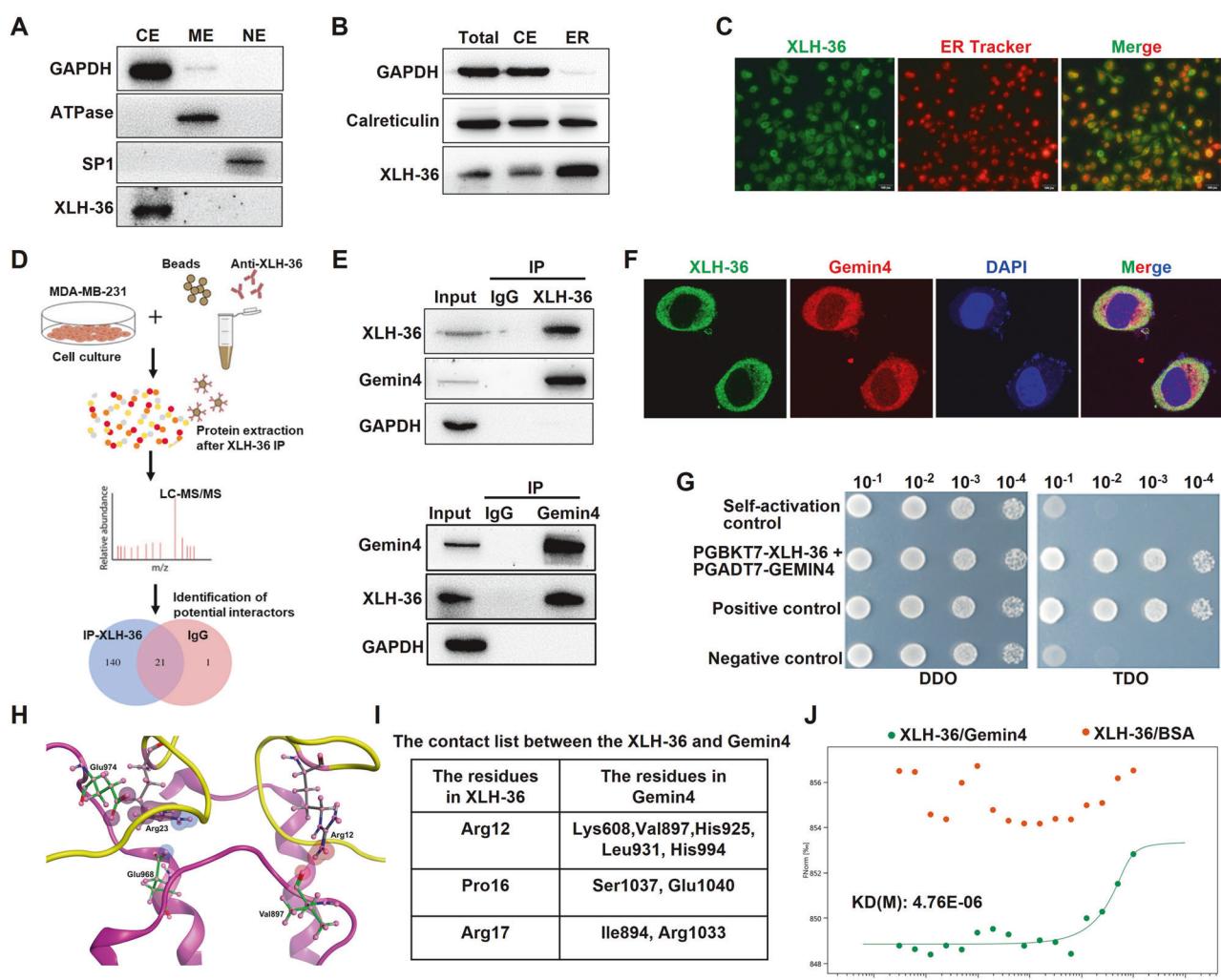
ICAM1 (Intercellular Adhesion Molecule 1), also known as CD54, is a member of the immunoglobulin superfamily (IGSF) of adhesion molecules [25]. It plays a crucial role in regulating various biological processes, including cell adhesion, angiogenesis, tumor evolution, apoptosis, and metastasis [26]. S100A4, also known as fibroblast-specific protein 1 (FSP1) or secreted calcium-binding protein S100A4, is a member of the S100 family of proteins, which plays a critical role in processes such as cell migration, invasion, angiogenesis, and epithelial-mesenchymal transition (EMT) [27, 28]. However, the exact mechanisms by which S100A4 promotes cancer progression are complex and not fully understood. To investigate the specific roles of ICAM1 and S100A4 involved in XLH-36 regulating the progression and metastasis of TNBC, we first conducted clinical database (TCGA and GTEx) analysis and loss-of-function experiments. The results showed that S100A4 expression was remarkably lower in BRCA tissues (Fig. 7E), and the knockdown of S100A4 led to an increase in cell proliferation and migration (Fig. 7F, G). Additionally, the inhibitory effects of silencing XLH-36 were counteracted by the knockdown of S100A4 (Fig. 7H and Supplementary Fig. 9E). Conversely, ICAM1 was highly expressed in BRCA tissues and TNBC samples (Fig. 7I, J). ICAM1 knockdown reduced proliferation and migration (Fig. 7K, L), which were reversed by XLH-36 overexpression (Fig. 7M and Supplementary Fig. 9F). These data show that ICAM1 and S100A4 are key genes regulated by XLH-36 in TNBC development.

### XLH-36/Gemin4 interaction promotes ICAM1 and EMT process via regulating S100A4 mRNA splicing

We next assessed the effect of XLH-36/Gemin4 interaction on the expression of S100A4 and ICAM1, and also investigated the



**Fig. 4** XLH-36 but not C5orf66-AS1 promotes TNBC progression. **A** The minimal free energy of RNA folding of wild-type and mutant (Mut) XLH-36 ORF, as predicted using the RNA-fold web server. **B** Immunoblot analysis of the expression levels of indicated XLH-36 variants in lentivirus-infected MDA-MB-231 and BT549 cells. Cell proliferation (**C**) and migration ability (**D**, **E**) analysis in the MDA-MB-231 and BT549 cells with the indicated constructs. Cell proliferation (**F**) and migration ability (**G**) assay was performed on WT and XLH-36 KO MDA-MB-231 cells with overexpression of XLH-36, XLH-36-OE-SNS, or XLH-36-OE-ATG-Mut. Cell proliferation (**H**) and migration ability (**I**) assay was performed on WT and XLH-36 KO BT549 cells with overexpression of XLH-36, XLH-36-OE-SNS, or XLH-36-OE-ATG-Mut. Data are presented as mean  $\pm$  SD from  $n = 3$  biologically independent experiments. ns non-significant,  $^*P < 0.05$ ,  $^{**}P < 0.01$ ,  $^{***}P < 0.001$ .

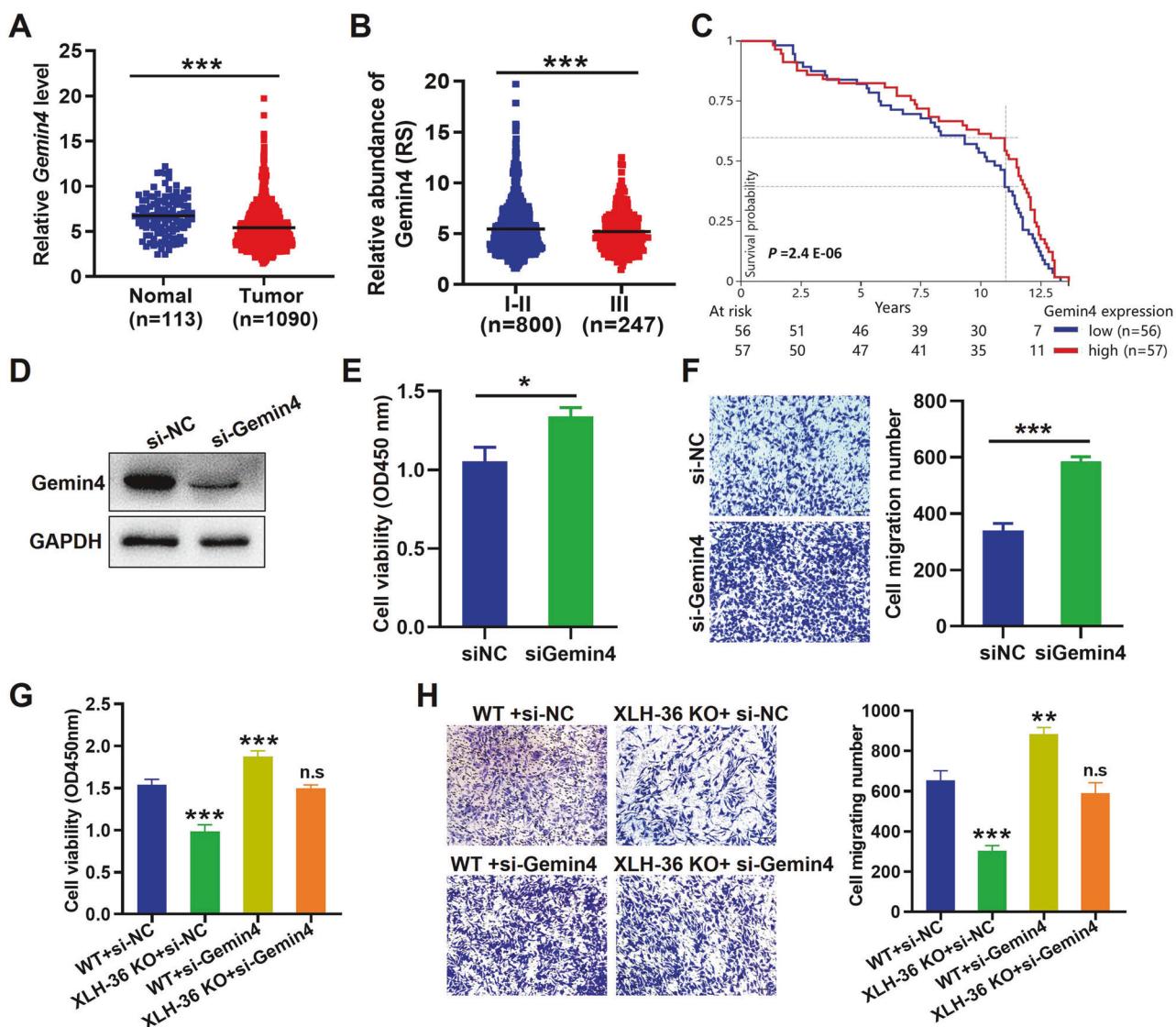


**Fig. 5** XLH-36 is localized in the endoplasmic reticulum and directly binds to Gemin4. **A, B** Immunoblot of XLH-36 in endoplasmic reticulum/ cytosolic fractionated MDA-MB-231 cells. ATPase: cell membrane protein; SP1: nuclear protein; GAPDH: cytosolic protein; Calreticulin: endoplasmic reticulum protein. **C** Immunofluorescence of XLH-36 (Green) colocalized with endoplasmic reticulum (Red) in MDA-MB-231 cells. **D** Schematic diagram showing the identification of XLH-36-interacting proteins using MS analysis following immunoprecipitation. **E** Co-IP assays showing the endogenous interaction between XLH-36 and Gemin4. **F** Representative images of MDA-MB-231 cells co-stained with XLH-36, Gemin4 and DAPI. **G** The direct interaction of Gemin4 and XLH-36 was analyzed by the yeast two-hybrid system. DDO medium, SD/Trp-/Leu-; TDO medium, SD/Trp-/Leu-/HIS-. **H** Molecular docking analysis of XLH-36 with Gemin4. **I** Tables showing the potential contact sites of XLH-36 binding with Gemin4. **J** MicroScale Thermophoresis (MST) analysis of the interaction between XLH-36 and Gemin4. Dose-response analysis of the MST traces allows for the determination of the XLH-36 KD values for its interactions with Gemin4.

regulatory relationship between ICAM1 and S100A4. The results indicated that knockdown of Gemin4 only decreased the levels of S100A4 and did not have any impact on the levels of ICAM1 (Fig. 8A and Supplementary Fig. 10A). Meanwhile, S100A4 knockdown promoted ICAM1 expression and did not affect XLH-36 or Gemin4 expression (Fig. 8B and Supplementary Fig. 10B), and the expression levels of XLH-36, Gemin4, or S100A4 were not affected upon ICAM1 knockdown (Fig. 8C and Supplementary Fig. 10C). Gemin4 is a protein that plays an important role in the process of pre-mRNA splicing, which is crucial for the proper maturation of mRNA. We therefore postulate that Gemin4 plays a role in the splicing process of S100A4 mRNA, potentially contributing to its stability. Wild-type cells (WT) and Gemin4 KD cells were treated with  $\alpha$ -amanitin to stop transcription by RNA polymerase II and RNA was isolated at 0, 2, 4, 6 and 8 h post-treatment. Then, the decay of S100A4 transcripts was measured using real-time polymerase chain reaction (RT-PCR). Our results showed that the S100A4 RNA levels was significantly decreased in Gemin4 knockdown cells (Fig. 8D, E). Moreover, cycloheximide was added to the

cell medium to block the synthesis of new proteins. Endogenous S100A4 protein levels were almost completely suppressed in Gemin4 knockdown cells from 0 h to 6 h time points (Fig. 8F). This phenomenon could be explained by the fact that S100A4 mRNA levels were significantly suppressed by the Gemin4 knockdown.

The SMN complex, including Gemin4, localizes primarily to the cytoplasm but shuttles to the nucleus, where it is involved in the biogenesis and maintenance of spliceosomal snRNPs for splicing [29]. As a micropeptide located in the cytoplasm, we hypothesize that XLH-36 binds to Gemin4 and prevents its transport into the nucleus, thereby inhibiting the transcription and translation of S100A4. Immunofluorescence results indicated that the absence of XLH-36 decreased the presence of Gemin4 in the cytoplasm, but increased its concentration in the nucleus (Fig. 8G). Inversely, the overexpression of XLH-36 shows an opposite trend (Supplementary Fig. 10D, E). These data collectively suggested that XLH-36 binds Gemin4 and retains it in the cytoplasm, thereby inhibiting the enhanced effect of Gemin4 on S100A4 mRNA splicing and stability.



**Fig. 6 Gemin4 serves as a tumor suppressor gene and is involved in XLH-36 regulating the progression of TNBC.** **A** Gemin4 is significantly downregulated in BRCA patients ( $n = 1090$ ) versus adjacent normal tissues ( $n = 113$ ) from TCGA. **B** Gemin4 is significantly related to clinical stage of BRCA patients. **C** Overall survival curves of BRCA patients grouped according to the Gemin4 expression. The overall survival curve was generated with Kaplan-Meier analysis and log-rank test. Gemin4 knockdown **(D)** promotes MDA-MB-231 cell proliferation **(E)** and migration **(F)** ability. Cell proliferation **(G)** and migration assay **(H)** were performed on WT and XLH-36 KO MDA-MB-231 cells with Gemin4 knockdown. Data are presented as mean  $\pm$  SD from  $n = 3$  biologically independent experiments.  $^*P < 0.05$ ,  $^{**}P < 0.01$ ,  $^{***}P < 0.001$ .

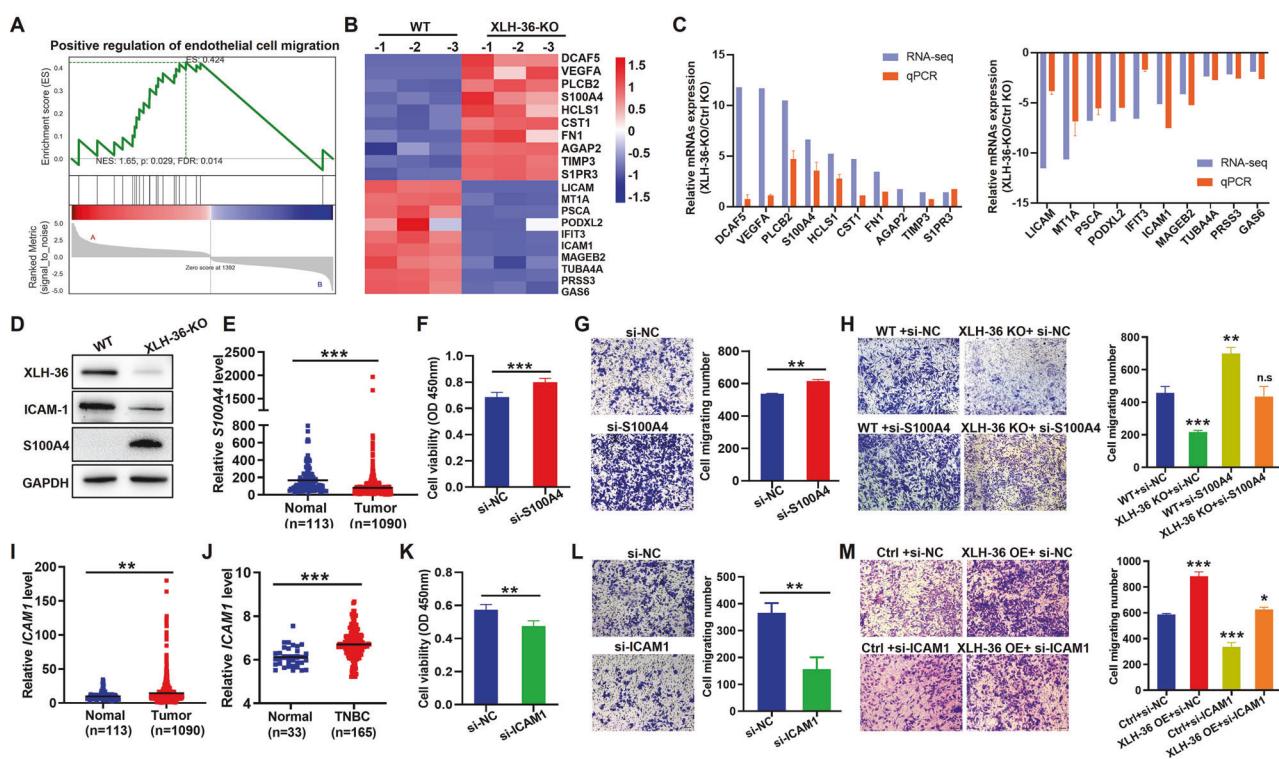
Notably, the regulatory roles of S100A4 and ICAM1 in tumor metastasis are associated with epithelial-mesenchymal transition (EMT) [30, 31]. Based on the results mentioned above, it can be inferred that decreased expression of S100A4 will hinder the impairment of TNBC cell function, leading to a compensatory increase in ICAM1 levels. However, the specific mechanism behind this relationship requires further investigation. We further explored the regulatory relationship between XLH-36, Gemin4, S100A4, and ICAM-1 and EMT. We found that the expression levels of EMT activator TWIST1 and SNAI1 were significantly reduced in XLH-36-KO or ICAM1-KD cells. In contrast, knockdown of Gemin4 or S100A4 enhanced TWIST1 and SNAI1 expression (Fig. 8H). Additionally, IF experiments in MDA-MB-231 cells and xenograft tumor tissues showed that silencing of XLH-36 inhibited TWIST1 expression (Fig. 8I) and XLH-36 overexpression promoted TWIST1 expression (Supplementary Fig. 10F, G). Taken together, these data demonstrate that XLH-36 is a high-affinity Gemin4 binding

micropeptide that enhances EMT by regulating ICAM1 and S100A4 expression.

## DISCUSSION

Here, we present evidence that the previously uncharacterized C5orf66-AS1 encodes the novel micropeptide XLH-36, which mediates TNBC progression and metastasis. Specifically, as an oncoprotein, XLH-36 directly binds Gemin4 to inactivate S100A4 and induce ICAM1 expression, consequently promoting EMT and cancer metastasis (Fig. 9). Together, these data provide new insight into the molecular regulatory network underlying EMT and metastasis in TNBC.

In this study, we found that C5orf66-AS1 encodes a novel endogenous micropeptide XLH-36. Conservation analysis across 101 species showed that micropeptide XLH-36 is highly conserved in different primates, suggesting that XLH-36 may be necessary for



**Fig. 7 XLH-36 regulates TNBC metastasis primarily through ICAM1 and S100A4. A** GSEA of cell migration signature between XLH-36-KO versus WT in MDA-MB-231 cells. **B** Heatmap of the top 10 up/down differentially expressed genes between XLH-36-KO versus WT in MDA-MB-231 cells. **C** Left, qRT-PCR validation of top 10 differentially upregulated genes between XLH-36-KO versus WT in MDA-MB-231 cells. Right, qRT-PCR validation of top 10 differentially downregulated genes between XLH-36-KO versus WT in MDA-MB-231 cells. **D** IB of the expression level of S100A4 and ICAM1 in WT versus XLH-36-KO MDA-MB-231 cells. **E** S100A4 is significantly downregulated in BRCA patients ( $n = 1090$ ) versus adjacent normal tissues ( $n = 113$ ) from TCGA. Cell proliferation (**F**) and migration capacity (**G**) are evaluated in MDA-MB-231 cells transfected with si-NC or si-S100A4. Cell migration assays (**H**) were performed on WT or XLH-36 KO MDA-MB-231 cells with S100A4 knockdown. **I** ICAM1 is significantly upregulated in BRCA patients ( $n = 1090$ ) versus adjacent normal tissues ( $n = 113$ ) from TCGA. **J** Data set from GSE76250 indicates ICAM1 is significantly upregulated in TNBC patients ( $n = 165$ ) versus adjacent normal tissues ( $n = 33$ ). Cell proliferation (**K**) and migration capacity (**L**) are evaluated in MDA-MB-231 cells transfected with si-NC or si-ICAM1. Cell migration assays (**M**) were performed on ctrl or XLH-36 overexpression MDA-MB-231 cells with ICAM1 knockdown. Data are presented as mean  $\pm$  SD from  $n = 3$  biologically independent experiments. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

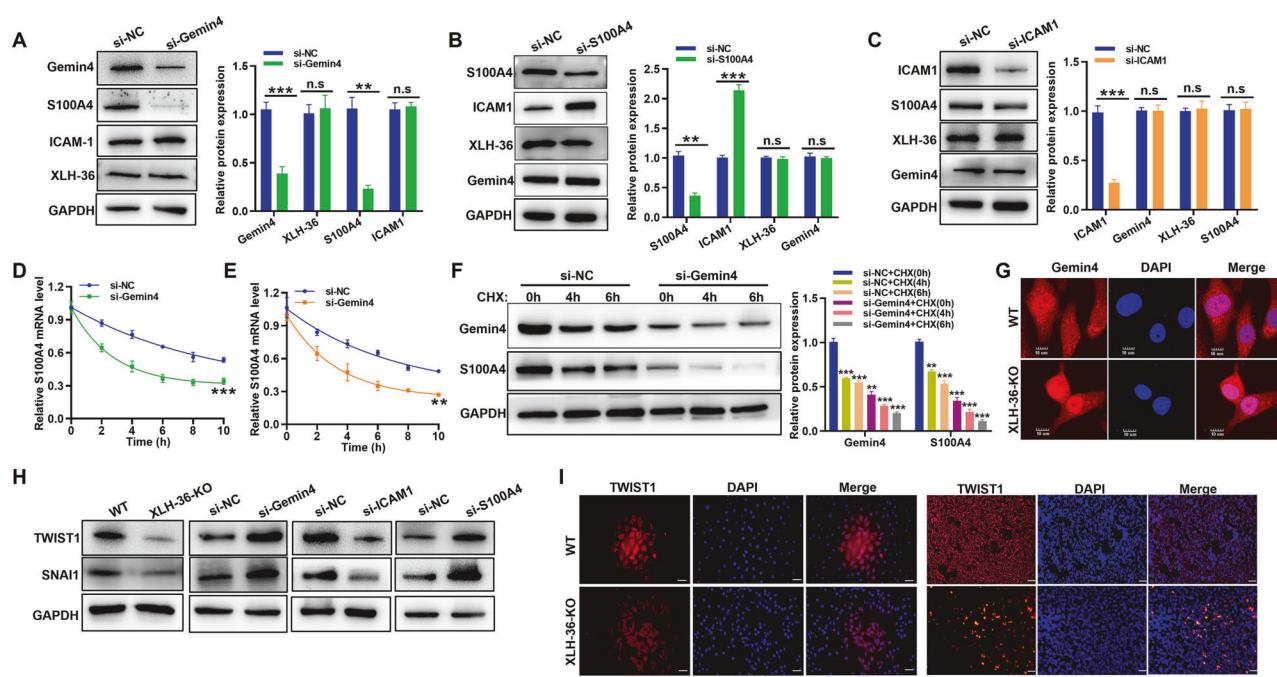
biological processes. Furthermore, through the analysis of 1295 clinical BRCA and TNBC samples from different sources, we proved that XLH-36 is highly expressed in BRCA and TNBC tissues. Importantly, the expression of XLH-36 in early patients was significantly lower than that in advanced-stage patients. In addition, RNA-seq analysis of 32 different cancer types from TCGA cohorts showed that XLH-36 behaves differently in various tumors (Supplementary Fig. 11), including Lung squamous cell carcinoma (LUSC), Head and Neck squamous cell carcinoma (HNSC), Bladder Urothelial Carcinoma (BLCA), Colon adenocarcinoma (COAD), Sarcoma (SARC), Stomach adenocarcinoma (STAD), and Glioblastoma multiforme (GBM), and further research is needed.

Recent studies have demonstrated that C5orf66-AS1, which is annotated as "non-coding RNA", has distinct biological functions in multiple cancers, including pituitary null cell adenoma, cervical cancer, and gastric cancer [32–34]. In this study, we showed that the micropeptide XLH-36, but not the lncRNA C5orf66-AS1, exerts specific pro-cancer effects in TNBC. Whether XLH-36 exerts similar effects in other tumors remains to be determined.

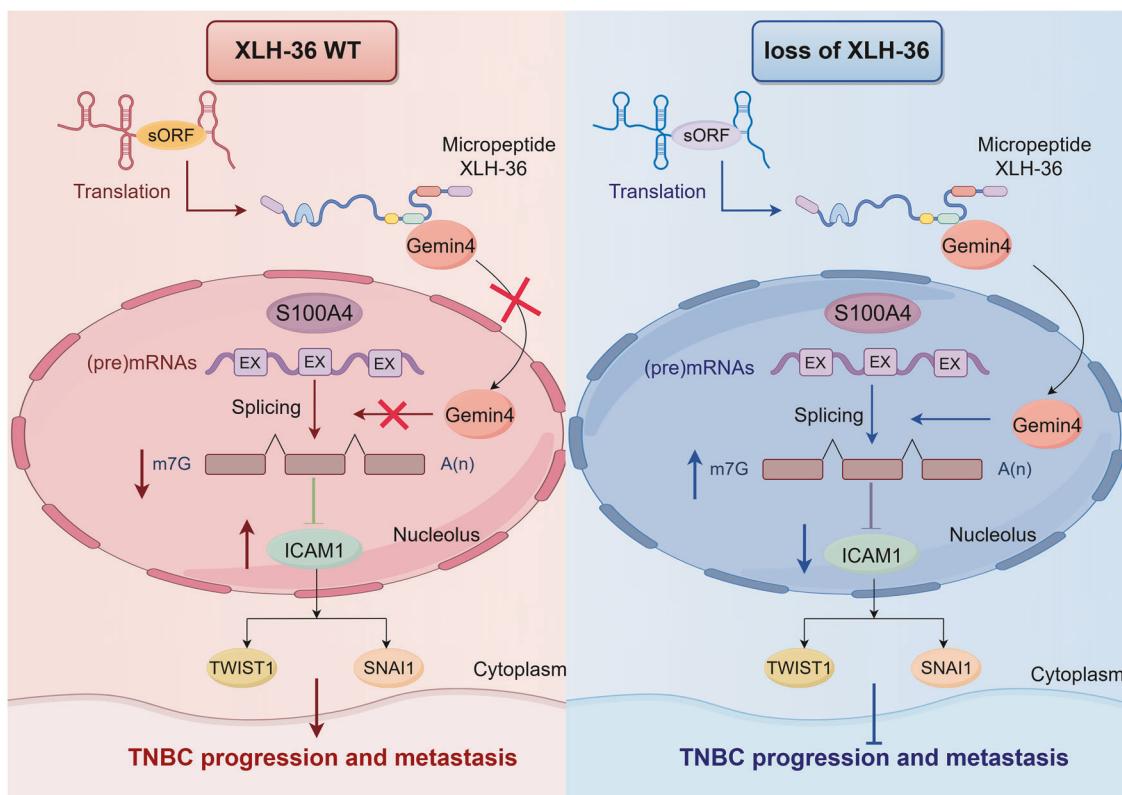
Moreover, TNBC is an exceptionally heterogeneous disease, and besides conventional chemotherapy, treatment options are limited [35]. Although some oncogenes (such as MMPs [36], CDK4/6 [37], SET [38], and mutant BRCA1 [39]) have been found in TNBC, few micropeptides have been found to play proto-oncogenic roles. Recent advances in novel immunotherapeutic approaches show promise in TNBC [40]. In the KEYNOTE-355 study [41], treatment of PD-L1-positive metastatic TNBC with immune

checkpoint inhibitors (ICIs) combined with chemotherapy reduced the risk of mortality by 9% at a 50-month follow-up. Here, we demonstrated that TNBC patients with low XLH-36 expression exhibited a 20% higher overall survival rate than those with high XLH-36 expression (85% vs. 65%) at the same follow-up time, supporting that XLH-36 may serve as a prognostic biomarker. In the future, further investigations will be needed to increase the number of clinical samples and use liquid biopsies to determine the value of XLH-36 in the diagnosis and prognosis of TNBC.

Recent studies report that Gemin4 gene polymorphism is associated with colon cancer, bladder cancer, liver cancer, and kidney cancer progression [42, 43], and high Gemin4 expression levels are associated with poor prognosis in Basal-like breast cancer (BLBC) patients [44], whereas its specific role in TNBC tumorigenesis is still unclear. In the present study, we showed that downregulation of Gemin4 expression is associated with higher overall survival rates in BRCA patients and Gemin4 can inhibit the proliferation and migration of TNBC cells. Furthermore, our Co-IP experiment, yeast two-hybrid analysis, molecular dynamics simulations, and affinity assay showed that XLH-36 directly binds to Gemin4. Of note, XLH-36 and Gemin4 interact, but do not affect each other's expression in TNBC cells. All binding sites between the two molecules could be accurately determined in the future using other analytical methods like cryo-EM or co-crystallization to accurately determine all the sites of binding between the two molecules in the future. Here, we document the critical role of Gemin4 in TNBC, and our results suggest that Gemin4 might serve



**Fig. 8 XLH-36/Gemin4 interaction promoted ICAM1 and EMT process via regulating S100A4 mRNA splicing.** IB of the expression level of XLH-36, Gemin4, S100A4, and ICAM1 in MDA-MB-231 cells transfected with si-NC, si-Gemin4 (A) si-S100A4 (B) or si-ICAM1(C). D, E The decay rates of total mRNAs were assessed using quantitative RT-PCR in si-NC and si-Gemin4 cells following transcription inhibition by  $\alpha$ -amanitin experiments. F Indicated MDA-MB-231 cells were treated with CHX (100  $\mu$ g/mL), and Gemin4 or S100A4 proteins were examined at indicated time point. G Immunofluorescence staining was performed to Gemin4 (red) in WT and XLH-36-KO MDA-MB-231 cells. Scale bar: 10  $\mu$ m. H IB of the expression level of TWIST1 and SNAI1 in the indicated MDA-MB-231 cells. I left, representative IF images of TWIST1 in the WT and XLH-36-KO MDA-MB-231 cells. Right, representative IF images of TWIST1 in the frozen tumor section from the mice subcutaneously injected with WT and XLH-36-KO MDA-MB-231 cells. Data are presented as mean  $\pm$  SD from  $n = 3$  biologically independent experiments. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .



**Fig. 9** The proposed model for the role of C5orf66-AS1-encoded XLH-36 in TNBC metastasis.

as a target for TNBC treatment, but this needs to be validated using more clinical samples.

Gemin4 specifically interacts with a DEAD-box protein and several core proteins of the spliceosome, indicating that it is directly involved in the assembly and regeneration of spliceosomes required for pre-mRNA splicing within the nucleus. S100A4 is a member of the S100 family of calcium-binding proteins, which has been shown to play an important role in tumor metastasis by regulating adhesion, extracellular matrix remodeling, and cell motility.

Given that XLH-36 colocalized to cytoplasm, directly bound to Gemin4, and inhibited the entry of Gemin4 into the nucleus, thereby preventing its promotion of S100A4 mRNA splicing, ICAM1 is an intercellular adhesion molecule that has been found to be involved in the development of a wide range of tumors through the regulation of cell adhesion, angiogenesis, tumor evolution, apoptosis, metastasis, and other biological processes [45]. It is not clear whether there is a regulatory relationship between ICAM1 and S100A4. We observed a significant upregulation of ICAM1 and EMT biomarkers (TWIST1 and SNAI1) in conjunction with a decrease in S100A4. This suggests that TNBC cells have developed compensatory mechanisms to promote survival and metastasis.

In summary, we reveal that the novel translated product micropeptide XLH-36 is significantly upregulated in TNBC, and that XLH-36 expression was significantly correlated with tumor stage and overall survival of patients. XLH-36 may serve as a target for the treatment, diagnosis, and prognosis of TNBC patients. Our study not only represents a new step in elucidating the functions of micropeptides and their translation to the clinic, but also will provide a new protein (peptide) for the human proteome.

## DATA AVAILABILITY

All data generated or analyzed in this study are included in this manuscript or in the supplementary files. Original data supporting the results of this study are available from the corresponding author upon reasonable request. The datasets analyzed during the current study are available in the GEO repository (<https://www.ncbi.nlm.nih.gov/geo>), GEO accession number: GSE76250), TCGA repository (<https://www.cancer.gov/tcga>).

## REFERENCES

- Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024;74:229–63.
- Yin L, Duan JJ, Bian XW, Yu SC. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res.* 2020;22:61.
- Howard FM, Olopade OI. Epidemiology of triple-negative breast cancer: a review. *Cancer J.* 2021;27:8–16.
- American Cancer Society. Survival rates for breast cancer. 2024. <https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html>.
- da Silva JL, Nunes NCC, Izetti P, de Mesquita GG, de Melo AC. Triple negative breast cancer: a thorough review of biomarkers. *Crit Rev Oncol Hematol.* 2020;145:102855.
- Jiang YZ, Liu Y, Xiao Y, Hu X, Jiang L, Zuo WJ, et al. Molecular subtyping and genomic profiling expand precision medicine in refractory metastatic triple-negative breast cancer: the FUTURE trial. *Cell Res.* 2021;31:178–86.
- Guo B, Wu S, Zhu X, Zhang L, Deng J, Li F, et al. Micropeptide CIP2A-BP encoded by LINC00665 inhibits triple-negative breast cancer progression. *EMBO J.* 2020;39:e102190.
- Wang Y, Wu S, Zhu X, Zhang L, Deng J, Li F, et al. LncRNA-encoded polypeptide ASRPS inhibits triple-negative breast cancer angiogenesis. *J Exp Med.* 2020;217:jem.20190950.
- Wu S, Guo B, Zhang L, Zhu X, Zhao P, Deng J, et al. A micropeptide XBP1SBM encoded by lncRNA promotes angiogenesis and metastasis of TNBC via XBP1s pathway. *Oncogene.* 2022;41:2163–72.
- Wright BW, Yi Z, Weissman JS, Chen J. The dark proteome: translation from noncanonical open reading frames. *Trends Cell Biol.* 2022;32:243–58.
- Chen J, Brunner AD, Cogan JZ, Nuñez JK, Fields AP, Adamson B, et al. Pervasive functional translation of noncanonical human open reading frames. *Science.* 2020;367:1140–6.
- Setrerrahmane S, Li M, Zogbhi A, Lv X, Zhang S, Zhao W, et al. Cancer-related micropeptides encoded by ncRNAs: promising drug targets and prognostic biomarkers. *Cancer Lett.* 2022;547:215723.
- Huang N, Li F, Zhang M, Zhou H, Chen Z, Ma X, et al. An upstream open reading frame in phosphatase and tensin homolog encodes a circuit breaker of lactate metabolism. *Cell Metab.* 2021;33:128–144.e9.
- Sun L, Wang W, Han C, Huang W, Sun Y, Fang K, et al. The oncomicropeptide APPLE promotes hematopoietic malignancy by enhancing translation initiation. *Mol Cell.* 2021;81:4493–508.e9.
- Zhang Q, Wei T, Yan L, Zhu S, Jin W, Bai Y, et al. Hypoxia-responsive lncRNA ac115619 encodes a micropeptide that suppresses M6A modifications and hepatocellular carcinoma progression. *Cancer Res.* 2023;83:2496–512.
- Morgado-Palacin L, Brown JA, Martinez TF, Garcia-Pedrero JM, Forouhar F, Quinn SA, et al. The TINCR ubiquitin-like microprotein is a tumor suppressor in squamous cell carcinoma. *Nat Commun.* 2023;14:1328.
- Li M, Li X, Zhang Y, Wu H, Zhou H, Ding X, et al. Micropeptide MIAC inhibits HNSCC progression by interacting with aquaporin 2. *J Am Chem Soc.* 2020;142:6708–16.
- Li M, Liu G, Jin X, Guo H, Setrerrahmane S, Xu X, et al. Micropeptide MIAC inhibits tumor progression by interacting with AQP2 and inhibiting EREG/EGFR signaling in renal cell carcinoma. *Mol Cancer.* 2022;21:181.
- Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. *Nat Protoc.* 2017;12:255–78.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids Res.* 2015;43:e47.
- Choi SW, Kim HW, Nam JW. The small peptide world in long noncoding RNAs. *Brief Bioinform.* 2019;20:1853–64.
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 2012;22:1775–89.
- Park JW, Voss PG, Grabski S, Wang JL, Patterson RJ. Association of galectin-1 and galectin-3 with Gemin4 in complexes containing the SMN protein. *Nucleic Acids Res.* 2001;29:3595–602.
- Rogalska ME, Vivori C, Valcárcel J. Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. *Nat Rev Genet.* 2023;24:251–69.
- Haydinger CD, Ashander LM, Tan ACR, Smith JR. Intercellular adhesion molecule 1: more than a leukocyte adhesion molecule. *Biology.* 2023;12:743.
- Zhou Q, Xu J, Xu Y, Sun S, Chen J. Role of ICAM1 in tumor immunity and prognosis of triple-negative breast cancer. *Front Immunol.* 2023;14:1176647.
- Allgöwer C, Kretz AL, von Karstedt S, Wittau M, Henne-Bruns D, Lemke J, Friend or Foe: S100 proteins in cancer. *Cancers.* 2020;12:2037.
- Zhou X, Zhao J, Yan T, Ye D, Wang Y, Zhou B, et al. ANXA9 facilitates S100A4 and promotes breast cancer progression through modulating STAT3 pathway. *Cell Death Dis.* 2024;15:260.
- Meier ID, Walker MP, Matera AG. Gemin4 is an essential gene in mice, and its overexpression in human cells causes relocalization of the SMN complex to the nucleoplasm. *Biol Open.* 2018;7:bio32409.
- Chow KH, Park HJ, George J, Yamamoto K, Gallup AD, Gruber JH, et al. S100A4 is a biomarker and regulator of glioma stem cells that is critical for mesenchymal transition in glioblastoma. *Cancer Res.* 2017;77:5360–73.
- Chen M, Wu C, Fu Z, Liu S. ICAM1 promotes bone metastasis via integrin-mediated TGF-β/EMT signaling in triple-negative breast cancer. *Cancer Sci.* 2022;113:3751–65.
- Yu G, Li C, Xie W, Wang Z, Gao H, Cao L, et al. Long non-coding RNA C5orf66-AS1 is downregulated in pituitary null cell adenomas and is associated with their invasiveness. *Oncol Rep.* 2017;38:1140–8.
- Rui X, Xu Y, Jiang X, Ye W, Huang Y, Jiang J. Long non-coding RNA C5orf66-AS1 promotes cell proliferation in cervical cancer by targeting miR-637/RING1 axis. *Cell Death Dis.* 2018;9:1175.
- Zhou Q, Li H, Jing J, Yuan Y, Sun L. Evaluation of C5orf66-AS1 as a potential biomarker for predicting early gastric cancer and its role in gastric carcinogenesis. *Oncot Targets Ther.* 2020;13:2795–805.
- Valencia GA, Rioja P, Morante Z, Ruiz R, Fuentes H, Castaneda CA, et al. Immunotherapy in triple-negative breast cancer: a literature review and new advances. *World J Clin Oncol.* 2022;13:219–36.
- Wu W, Warner M, Wang L, He WW, Zhao R, Guan X, et al. Drivers and suppressors of triple-negative breast cancer. *Proc Natl Acad Sci USA.* 2021;118:e2104162118.
- Saleh L, Wilson C, Holen I. CDK4/6 inhibitors: a potential therapeutic approach for triple-negative breast cancer. *MedComm.* 2021;2:514–30.

38. Liu CY, Huang TT, Chen YT, Chen JL, Chu PY, Huang CT, et al. Targeting SET to restore PP2A activity disrupts an oncogenic CIP2A-feedforward loop and impairs triple-negative breast cancer progression. *EBioMedicine*. 2019;40:263–75.
39. Chen H, Wu J, Zhang Z, Tang Y, Li X, Liu S, et al. Association between brca status and triple-negative breast cancer: a meta-analysis. *Front Pharmacol*. 2018;9:909.
40. Abdou Y, Goudarzi A, Yu JX, Upadhyaya S, Vincent B, Carey LA. Immunotherapy in triple-negative breast cancer: beyond checkpoint inhibitors. *NPJ Breast Cancer*. 2022;8:121.
41. Schmid P, Cortes J, Pusztai L, McArthur H, Kümmel S, Bergh J, et al. Pembrolizumab for early triple-negative breast cancer. *N Engl J Med*. 2020;382:810–21.
42. Cieślak A, Galita G, Mik M, Dziki Ł, Dziki A, Sokołowski I, et al. Association of GEMIN4 gene polymorphisms with the risk of colorectal cancer in the Polish population. *Pol Prz Chir*. 2021;93:40–45.
43. Wu N, Zhang X, Tian J, Yu S, Qiao Y. Association of GEMIN4 gene polymorphism and the risk of cancer: a meta-analysis. *Oncotargets Ther*. 2017;10:5263–71.
44. Wu L, Zhang Y, Zheng C, Zhao F, Lin Y. GEMIN4, a potential therapeutic target for patients with basal-like subtype breast cancer. *BMC Women's Health*. 2023;23:396.
45. Taftaf R, Liu X, Singh S, Jia Y, Dashzeveg NK, Hoffmann AD, et al. ICAM1 initiates CTC cluster formation and trans-endothelial migration in lung metastasis of breast cancer. *Nat Commun*. 2021;12:4867.

## ACKNOWLEDGEMENTS

We thank the TCGA and GEO research network for providing us access to genomic and transcriptomic data. Figure 9 was created by Figdraw. This work is mainly funded by the University-Industry Collaboration Program (8210040001, China Pharmaceutical University-NANJING ANJI BIOTECHNOLOGY). The mechanism studies for this work were supported by the National Natural Science Foundation of China (82373039, 82103118), the National Key R&D Program of China (2021YFA0911700), the 2024 Multi-target Natural Medicine Full-weight Research Project (SKLMZZ2024JS08), and the Natural Science Foundation of Jiangsu Province of China (BK20210418).

## AUTHOR CONTRIBUTIONS

HMX supervised the research, HMX and MWL designed the study, wrote the manuscript and coordinated the work. HMX, MWL, and LX discovered XLH-36. WWX, KXH and TTL performed cellular experiments, animal experiments, RNA-sequencing, and Western blot assay. WWX, XL, and MWL contributed to bioinformatic analyses. MWL and TTL established mouse metastasis models and histopathology analyses. XWX contributed to the expression and purification of the Gemin4 protein. WWX and SL provided technical support. WWX and JHC revised the manuscript. All authors read and approved the final manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ETHICAL DECLARATIONS

All research involving human participants, human materials, or human data was conducted in accordance with the Declaration of Helsinki. The study protocol was reviewed and approved by the Ethics Committee of Hunan Aifang Biotechnology Co., Ltd. (Approval No. HN20250401). Written informed consent was obtained from all participants prior to their inclusion in the study.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41388-025-03621-8>.

**Correspondence** and requests for materials should be addressed to Hanmei Xu.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025