Society of Medical Education & Research

2022, Volume 2, Issue 1, Page No: 65-85

ISSN: 3108-4834

Archive of International Journal of Cancer and Allied Science

The Newly Identified Endogenous Micropeptide XLH-36 Interacts with Gemin4 to Drive Metastasis in Triple-Negative Breast Cancer

Laura Jane Davies[1]*, Helen Elizabeth Brown[1], Michael Eric Dubois[2], Pierre Andre Lefevre[2]

[1]Department of Clinical Oncology, Trinity College Dublin, Dublin, Ireland.
[2]Department of Cancer Immunology, Université Paris-Saclay, Paris, France.

***E-mail** ✉ l.davies.tcd@gmail.com

Abstract

Triple-negative breast cancer (TNBC) is recognized as the most aggressive form of breast malignancy. Although TNBC patients have a 5-year overall survival rate of 77%, the prognosis drops sharply to 12% once metastasis develops. The identification of biomarkers that accurately predict TNBC metastasis remains an urgent unmet need. In this study, we discovered a previously uncharacterized micropeptide, XLH-36, derived from the long non-coding RNA (lncRNA) C5orf66-AS1, which functions as a potent oncogenic factor in TNBC. Comparative sequence analyses across 101 species revealed that XLH-36 is highly conserved in humans and primates. Analysis of RNA-seq datasets from 1295 breast cancer patients, including 165 TNBC cases across multiple cohorts, demonstrated elevated XLH-36 expression in both breast cancer overall and TNBC specifically. Importantly, patients with lower C5orf66-AS1 expression, encoding XLH-36, exhibited a 20% survival advantage over 50 months compared with those with higher expression. Functional experiments showed that loss of XLH-36 significantly reduced tumor growth and metastasis in TNBC xenografts in vivo. Mechanistically, XLH-36 binds to Gemin4, retaining it in the cytoplasm and thereby disrupting its nuclear role in S100A4 mRNA splicing. This interaction triggers a compensatory upregulation of ICAM1, which facilitates epithelial-to-mesenchymal transition (EMT) and promotes metastatic dissemination. These findings establish XLH-36 as a critical regulator of TNBC metastasis and suggest its potential as both a biomarker and therapeutic target.

Keywords: XLH-36, Gemin4, Breast cancer, Endogenous micropeptide

Introduction

Breast cancer remains the most common malignancy among women, accounting for 23.8% of all female cancer cases worldwide [1]. TNBC represents 15–20% of these cases and is characterized by the absence of HER2, estrogen, and progesterone receptors, limiting treatment options primarily to non-specific chemotherapy [2, 3]. Although the general 5-year survival rate for TNBC is 77%, metastatic progression reduces survival to a mere 12% [4]. While certain molecules, such as EGFR, VEGF, PARP1, and C-kit, have been associated with TNBC metastasis [5, 6], effective targeted therapies remain limited. Recently, micropeptides including CIP2A-BP, ASRPS, and XBP1SBM were shown to influence TNBC progression [7–9], but the functions of many novel micropeptides remain largely uncharacterized. Increasing evidence indicates that small open reading frames (sORFs) within previously annotated non-coding regions—including lncRNAs, circRNAs, pri-miRNAs, and untranslated regions (5'-UTR and 3'-UTR)—can encode biologically active micropeptides [10–12]. Such micropeptides have been implicated as proto-oncogenes or oncogenes, influencing proliferation, metastasis, angiogenesis, metabolism, and therapy resistance in various cancers [13–16].

Building upon our previously established micropeptide discovery platform integrating Ribo-seq, mass spectrometry, machine learning, and gene editing, we

Access this article online https://smerpub.com/

Received: 12 November 2021; **Accepted:** 28 January 2022

How to cite this article: Davies LJ, Brown HI, Dubois ME, Lefevre PA. The Newly Identified Endogenous Micropeptide XLH-36 Interacts with Gemin4 to Drive Metastasis in Triple-Negative Breast Cancer. Arch Int J Cancer Allied Sci. 2022;2(1):65-85. https://doi.org/10.51847/k5K4oi1ShJ

identified MIAC, encoded by lncRNA AC025154.2, as a key regulator in head and neck squamous cell carcinoma and kidney cancer [17, 18]. Using the same approach, we identified a novel 36-amino-acid micropeptide, XLH-36, derived from lncRNA C5orf66-AS1. In this study, we interrogated XLH-36's binding partners and molecular mechanisms, assessed its impact on TNBC cell behavior, and explored its potential as a biomarker and therapeutic target.

Materials and Methods

Animal models

Female BALB/C nude mice were sourced from Jiangsu Gempharmatech and maintained under specific pathogen-free (SPF) conditions at the Pharmaceutical Animal Center, China Pharmaceutical University.

Tumor xenograft experiments

For subcutaneous xenografts, five-week-old female nude mice were injected with 6.5×10^6 MDA-MB-231 cells suspended in 100 µl medium into the left para-mammary region. Tumor dimensions were recorded every three days, and volume was calculated as (length × width²)/2. Mice were euthanized when necessary, and tumors were harvested for downstream analyses.

For metastasis studies, 3×10^6 MDA-MB-231 cells with XLH-36 overexpression or knockout were delivered via tail vein injection. Tumor dissemination was monitored biweekly for four weeks using the IVIS Spectrum live imaging system (PerkinElmer). All procedures adhered to NIH and institutional guidelines and were approved by the Institutional Animal Care and Use Committee (IACUC-202312016).

Cell culture

MCF10A, BT474, MCF7, MDA-MB-231, MDA-MB-468, and 293T cells were cultured in DMEM (GIBCO) supplemented with 10% FBS (Bioind), while T47D cells were maintained in RPMI 1640 (GIBCO) with 10% FBS. Cultures were incubated at 37°C in a 5% CO_2 atmosphere.

RNA extraction and quantification

Total RNA was isolated using TRIzol (TIANGEN Biotech #G1012). Two micrograms of RNA were reverse-transcribed into cDNA using 5X All-In-One MasterMix (with AccuRT Genomic DNA Removal Kit, Applied Biological Materials #G492). Gene expression was quantified by qRT-PCR using Blastaq 2X qPCR MasterMix (Applied Biological Materials #G891), with GAPDH as the reference gene. Relative expression was calculated using the $2^{-\Delta\Delta CT}$ method.

Transfection and lentiviral infection

siRNAs targeting ICAM1, S100A4, and XLH-36, along with non-targeting controls, were obtained and delivered using Lipofectamine 8000 (Beyotime Biotechnology). Lentiviral particles were produced in HEK293T cells with packaging plasmids pMD2.G and psPAX2, and target cells were transduced and selected with puromycin (2 µg/ml).

CRISPR/Cas9-mediated gene editing

XLH-36 knock-in (KI) was achieved by inserting a FLAG epitope at the C-terminal via an sgRNA targeting the stop codon, cloned into pSpCas9(BB)-2A-GFP (PX458, Addgene). A single-stranded oligodeoxynucleotide (ssODN) served as the donor template. Knock-in efficiency was confirmed by PCR, sequencing, immunofluorescence, and IP-MS.

For knockout (KO), two sgRNAs targeting the start and stop codons were cloned into pYSY-U6-sgRNA-EFla-mCherry and pYSY-U6-sgRNA-EF1a-eGFP. Lentiviral Cas9 (Hanbio) was delivered into MDA-MB-231-Cas9 cells, followed by sgRNA co-transfection and FACS sorting. Monoclonal XLH-36-KO lines were validated by sequencing and Western blot.

Generation of XLH-36 monoclonal antibody

The XLH-36 peptide, synthesized by Abmart, served as the immunogen. Groups of 12 Balb/c mice (aged 8–12 weeks) were immunized, and serum titers were periodically monitored to determine the optimal immunization schedule. Using a combination of optimized adjuvants and immunization techniques, high-affinity IgG antibodies targeting the XLH-36 peptide were induced. After an initial immunization, mice received three to four booster injections. Serum reactivity against recombinant XLH-36 protein was assessed via ELISA, with a titer threshold of 10,000; if the threshold was not met, additional boosts were administered. Following successful immunization, spleens and half of the lymph nodes were harvested and fused with SP2/0 myeloma cells. The resulting hybridomas were plated into four 384-well plates at 10^2–10^4 cells per well and cultured. Supernatants were screened by ELISA for reactivity to the XLH-36 peptide, and positive wells with

viable cells were transferred to 96-well plates for expansion. Subsequent rounds of ELISA screening and subcloning continued until monoclonal populations showed 100% reactivity. Positive hybridoma clones were expanded, and ascites were collected, purified using Protein A/G, and used in downstream assays.

Protein extraction and western blotting

Cells were harvested, and total protein was extracted using RIPA lysis buffer (Beyotime Biotechnology) supplemented with Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific). Lysates were centrifuged at 12,000 × g for 10 minutes at 4°C to remove debris. Protein samples were separated on 12% SDS-PAGE gels and transferred to PVDF membranes (Millipore Sigma). Membranes were incubated overnight with primary antibodies, followed by HRP-conjugated goat anti-rabbit or anti-mouse secondary antibodies. GAPDH was used as the loading control. Chemiluminescence detection was performed using ECL reagents, and images were captured with Tanon software (TANON-5200 system, Shanghai, China).

Immunoprecipitation (IP)

Cells were lysed, and 2 µl of primary antibody was added to 1.5 mg of protein solution. For control reactions, 0.4 µl of IgG was incubated with 400 µg lysate. Samples were incubated overnight at 4°C, then 35 µl of protein A/G magnetic beads were added for 3 hours. Beads were washed three times with 500 µl RIPA-containing protease solution at 4°C and 6000 × g for 1 minute each. Bound proteins were eluted and analyzed by Western blot.

Subcellular fractionation

Cytoplasmic, membrane, nuclear, chromatin, and cytoskeletal fractions were isolated using the Subcellular Fractionation Isolation Kit (Thermo Fisher Scientific) according to the manufacturer's protocol.

Endoplasmic reticulum and mitochondrial sub-fractionation

For ER isolation, 2 × 10^8 cells were lysed and resuspended in hypotonic buffer at 4°C for 20 minutes. Following centrifugation at 600 × g for 5 minutes, cells were homogenized and centrifuged at 1000 × g for 25 minutes. Supernatants were treated with $CaCl_2$, stirred, and centrifuged at 8000 × g to obtain ER-enriched precipitates, which were homogenized in isotonic solution.

Mitochondria were isolated from 2 × 10^7 cells using the Thermo Fisher Mitochondrial Isolation Kit. Cells were incubated sequentially with reagents A, B, and C, with intermittent vortexing and incubation on ice. Centrifugation steps at 700 × g, 1200 × g, and 12,000 × g were performed to separate the mitochondrial fraction.

Immunofluorescence and confocal microscopy

Cells were fixed in β-galactosidase staining fixative (Beyotime Biotechnology #C0602) for 15 minutes at room temperature, followed by two PBS washes. Membranes were permeabilized with Triton X-100 (Beyotime Biotechnology #P0096) for 15 minutes, washed, and blocked with 5% BSA for 30 minutes. Primary antibodies were applied and incubated at 4°C overnight. After washing, secondary antibodies were added for 2 hours at room temperature in the dark. Cells were counterstained with DAPI (Servicebio Biotechnology #GDP1024) for 10 minutes and imaged under a fluorescence microscope.

Cell proliferation assay (CCK-8)

Cell growth was assessed using the CCK-8 kit (Yeasen Biotechnology #40203ES76). Cells were seeded at 2,000 cells per well in 96-well plates. At designated time points, 10 µl of CCK-8 reagent was added to each well, and absorbance was measured at 450 nm using a spectrophotometer (Thermo Scientific). All experiments were conducted in triplicate.

Flow cytometry-based cell cycle analysis

Cell cycle distribution was assessed using the Cell Cycle Assay Kit-PI/RNase Staining (CCS012, MULTI SCIENCES (LIANKE) BIOTECH CO., LTD) according to the manufacturer's protocol. Approximately 5 × 10^5 cells were collected and fixed in pre-chilled 70% ethanol, then incubated overnight at –20°C. After centrifugation at 2,000 rpm for 5 minutes, cells were treated with 100 µl RNase solution at 37°C for 30 minutes, followed by 900 µl propidium iodide (PI) staining for another 30 minutes at 37°C. Cell cycle profiles were analyzed using a CytoFLEX flow cytometer (Beckman Coulter) and FlowJo software.

Cell migration and invasion assays

For migration assays, 1.5 × 10^4 cells were seeded into the upper chamber of a transwell system in 200 µl serum-free

DMEM, while 500 µl of DMEM containing 10 percent FBS was placed in the lower chamber. After 48 hours, non-migrated cells on the upper surface were removed, and migrated cells were fixed with precooled formaldehyde for 30 minutes, stained with 0.1 percent crystal violet for 15 minutes, and counted in five random fields (10× magnification).

For invasion assays, OrganoGel Matrigel (CELLada #OM-1) was thawed at 4°C, diluted 1:7 in ice-cold serum-free DMEM, and 10 µl was applied to each transwell chamber, which was incubated at 37°C for 30 minutes. Cells were digested and prepared as a suspension, and 3×10^4 cells were added per chamber. Migration distance was monitored at 0, 12, 24, and 48 hours under an inverted microscope and quantified using ImageJ-win64 software.

Cell adhesion assay

Fibrin was dissolved in sterile PBS at 10 µg/ml and applied to 96-well plates (50 µl/well), incubated at 37°C for 1 hour, and washed twice with serum-free medium. A total of 1×10^4 cells were seeded in serum-free medium and allowed to adhere for 1 hour at 37°C. After three PBS washes, CCK-8 reagent was used to quantify adherent cells.

Wound healing assay

Cells were grown to ~90% confluence in 24-well plates, and a linear scratch was made using a sterile pipette tip. The medium was replaced with DMEM containing 2% FBS, and wound closure was monitored over time.

Bioinformatic analysis

Homologous sequences of C5orf66-AS1 ORF1 were identified via NCBI-BLAST and UCSC Genome Browser BLAT searches (https://genome.ucsc.edu/). XLH-36 structural and domain predictions were performed using 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 XLH-36 was modeled with HDOCK, while the 3D structure of Gemin4 was obtained from AlphaFold (AF-P57678-F1). Docking simulations were performed using the ClusPro server [19], and results were analyzed in MOE 2019.1.

MicroScale thermophoresis (MST)

Recombinant human Gemin4 protein, expressed in HEK293 cells, was fluorescently labeled with CY5 via cysteine sulfhydryl groups (Xi'an Ruixi Biological Technology Co., Ltd). Labeled Gemin4 was incubated with synthetic XLH-36 peptide (Shanghai GL Biochem Ltd) for 5 minutes at room temperature, loaded into premium-coated capillaries, and analyzed using the Monolith NT.115 instrument (NanoTemper, Germany). Dissociation constants (Kd) were calculated using the NanoTemper MO. Affinity analysis tool.

Hematoxylin–eosin (H&E) staining

Mouse tissues were fixed in 4 percent formalin for 48 hours and cryosectioned at 9 µm thickness using Tissue-TEK O.C.T. Compound (SAKURA #4583). Sections were sequentially dehydrated in 70 percent, 85 percent, and 100 percent ethanol, stained with hematoxylin (BBI #E607317-0100) for 3 minutes, rinsed with water for 5 minutes, then stained with eosin (BBI #E607321-0100) for 3 minutes. Sections were dehydrated through graded ethanol and xylene, air-dried, and mounted with neutral balsam (Solarbio #G8590).

LC-MS/MS analysis

Immunoprecipitation complexes from MDA-MB-231 cells using XLH-36 antibody were analyzed by LC-MS/MS as previously described [17]. Proteins were separated on 10–17% Bis-Tris SDS-PAGE gels and silver-stained. Peptides were resuspended in 0.1% TFA with 2% acetonitrile and analyzed using a QExactive mass spectrometer (Thermo Fisher Scientific) coupled with Dionex NCS3500 HPLC. Spectra were searched against the UniProt database, allowing up to two missed cleavages.

RNA sequencing sample and library preparation

MDA-MB-231 WT and XLH-36 KO cells (1×10^6) were washed with cold PBS and collected by centrifugation at 500 × g for 5 minutes at 4°C. Total RNA was extracted using TRIzol (Invitrogen) and quality assessed using Agilent 2100 BioAnalyzer (RIN ≥ 7.0, rRNA ratio ≥ 1.5). Libraries were prepared with the Illumina NEBNext Ultra RNA Library Prep Kit and sequenced on Illumina NextSeq 500 (1100-bp single-end reads). Library construction and QC were performed at CapitalBio Technology following standard protocols. Replicates for RNA-seq experiments are indicated in **Figure 7b**.

Statistical analysis

Experimental results are reported as means ± standard deviation (SD) from at least three independent biological experiments. Data comparisons employed Student's t-test for pairwise groups. For analyses involving more than two groups, unpaired two-tailed Student's t-tests or one-way analysis of variance (ANOVA) were applied. Survival analysis utilized Kaplan-Meier estimates, with differences evaluated via the log-rank test. All statistical computations were performed in GraphPad Prism software (version 8; La Jolla, CA, USA), except where specified otherwise. Significance levels: NS (not significant), *P < 0.05, **P < 0.01, ***P < 0.001.

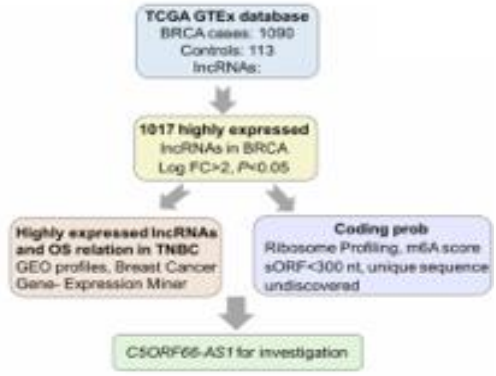
RNA-sequencing data processing

Transcript abundance was quantified as gene-level read counts using Hisat2 alignment. Normalization was carried out to counts per million (CPM), followed by differential expression analysis with LIMMA [20] and DESeq tools, applying thresholds of $P < 0.05$ and $|\log2(\text{fold change})| \geq 1$. Gene Ontology (GO) term enrichment was conducted via PANTHER to identify enriched biological processes and molecular functions among differentially expressed genes (DEGs) in XLH-36-knockout cells. DEG lists comprised genes up- or down-regulated in comparisons (MDA-MB-231 wild-type versus MDA-MB-231/XLH-36-KO), with adjusted $P < 0.05$ and $|\log2(\text{fold change})| \geq 1$. A total of 640 DEGs were detected. Highly enriched GO categories were selected using a false discovery rate (FDR) cutoff of < 0.05. The RNA-seq datasets are archived in the GSA repository under accession HRA008378.

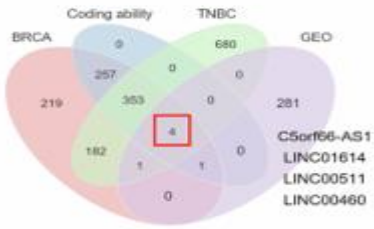
Results and Discussion

Screening for differentially expressed long non-coding RNAs and candidate small open reading frames

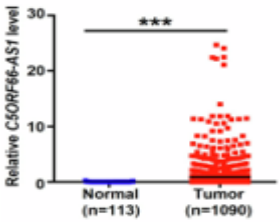
Recent studies indicate that certain long non-coding RNAs (lncRNAs) can encode functional micropeptides through small open reading frames (sORFs) [21], sharing transcriptional and structural similarities with messenger RNAs [22]. To pinpoint lncRNAs with translational potential that are overexpressed in triple-negative breast cancer (TNBC), we examined expression profiles from 1090 breast invasive carcinoma (BRCA) cases and 113 adjacent normal breast samples in the TCGA dataset, supplemented by RNA-seq from 165 TNBC cases in the GEO repository (GSE76250). This yielded 1017 candidate lncRNAs with elevated expression (log(fold change) ≥ 2, P ≤ 0.01), selected based on these criteria: (1) elevated levels in breast cancer overall and specifically in TNBC, alongside significant prognostic associations; (2) strong translational likelihood inferred from ribosome profiling data and m6A modification scores in the Translnc database; and (3) sORF regions shorter than 300 nucleotides encoding novel, unique peptide sequences **(Figure 1a)**. Four lncRNAs satisfied all criteria concurrently (C5ORF66-AS1, LINC01614, LINC00511, LINC00460; **(Figure 1b)**), with C5orf66-AS1 ranking highest in composite scoring **(Table 1 and Figure 1c)**. Additionally, C5orf66-AS1 showed markedly higher expression in TNBC patient specimens **(Figure 1d)** and in TNBC cell lines (MDA-MB-231, MDA-MB-468, BT549; **(Figure 1e)**). Accordingly, this investigation centered on the contributions of this micropeptide-producing lncRNA to TNBC advancement.



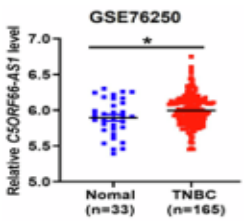
a)



b)



c)



d)

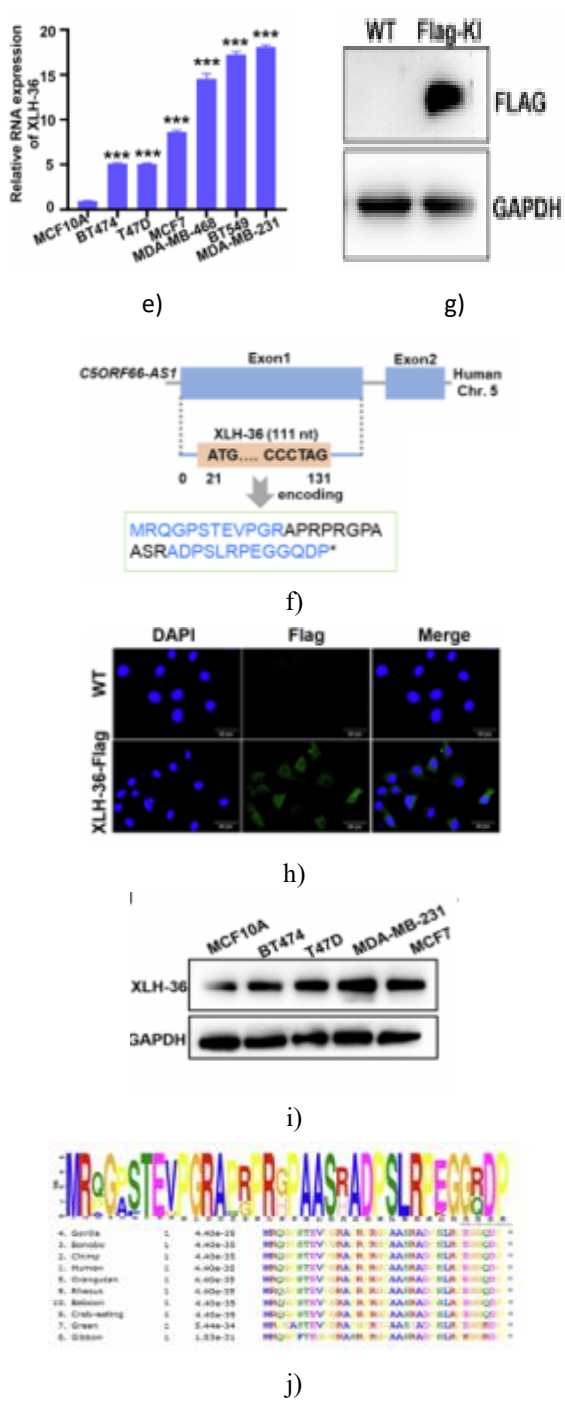


Figure 1. Comprehensive screening reveals a new micropeptide, XLH-36, in triple-negative breast cancer

(a) Overview of the pipeline for identifying potentially translatable long non-coding RNAs with differential expression, integrating data from TCGA, GEO, and bioinformatics resources.

(b) Venn diagram illustrating the intersection of overexpressed lncRNAs across general breast cancer cases, TNBC specimens, and those with predicted translational capacity.

(c) Levels of C5orf66-AS1 (which encodes XLH-36) were markedly elevated in invasive breast carcinoma patients (n = 1090) relative to matched non-tumor tissues (n = 113) based on TCGA data.

(d) Analysis of the GSE76250 cohort showed substantial overexpression of C5orf66-AS1 in TNBC cases (n = 165) versus paired normal samples (n = 33).

(e) Relative expression of the XLH-36-producing lncRNA C5orf66-AS1 across various breast cancer cell lines compared to non-malignant mammary cells, with normalization against GAPDH.

(f) Schematic representation of the XLH-36 coding sequence position within the C5orf66-AS1 transcript.

(g–h) Validation of XLH-36-Flag fusion protein production in transfected MDA-MB-231 cells via Western blotting (g) and immunofluorescence staining (h), where nuclear counterstaining was performed using DAPI (blue signal).

(i) Endogenous XLH-36 protein detection in breast cancer cell lines versus normal breast cells using a custom anti-XLH-36 antibody in immunoblotting.

(j) Sequence conservation analysis of the XLH-36 peptide across multiple vertebrate species.

Results are shown as mean ± SD derived from three independent biological replicates. Significance indicators: *P < 0.05, **P < 0.01, ***P < 0.001.

Table 1. Summary of high-potential coding lncrnas differentially overexpressed in breast cancer and TNBC

Rank	Log₂(Fold Change)	lncRNA Symbol	P-value	Overexpression in TNBC	Association with Overall Survival	sORF Peptide Length	Novel Peptide Sequence	Coding Probability Score

1	6.04	C5ORF66-AS1	P < 0.001	P < 0.001	P < 0.05	36 amino acids	Yes	0.4982
2	5.56	LINC01614	P < 0.001	P < 0.001	P < 0.05	27 amino acids	Yes	0.2824
3	2.96	LINC00511	P < 0.001	P < 0.001	P < 0.05	73 amino acids	Yes	0.6387
4	2.60	LINC00460	P < 0.001	P < 0.001	P < 0.001	18 amino acids	Yes	0.567

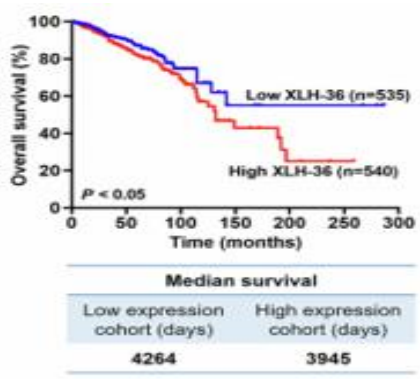
Identification of the conserved novel endogenous micropeptide XLH-36 in primates

To evaluate the coding potential of C5orf66-AS1, we applied PhyloCSF, the Coding Potential Calculator (CPC), and ORF finder, which revealed a sORF within exon 1 (21–131 bp) capable of encoding a 36-amino acid micropeptide, designated XLH-36 **(Figure 1f)**. To verify endogenous expression, a Flag-tag knock-in (Flag KI) was introduced upstream of the XLH-36 stop codon, and both immunoblotting (IB) and immunofluorescence (IF) using anti-Flag confirmed that XLH-36 is indeed translated in cells **(Figures 1f and 1g)**. Additionally, we developed a monoclonal anti-XLH-36 antibody and detected XLH-36 in various BRCA cell lines via IB **(Figure 1h)**. XLH-36-specific peptide sequences were further identified through immunoprecipitation (IP) of Flag-KI-enriched lysates followed by mass spectrometry analysis.

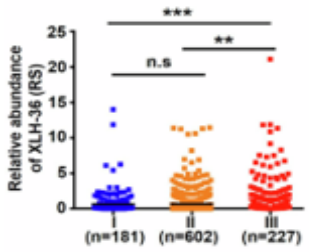
Considering that evolutionary conservation is a hallmark of functional ORFs, we examined XLH-36 codon conservation and found that its amino acid sequence is relatively conserved among primates **(Figure 1i)**, suggesting potential functional significance. Collectively, these findings demonstrate that XLH-36 is naturally and endogenously translated from the sORF within lncRNA C5orf66-AS1 (Chr.5 135,040,047–135,039,596).

XLH-36 as an oncogenic micropeptide and prognostic biomarker in TNBC

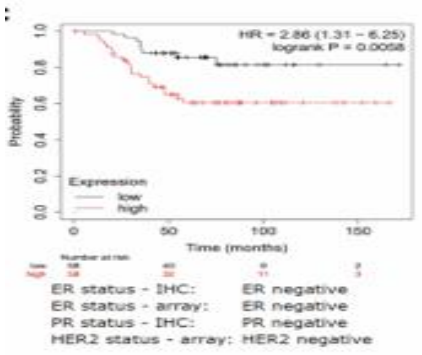
To explore the clinical significance of XLH-36, patients were stratified into XLH-36-low and XLH-36-high groups based on the median expression of C5orf66-AS1. Kaplan–Meier survival analysis revealed that elevated XLH-36 expression correlates with poor prognosis **(Figure 2a)** and is upregulated in advanced-stage BRCA patients **(Figure 2b)**. In a separate TNBC cohort, high XLH-36 expression was also associated with worse survival outcomes **(Figure 2c)**, whereas its expression showed limited correlation with prognosis in ER-, PR-, or HER2-positive patients. Immunohistochemical analysis using the anti-XLH-36 antibody confirmed that protein levels were significantly higher in BRCA tissues compared with normal breast samples **(Figure 2d)**, with TNBC patients displaying particularly elevated XLH-36 expression **(Figure 2e)**. Furthermore, patients with high XLH-36 expression exhibited more advanced clinical stages. These results collectively indicate that XLH-36 overexpression is linked to poor clinical outcomes in TNBC.



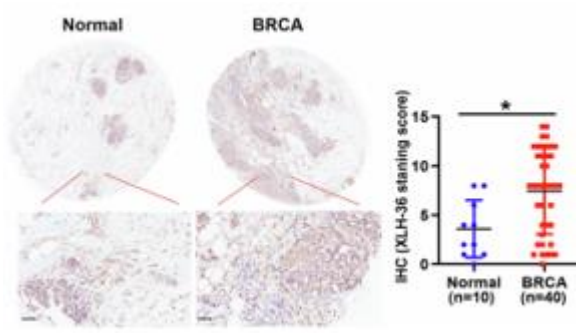
a)



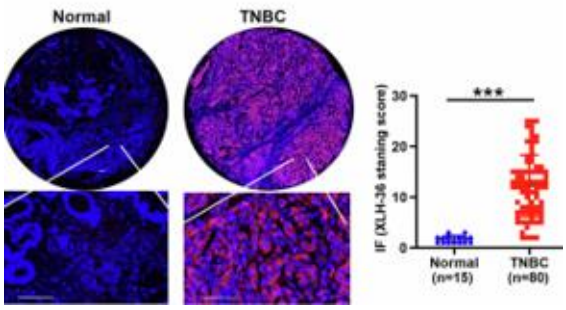
b)



c)



d)



e)

Figure 2. XLH-36 Exhibits Elevated Expression in TNBC Specimens and Serves as a Strong Indicator of Unfavorable Prognosis

(a) Kaplan–Meier overall survival plots for breast cancer (BRCA) patients stratified by XLH-36 expression levels.

(b) Expression of the XLH-36-encoding lncRNA C5orf66-AS1 shows a significant correlation with advanced clinical stages in BRCA patients.

(c) Reduced XLH-36 expression correlates with improved overall survival in patients with triple-negative breast cancer (TNBC).

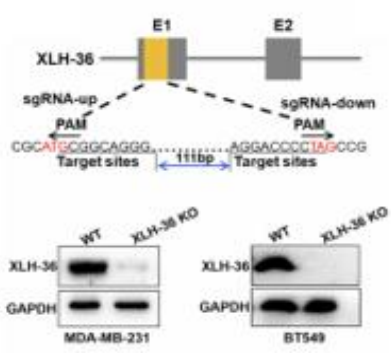
(d) Immunohistochemistry (IHC) staining using anti-XLH-36 antibodies reveals markedly higher XLH-36 protein levels in BRCA tissues compared to normal controls. Scale bar: 200 μm.

(e) Immunofluorescence (IF) assay with anti-XLH-36 antibodies demonstrates substantially elevated XLH-36 levels in TNBC tissues. Scale bar: 50 μm.

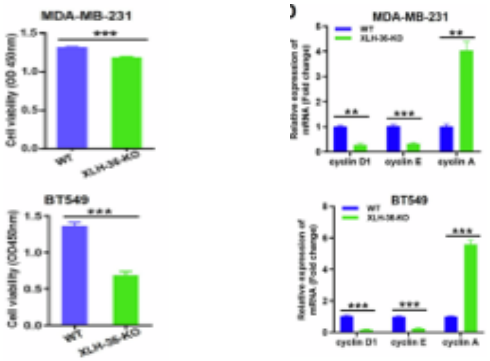
Survival analyses were performed using Kaplan–Meier method with log-rank testing. Data are represented as mean ± standard deviation (SD). Statistical significance: *P < 0.05, **P < 0.01, ***P < 0.001.

XLH-36 drives TNBC metastasis both in vitro and in vivo

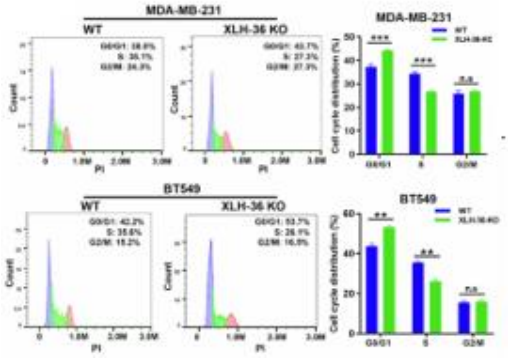
To explore the functional role of XLH-36 in TNBC progression, knockdown and overexpression studies were conducted in two TNBC cell lines (MDA-MB-231 and BT549; **(Figure 3a)**. Knockdown of XLH-36 (XLH-36 KO) in these cells substantially suppressed cell proliferation **(Figures 3b–3g)**, cell cycle progression, migration, invasion, and wound-healing ability, whereas XLH-36 overexpression (XLH-36 OE) markedly enhanced proliferation, cell cycle advancement, migration, invasion, and wound closure. These findings indicate that XLH-36 acts as an oncogenic micropeptide, facilitating TNBC tumor progression.



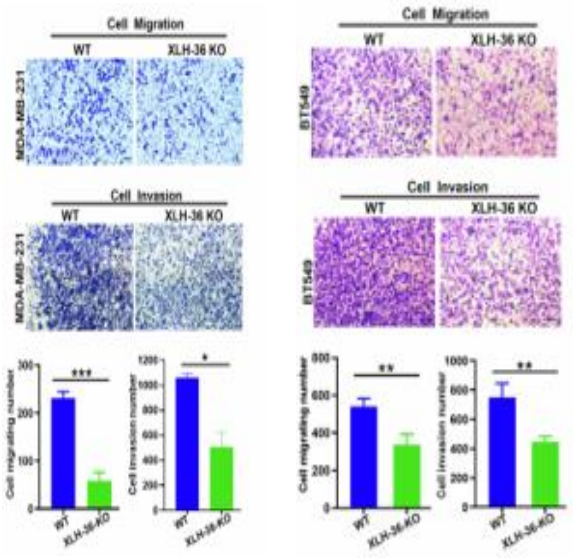
a)



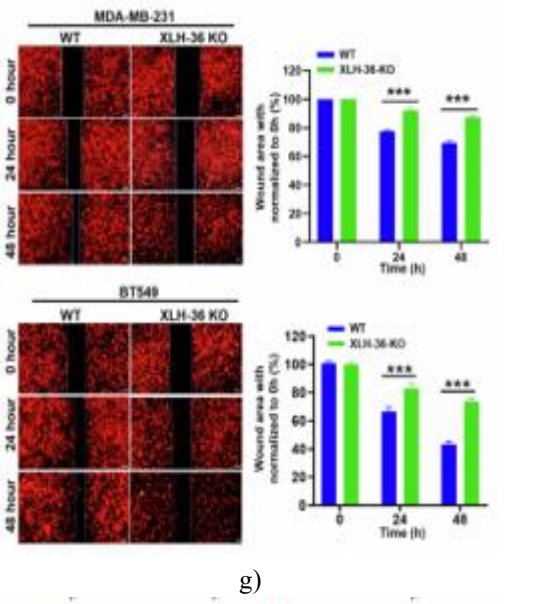
b) d)



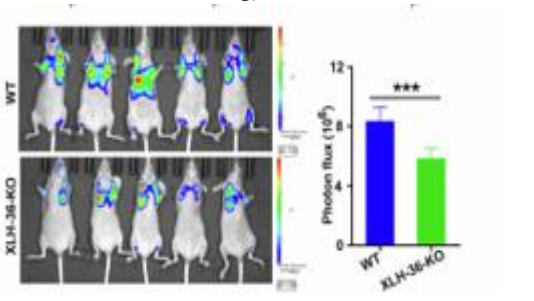
c)



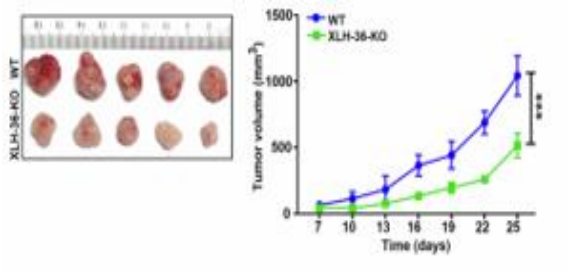
e) f)



g)



h)



i)

Figure 3. XLH-36 Serves as an oncogenic micropeptide facilitating triple-negative breast cancer metastasis in vitro and in vivo

(a) Top: Illustration of the CRISPR/Cas9 approach for generating XLH-36 knockout (KO) in MDA-MB-231 and BT549 cell lines. Bottom: Western blot confirming absence of XLH-36 protein in knockout cells compared to wild-type controls.

(b) Assessment of cell growth using the CCK-8 proliferation assay across the specified groups.

(c) Cell cycle profiling via flow cytometry in the indicated cell populations.

(d) Quantitative RT-PCR measurement of key cell cycle regulators in wild-type versus XLH-36-KO MDA-MB-231 and BT549 cells.

Cell motility and invasiveness were evaluated through Transwell migration (e), Matrigel-based invasion (f), and scratch-wound healing assays (g).

(h) Representative lung metastasis images and quantification in mice following intravenous injection of the designated MDA-MB-231 cells.

(i) Tumor growth curves and volume measurements from subcutaneous xenografts.

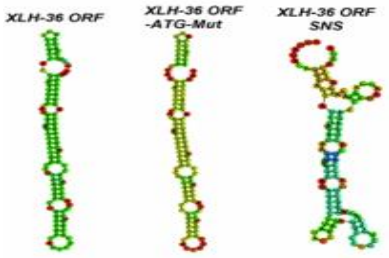
Results are expressed as mean ± SD (n = 5 mice/group). Two-way ANOVA was applied where appropriate. *P < 0.05, **P < 0.01, ***P < 0.001.

In vivo studies further validated the pro-tumorigenic activity of XLH-36. Delivery of XLH-36-KO cells via tail-vein injection led to a sharp decline in lung colonization **(Figure 3h)** and impaired subcutaneous tumor development **(Figure 3i)**, with lower proliferation rates shown by reduced Ki-67 positivity. H&E staining highlighted fewer and smaller metastatic nodules in lungs from the knockout group. Overexpression of XLH-36 yielded contrasting results, promoting metastasis and growth. Overall, these observations reinforce XLH-36's oncogenic properties in TNBC and highlight its potential as a therapeutic target for inhibiting tumor spread.

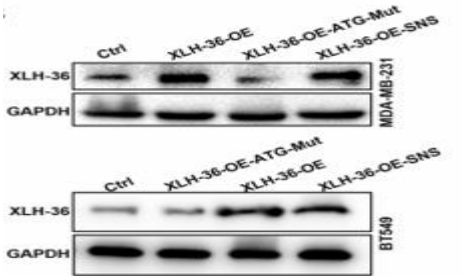
The pro-oncogenic activity in TNBC originates from the XLH-36 micropeptide, independent of its host lncRNA C5orf66-AS1

To clarify whether the cancer-promoting effects stem from the XLH-36 peptide or its encoding long non-coding RNA C5orf66-AS1, three distinct overexpression vectors were introduced into MDA-MB-231 and BT549 cells **(Figure 4a)**: the native XLH-36 open reading frame (XLH-36-OE), a start-codon-mutated version of the small ORF from C5orf66-AS1 (XLH-36-OE-ATG-Mut), and a synonymous-codon variant of XLH-36-ORF that yields the identical protein without altering non-coding sequences (XLH-36-OE-SNS).

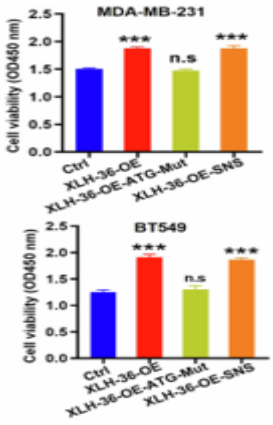
Critically, only XLH-36-OE and XLH-36-OE-SNS constructs elevated XLH-36 protein abundance **(Figure 4b)** and boosted proliferation and migratory capacity **(Figures 4c–4e)**. The ATG-mutant failed to do so. Additionally, defects in growth and motility caused by XLH-36 knockout were fully reversed by XLH-36-OE or XLH-36-OE-SNS, but not by the translation-defective ATG-mutant **(Figures 4f–4i)**. These experiments establish that the tumor-driving functions in TNBC are mediated directly by the translated XLH-36 micropeptide, rather than RNA-level effects from C5orf66-AS1.



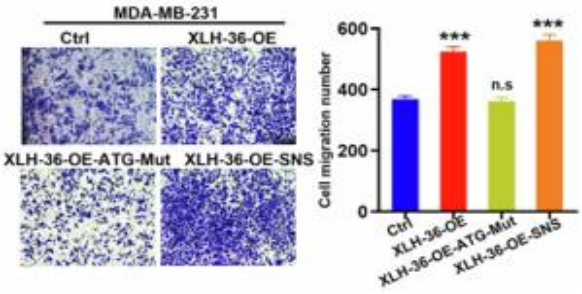
a)



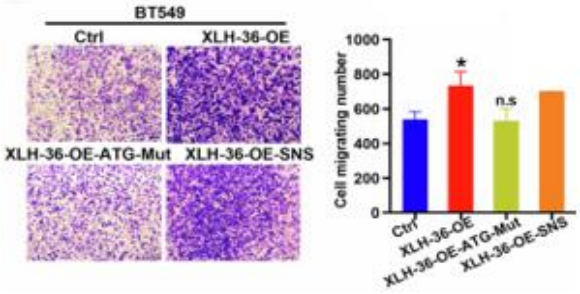
b)



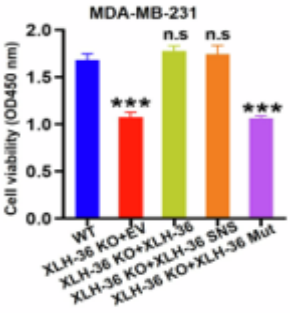
c)



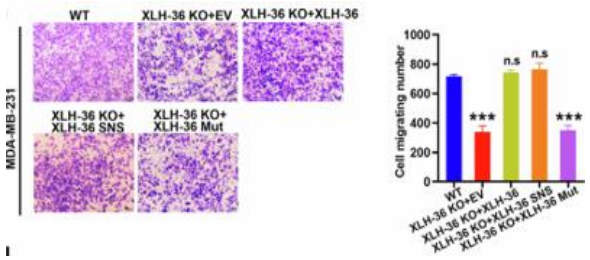
d)



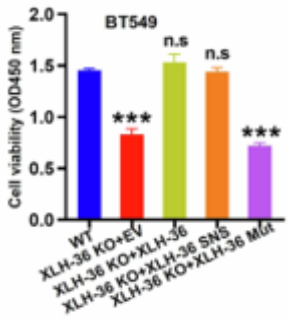
e)



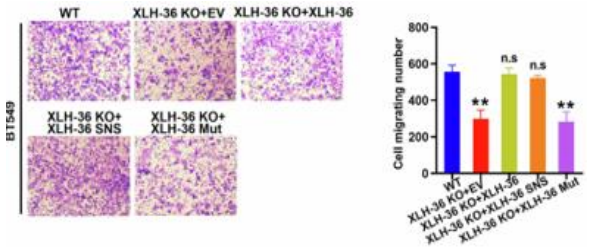
f)



g)



h)



i)

Figure 4. The Pro-tumorigenic effects in TNBC are driven by the XLH-36 micropeptide, not by its encoding lncRNA C5orf66-AS1

(a) Predicted minimum free energy secondary structures for RNA folding of the wild-type XLH-36 open reading frame (ORF) and its mutant variant, generated using the RNAfold online tool.

(b) Western blot evaluation of XLH-36 protein expression from the specified constructs following lentiviral transduction in MDA-MB-231 and BT549 cells.

Assessment of cell proliferation (c) and migratory capacity (d, e) in MDA-MB-231 and BT549 cells expressing the indicated variants.

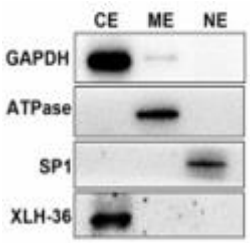
Rescue experiments measuring proliferation (f) and migration (g) in wild-type and XLH-36 knockout MDA-MB-231 cells after overexpression of XLH-36, the synonymous substitution variant (XLH-36-OE-SNS), or the start-codon mutant (XLH-36-OE-ATG-Mut).

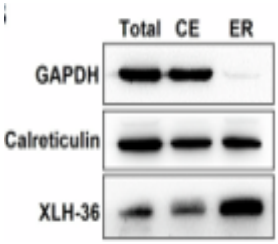
Similar rescue assays for proliferation (h) and migration (i) conducted in wild-type and XLH-36 knockout BT549 cells with the same overexpression constructs.

Data represent mean ± SD from three independent biological replicates. ns: not significant, *P < 0.05, **P < 0.01, ***P < 0.001.

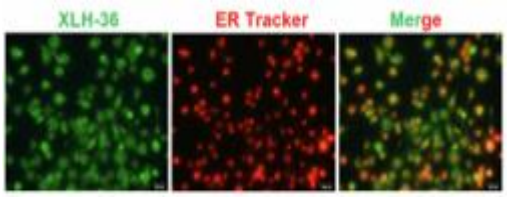
XLH-36 predominantly localizes to the endoplasmic reticulum

Protein function is often dictated by its subcellular location. To elucidate the mechanism by which XLH-36 drives TNBC progression, we investigated its distribution within the cell. Subcellular fractionation combined with in silico structural predictions (via TMHMM and HDOCK tools) revealed predominant cytoplasmic localization **(Figure 5a)**. Immunofluorescence studies further demonstrated clear co-localization of XLH-36 with endoplasmic reticulum markers, but not with mitochondrial markers **(Figures 5b and 5c)**.

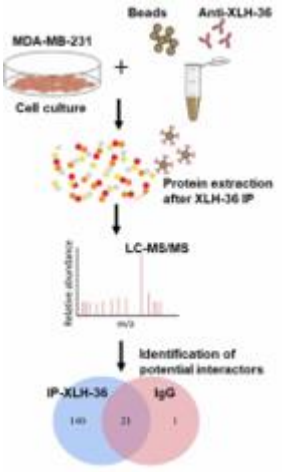




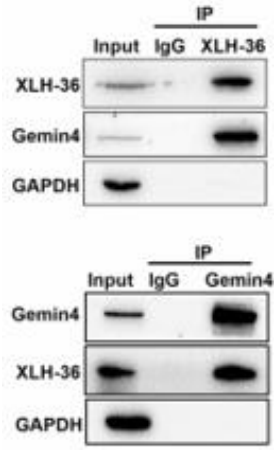
a) b)



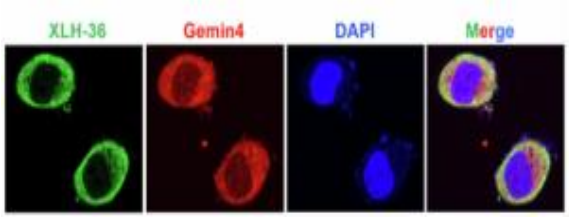
c)



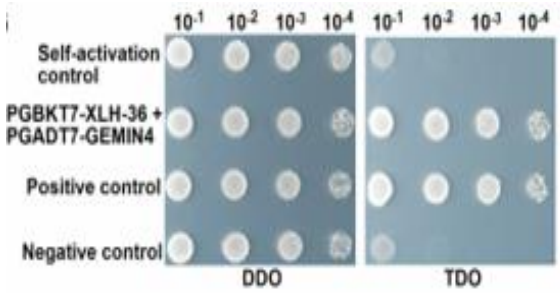
d)



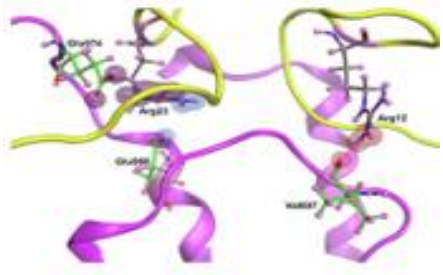
e)



f)



g)

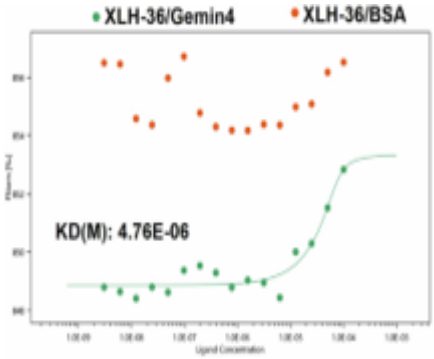


h)

The contact list between the XLH-36 and Gemin4

The residues in XLH-36	The residues in Gemin4
Arg12	Lys608,Val897,His925, Leu931, His994
Pro16	Ser1037, Glu1040
Arg17	Ile894, Arg1033

i)



j)

Figure 5. XLH-36 is localized to the endoplasmic reticulum and directly interacts with Gemin4. (a and b) Western blot analysis of XLH-36 in endoplasmic reticulum and cytosolic fractions from MDA-MB-231 cells. Markers include ATPase (plasma membrane), SP1 (nuclear), GAPDH (cytosolic), and Calreticulin (endoplasmic reticulum). (c) Immunofluorescence images showing colocalization of XLH-36 (green) with the endoplasmic reticulum marker (red) in MDA-MB-231 cells. (d) Schematic illustrating the identification of XLH-36-interacting proteins via mass spectrometry (MS) following immunoprecipitation. (e) Co-immunoprecipitation (Co-IP) experiments confirming the endogenous interaction between XLH-36 and Gemin4. (f) Representative immunofluorescence images of MDA-MB-231 cells co-stained for XLH-36, Gemin4, and DAPI (nuclear stain). (g) Yeast two-hybrid assay demonstrating direct binding between Gemin4 and XLH-36 (growth on DDO: SD/Trp-/Leu-; TDO: SD/Trp-/Leu-/His-). (h) Molecular docking simulation of XLH-36 with Gemin4.

(i) Table listing potential interaction sites, with residues R12, P16, and R17 in XLH-36 forming salt bridges and hydrogen bonds with residues K608, V897, H925, L931, H994, S1037, E1040, I894, and R1033 in Gemin4.

(j) MicroScale Thermophoresis (MST) assay measuring the binding affinity, yielding a KD value of 4.76×10^{-6} for XLH-36 and Gemin4.

XLH-36 directly interacts with gemin4

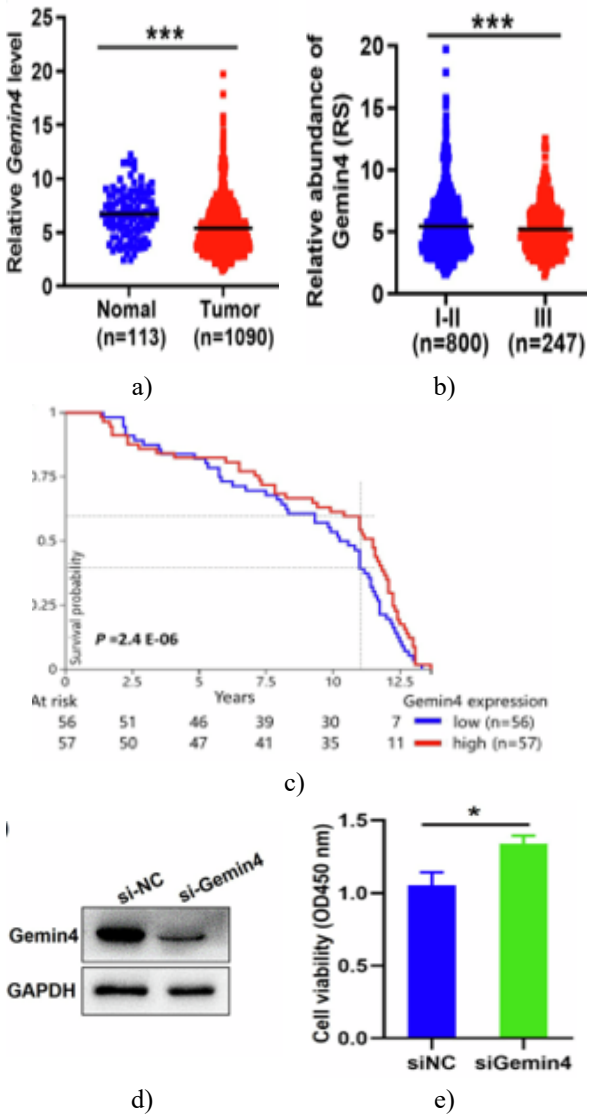
To screen for potential binding partners of XLH-36, co-immunoprecipitation followed by mass spectrometry analysis was conducted **(Figure 5d)**. A total of 140 proteins co-precipitated with XLH-36 were detected. Among the top-scoring candidates (Gemin4, EEF2, and ACLY) validated by Co-IP, only Gemin4 showed confirmed binding to XLH-36 **(Figure 5e)**. Immunofluorescence in MDA-MB-231 cells further demonstrated colocalization and interaction between XLH-36 and Gemin4 **(Figure 5f)**.

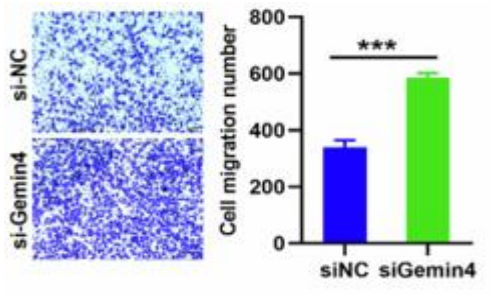
A yeast two-hybrid assay provided additional evidence of direct binding between XLH-36 and Gemin4 **(Figure 5g)**. Molecular docking was performed using predicted structures of XLH-36 and Gemin4, revealing involvement of XLH-36 residues R12, P16, and R17 in interactions with Gemin4 residues K608, V897, H925, L931, H994, S1037, E1040, I894, and R1033 via salt bridges and hydrogen bonds **(Figures 5h and 5i)**. MST experiments with synthetic XLH-36 and purified Gemin4 determined a KD of 4.76×10^{-6}, indicating high binding affinity **(Figure 5j)**. Mutation of R12, P16, and R17 to alanine reduced binding affinity, with the P16 mutation causing the greatest reduction, highlighting its critical role in the interaction.

XLH-36 drives TNBC progression through binding to Gemin4

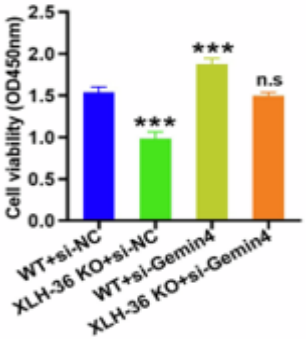
Gemin4 is a key component of the survival of motor neuron (SMN) complex [23], which associates with other proteins including SMN and GEMIN2, 3, 5, 6, 7, and 8 to maintain snRNP integrity essential for spliceosome assembly and pre-mRNA processing [24]. Prior to this study, no links between Gemin4 and TNBC pathogenesis had been reported. Further exploration of Gemin4's role in triple-negative breast cancer is essential for a fuller understanding of its functions.

TCGA data analysis showed significantly reduced Gemin4 expression in breast cancer (BRCA) tissues relative to normal tissues **(Figure 6a)**, with lower Gemin4 levels strongly associated with advanced clinical stages **(Figure 6b)**. Kaplan–Meier analysis indicated that higher Gemin4 expression correlated with reduced risk of BRCA-related mortality **(Figure 6c)**. Functional studies demonstrated that Gemin4 knockdown markedly enhanced proliferation and migration in MDA-MB-231 cells **(Figures 6d–6f)** and abrogated the suppressive effects of XLH-36 on TNBC cell proliferation and migration **(Figures 6g and 6h)**. Although XLH-36 directly binds Gemin4, overexpression of XLH-36 did not alter Gemin4 levels, and Gemin4 knockdown had no impact on XLH-36 expression, suggesting no mutual regulatory effects. Collectively, these findings suggest that XLH-36 promotes oncogenic activity in TNBC by interacting with Gemin4, which functions as a novel tumor suppressor in this context.

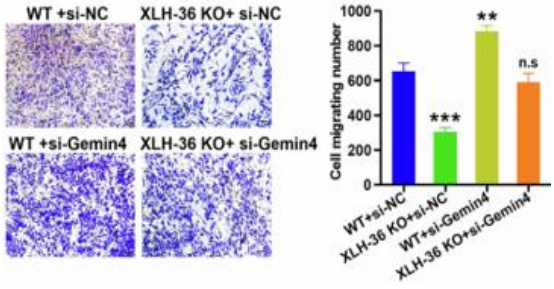




f)



g)



h)

Figure 6. Gemin4 functions as a tumor suppressor and plays a key role in XLH-36-driven TNBC advancement.

(a) Analysis of TCGA data shows markedly lower Gemin4 expression in breast cancer (BRCA) samples from patients (n = 1090) relative to matched normal adjacent tissues (n = 113).

(b) Reduced Gemin4 expression correlates significantly with more advanced clinical stages in BRCA cases.

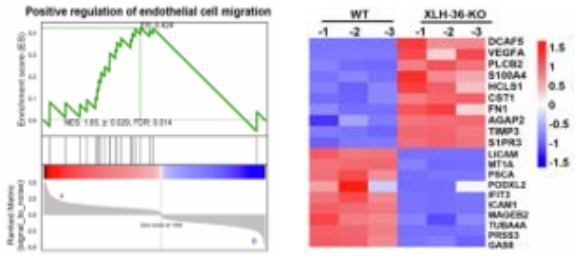
(c) Kaplan-Meier plots of overall survival in BRCA patients divided by high versus low Gemin4 expression, with statistical comparison via log-rank test.

(d) Depletion of Gemin4 enhances cell proliferation (e) and migratory capacity (f) in MDA-MB-231 cells.

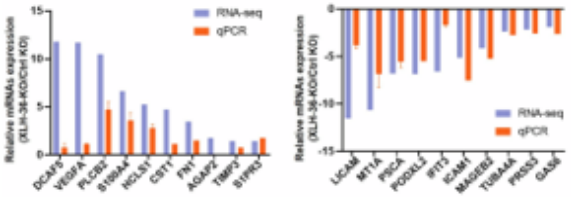
(g) Proliferation assays and (h) migration assays in control versus XLH-36 knockout MDA-MB-231 cells after Gemin4 silencing. Results are expressed as mean ± SD across three independent biological replicates. *P < 0.05, **P < 0.01, ***P < 0.001.

XLH-36 modulates TNBC metastasis mainly via ICAM1 and S100A4

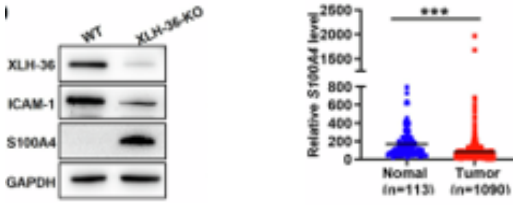
To elucidate how XLH-36 facilitates metastasis in TNBC, RNA sequencing was carried out on cells lacking XLH-36. This revealed 640 genes with significant differential expression (fold change ≥ 2, P < 0.05), of which 310 were increased and 330 were decreased. Enrichment analyses using KEGG pathways and GSEA highlighted involvement of these genes in extracellular matrix-receptor interactions, focal adhesions, and cell adhesion molecules—processes critical for cellular motility and invasiveness **(Figure 7a)**. From the RNA-seq results, the 10 most upregulated and 10 most downregulated genes **(Figure 7b)** were confirmed through qRT-PCR **(Figure 7c)**. Notably, ICAM1 and S100A4 showed the strongest dependence on XLH-36 for their expression, as evidenced at both transcript and protein levels **(Figure 7d)**.



a) b)



c)



d) e)

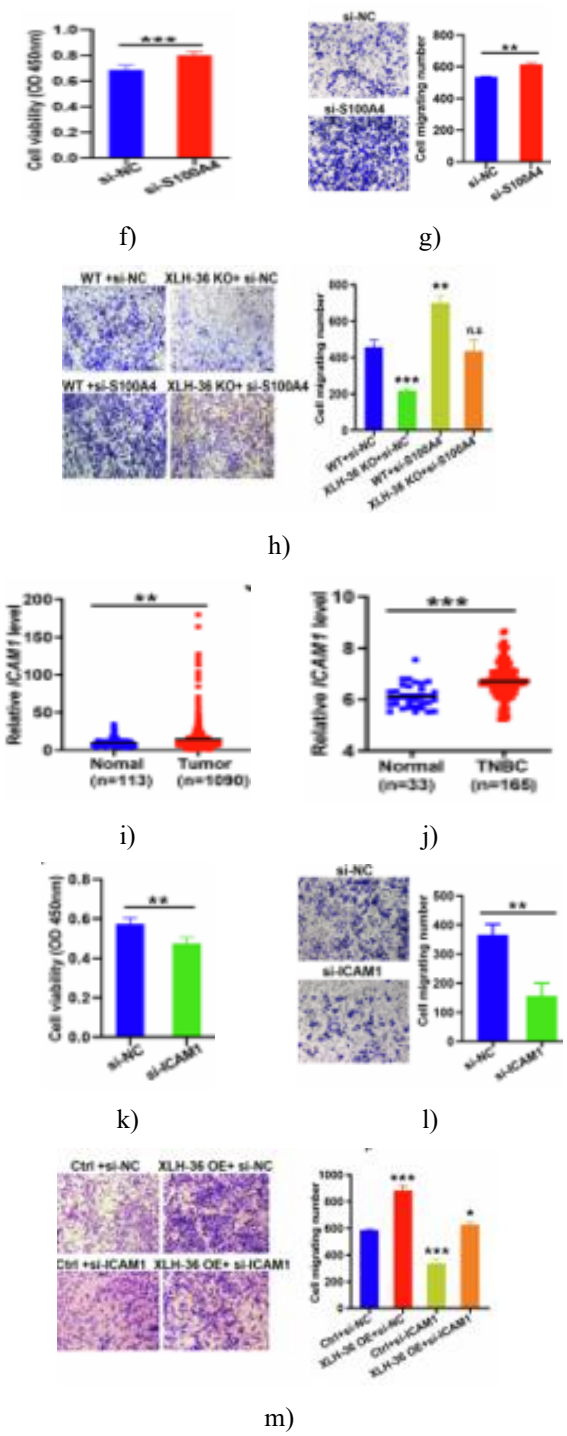


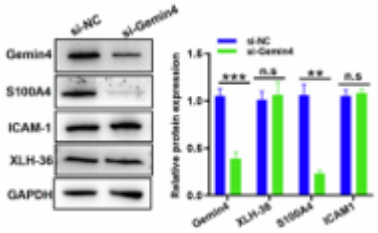
Figure 7. XLH-36 drives TNBC metastasis predominantly via ICAM1 and S100A4.
(a) Gene set enrichment analysis (GSEA) of cell migration signatures comparing XLH-36 knockout (KO) to wild-type (WT) MDA-MB-231 cells.
(b) Heatmap displaying the top 10 upregulated and top 10 downregulated differentially expressed genes in XLH-36-KO versus WT MDA-MB-231 cells.
(c) Left: qRT-PCR confirmation of the top 10 upregulated genes in XLH-36-KO relative to WT cells. Right: qRT-PCR confirmation of the top 10 downregulated genes in XLH-36-KO relative to WT cells.
(d) Immunoblot showing protein levels of S100A4 and ICAM1 in WT versus XLH-36-KO MDA-MB-231 cells.
(e) TCGA data reveal significant downregulation of S100A4 in breast cancer (BRCA) tissues (n = 1090) compared to adjacent normal tissues (n = 113).
(f) Proliferation and (g) migration assays in MDA-MB-231 cells after transfection with control siRNA (si-NC) or si-S100A4.
(h) Migration assays in WT or XLH-36 KO MDA-MB-231 cells following S100A4 knockdown.
(i) TCGA analysis shows significant upregulation of ICAM1 in BRCA tissues (n = 1090) versus adjacent normal tissues (n = 113).
(j) GSE76250 dataset demonstrates elevated ICAM1 expression in TNBC samples (n = 165) compared to adjacent normal tissues (n = 33).
(k) Proliferation and (l) migration assays in MDA-MB-231 cells transfected with si-NC or si-ICAM1.
(m) Migration assays in control or XLH-36-overexpressing MDA-MB-231 cells after ICAM1 knockdown. Data are shown as mean ± SD from three independent biological experiments. *P < 0.05, **P < 0.01, ***P < 0.001.

ICAM1 (Intercellular Adhesion Molecule 1), also referred to as CD54, belongs to the immunoglobulin superfamily of adhesion molecules [25] and is involved in diverse processes such as cell-cell adhesion, angiogenesis, tumor development, apoptosis, and metastatic spread [26]. S100A4, known alternatively as fibroblast-specific protein 1 (FSP1) or metastasin, is a member of the S100 calcium-binding protein family and contributes to cell motility, invasion, angiogenesis, and epithelial-mesenchymal transition (EMT) [27, 28]. The precise ways in which S100A4 drives malignancy remain incompletely defined. To clarify how ICAM1 and S100A4 mediate XLH-36's effects on TNBC progression and metastasis, clinical databases (TCGA and GTEx) were queried alongside loss-of-function studies. Findings indicated markedly reduced S100A4 in BRCA tissues **(Figure 7e)**, with S100A4 depletion increasing cell proliferation and migration **(Figures 7f and 7g)**. Furthermore, S100A4 silencing abolished the

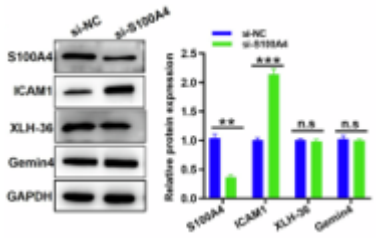
suppressive impact of XLH-36 loss on migration **(Figure 7h)**. In contrast, ICAM1 was overexpressed in BRCA and TNBC specimens **(Figures 7i and 7j)**; its knockdown impaired proliferation and migration **(Figures 7k and 7l)**, effects that were overridden by XLH-36 overexpression **(Figure 7m)**. Together, these results establish ICAM1 and S100A4 as central downstream effectors of XLH-36 in TNBC pathogenesis.

The XLH-36/Gemin4 complex enhances ICAM1 expression and EMT by modulating S100A4 mRNA splicing

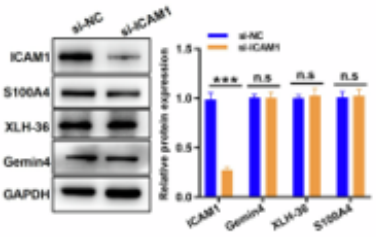
The influence of the XLH-36/Gemin4 binding on S100A4 and ICAM1 levels was examined next, along with potential crosstalk between ICAM1 and S100A4. Gemin4 depletion selectively reduced S100A4 abundance without altering ICAM1 **(Figure 8a)**. S100A4 knockdown, in turn, elevated ICAM1 while leaving XLH-36 and Gemin4 unchanged **(Figure 8b)**. ICAM1 depletion had no effect on XLH-36, Gemin4, or S100A4 levels **(Figure 8c)**. Given Gemin4's established role in pre-mRNA splicing essential for mRNA maturation, we hypothesized that it influences S100A4 transcript processing and stability. To test this, transcription was inhibited with α-amanitin in WT and Gemin4-knockdown cells, followed by RNA isolation at 0, 2, 4, 6, and 8 hours. RT-PCR revealed accelerated decay of S100A4 transcripts upon Gemin4 loss **(Figures 8d and 8e)**. Protein synthesis blockade with cycloheximide nearly eliminated endogenous S100A4 in Gemin4-depleted cells across 0–6 hours **(Figure 8f)**, consistent with diminished S100A4 mRNA resulting from impaired splicing due to Gemin4 knockdown.



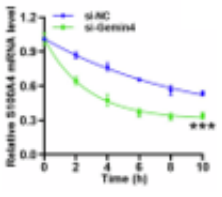
a)



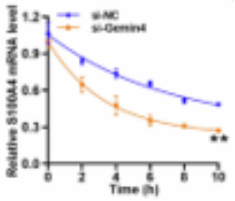
b)



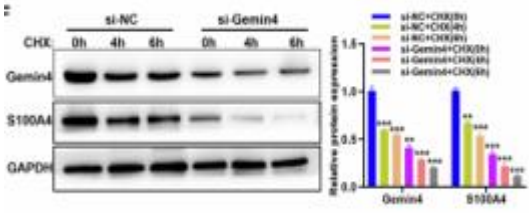
c)



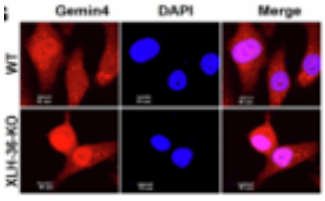
d)



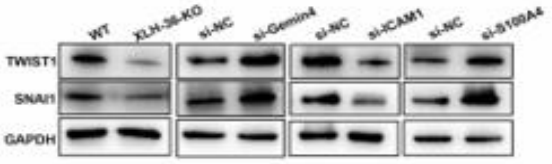
e)



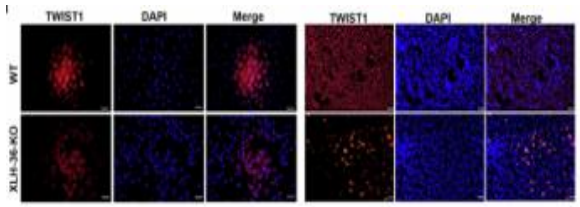
f)



g)



h)



i)

Figure 8. The XLH-36/Gemin4 interaction facilitates ICAM1 expression and EMT by controlling S100A4 mRNA splicing.

(a) Immunoblot analysis of XLH-36, Gemin4, S100A4, and ICAM1 protein levels in MDA-MB-231 cells transfected with control siRNA (si-NC) or si-Gemin4.

(b) Immunoblot analysis in cells transfected with si-NC or si-S100A4.

(c) Immunoblot analysis in cells transfected with si-NC or si-ICAM1.

(d, e) Assessment of total mRNA decay rates by quantitative RT-PCR in si-NC and si-Gemin4 cells after blocking transcription with α-amanitin.

(f) MDA-MB-231 cells were treated with cycloheximide (CHX, 100 μg/mL), and Gemin4 and S100A4 protein levels were monitored at the indicated time points.

(g) Immunofluorescence staining for Gemin4 (red) in wild-type (WT) and XLH-36 knockout (KO) MDA-MB-231 cells. Scale bar: 10 μm.

(h) Immunoblot showing protein levels of EMT markers TWIST1 and SNAI1 in the indicated MDA-MB-231 cell lines.

(i) Left: Representative immunofluorescence images of TWIST1 in WT and XLH-36-KO MDA-MB-231 cells. Right: Representative immunofluorescence images of TWIST1 in frozen tumor sections from mice subcutaneously injected with WT or XLH-36-KO MDA-MB-231 cells. Data are shown as mean ± SD from three independent biological replicates. *P < 0.05, **P < 0.01, ***P < 0.001.

The survival of motor neuron (SMN) complex, which includes Gemin4, is mainly cytoplasmic but shuttles into the nucleus to participate in spliceosomal snRNP assembly and maintenance for pre-mRNA splicing [29]. As a cytoplasmic micropeptide, XLH-36 is proposed to bind Gemin4 and block its nuclear translocation, thereby disrupting S100A4 transcription and translation. Immunofluorescence revealed that XLH-36 deficiency reduced cytoplasmic Gemin4 while enriching it in the nucleus **(Figure 8g)**. Conversely, XLH-36 overexpression produced the opposite pattern. These findings indicate that XLH-36 anchors Gemin4 in the cytoplasm, preventing its nuclear function in promoting S100A4 mRNA splicing and stability.

Both S100A4 and ICAM1 contribute to tumor metastasis through epithelial-mesenchymal transition (EMT) [30, 31]. Given the earlier observations, reduced S100A4 may relieve suppression of TNBC cellular functions, resulting in compensatory upregulation of ICAM1, although the underlying mechanism awaits further clarification. To examine links between XLH-36, Gemin4, S100A4, ICAM1, and EMT, expression of the EMT drivers TWIST1 and SNAI1 was evaluated. Levels of both markers were markedly lowered in XLH-36 knockout or ICAM1 knockdown cells, whereas depletion of Gemin4 or S100A4 increased their abundance **(Figure 8h)**. Immunofluorescence in MDA-MB-231 cells and xenograft tumors confirmed that XLH-36 loss suppressed TWIST1 **(Figure 8i)**, while XLH-36 overexpression elevated it. Overall, these results establish XLH-36 as a potent Gemin4-interacting micropeptide that drives EMT by modulating ICAM1 and S100A4 expression.

In this study, we demonstrate that the previously unstudied C5orf66-AS1 encodes a novel micropeptide, XLH-36, which drives TNBC progression and metastasis. Functioning as an oncopeptide, XLH-36 directly interacts with Gemin4, leading to S100A4 inactivation and upregulation of ICAM1, thereby facilitating EMT and metastatic behavior **(Figure 9)**. Collectively, these findings reveal a previously unrecognized molecular pathway regulating EMT and metastasis in TNBC.

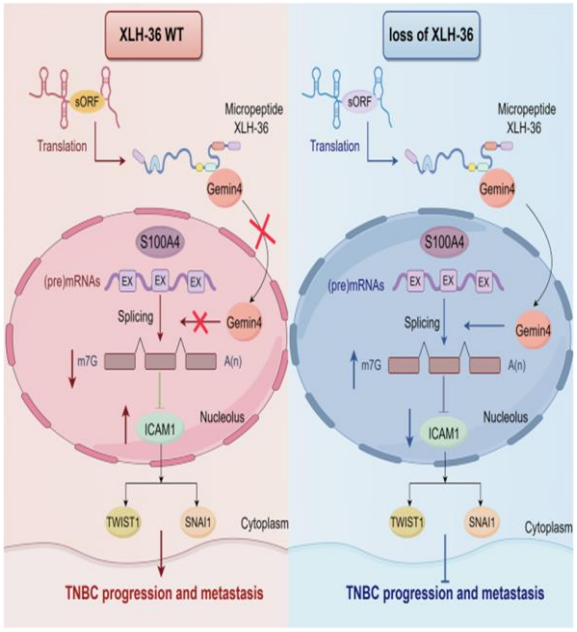


Figure 9. The proposed model for the role of C5orf66-AS1-encoded XLH-36 in TNBC metastasis.

In this investigation, we identified that C5orf66-AS1 encodes a novel endogenous micropeptide termed XLH-36. Evolutionary conservation analysis across 101 species revealed that XLH-36 is highly conserved among various primates, indicating its potential importance in essential biological functions. Additionally, examination of 1295 clinical samples from breast cancer (BRCA) and TNBC cases across multiple datasets demonstrated elevated XLH-36 expression in BRCA and TNBC tissues. Notably, XLH-36 levels were substantially lower in early-stage patients compared to those with advanced disease. Moreover, RNA-seq data from 32 cancer types in TCGA datasets indicated variable behavior of XLH-36 across tumors, 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), warranting additional studies.

Prior research has shown that C5orf66-AS1, previously classified as a “non-coding RNA”, exhibits unique roles in several malignancies, such as pituitary null cell adenomas, cervical cancer, and gastric cancer [32–34]. Here, we established that the micropeptide XLH-36, rather than the lncRNA C5orf66-AS1 itself, drives specific oncogenic activities in TNBC. The potential effects of XLH-36 in other cancer types require further exploration.

TNBC represents a highly heterogeneous malignancy with restricted therapeutic options beyond standard chemotherapy [35]. While certain oncogenes (e.g., MMPs [36], CDK4/6 [37], SET [38], and mutant BRCA1 [39]) have been identified in TNBC, few micropeptides have been implicated as proto-oncogenes. Emerging immunotherapies hold potential for TNBC [40]. In the KEYNOTE-355 trial [41], combining immune checkpoint inhibitors (ICIs) with chemotherapy in PD-L1-positive metastatic TNBC patients lowered mortality risk by 9% over a 50-month period. In our analysis, TNBC patients with low XLH-36 expression showed a 20% improved overall survival compared to those with high expression (85% vs. 65%) at the same timeframe, suggesting XLH-36 as a possible prognostic indicator. Future work should expand clinical sample sizes and employ liquid biopsies to assess XLH-36's utility in TNBC diagnosis and prognosis.

Previous reports link Gemin4 gene polymorphisms to progression in colon, bladder, liver, and kidney cancers [42, 43], with elevated Gemin4 levels correlating with unfavorable outcomes in basal-like breast cancer (BLBC) patients [44], though its precise involvement in TNBC oncogenesis remains undefined. We found that reduced Gemin4 expression correlates with improved overall survival in BRCA patients and suppresses proliferation and migration in TNBC cells. Co-immunoprecipitation, yeast two-hybrid screening, molecular dynamics modeling, and binding assays confirmed direct interaction between XLH-36 and Gemin4. Notably, this binding does not alter mutual expression levels in TNBC cells. Advanced techniques such as cryo-EM or co-crystallization could precisely map all interaction sites in future studies. Our findings highlight Gemin4's key role in TNBC and propose it as a therapeutic target, pending validation in larger clinical cohorts.

Gemin4 interacts specifically with a DEAD-box protein and multiple spliceosome core components, playing a direct part in spliceosome assembly and regeneration for nuclear pre-mRNA splicing. S100A4, a calcium-binding protein in the S100 family, contributes significantly to tumor metastasis via modulation of adhesion, extracellular matrix remodeling, and motility.

XLH-36 localizes to the cytoplasm, directly interacts with Gemin4, and blocks its nuclear translocation, thus impairing Gemin4-mediated splicing of S100A4 mRNA. ICAM1, an intercellular adhesion molecule, participates

in tumor progression through control of adhesion, angiogenesis, evolution, apoptosis, metastasis, and related processes [45]. The regulatory link between ICAM1 and S100A4 is unclear. We noted marked increases in ICAM1 and EMT markers (TWIST1 and SNAI1) alongside reduced S100A4, implying compensatory adaptations in TNBC cells to sustain survival and metastasis.

Conclusion

In conclusion, we uncover that the newly identified micropeptide XLH-36 is markedly overexpressed in TNBC, with its expression strongly associated with tumor stage and patient overall survival. XLH-36 could emerge as a target for TNBC therapy, diagnosis, and prognostication. This work advances understanding of micropeptide functions and their clinical translation, while adding a novel peptide to the human proteome.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. 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.
2. Yin L, Duan JJ, Bian XW, Yu SC. Triple-negative breast cancer molecular subtyping and treatment progress. Breast Cancer Res. 2020;22:61.
3. Howard FM, Olopade OI. Epidemiology of triple-negative breast cancer: a review. Cancer J. 2021;27:8–16.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. Wright BW, Yi Z, Weissman JS, Chen J. The dark proteome: translation from noncanonical open reading frames. Trends Cell Biol. 2022;32:243–58.
11. 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.
12. Setrerrahmane S, Li M, Zoghbi 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.
13. 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.
14. 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.
15. 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.
16. 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.

17. 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.
18. 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.
19. 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.
20. 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.
21. Choi SW, Kim HW, Nam JW. The small peptide world in long noncoding RNAs. Brief Bioinform. 2019;20:1853–64.
22. 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.
23. 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.
24. 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.
25. Haydinger CD, Ashander LM, Tan ACR, Smith JR. Intercellular adhesion molecule 1: more than a leukocyte adhesion molecule. Biology. 2023;12:743.
26. 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.
27. 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.
28. 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.
29. 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:bio032409.
30. Chow KH, Park HJ, George J, Yamamoto K, Gallup AD, Graber 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.
31. 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.
32. 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.
33. 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.
34. 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. Onco Targets Ther. 2020;13:2795–805.
35. 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.
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.
37. 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, Upadhaya 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 Przegl 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. Onco Targets 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.