

The microprotein HDSP promotes gastric cancer progression through activating the MECOM-SPINK1-EGFR signaling axis

Received: 7 July 2023

Accepted: 27 July 2024

Published online: 27 September 2024

 Check for updates

Yuli Chen^{1,2,9}, Qiuhui Li^{1,9}, Xiang Yu^{3,9}, Lu Lu¹, Zihan Zhou⁴, Mingjie Li⁵, Rui Xia⁶, Xiongkang Gan⁷, Yanming Hu², Guoqing Guo¹, Jiahao Guo², Hanyang Li¹, Qiunuo Li⁴, Yanwen Liu⁸, Xianghua Liu^{1,10}✉ & Ming Sun^{2,10}✉

The presence of noncanonical open reading frames within lncRNAs (long non-coding RNAs) suggests their potential for translation, yielding various functional peptides or proteins. However, the existence and specific roles of these products in gastric cancer remain largely unclear. Here we identify the HOXA10-HOXA9-derived small protein (HDSP) in gastric cancer through comprehensive analysis and experimental validation, including mass spectrometry and western blotting. HDSP exhibits high expression and oncogenic roles in gastric cancer. Mechanistically, HDSP blocks TRIM25-mediated ubiquitination and degradation by interacting with MECOM, leading to MECOM accumulation and enhanced SPINK1 transcription—a gene promoting cancer via the EGFR signaling pathway. Furthermore, MECOM fosters HOXA10-HOXA9 transcription, establishing a feedback loop activating SPINK1-EGFR signaling. HDSP knockdown inhibits tumor growth in a PDX (patient-derived xenograft) model, and infusion of an artificially synthesized HDSP peptide as a neoantigen enhances immune cell-mediated anti-tumor efficacy against gastric cancer *in vitro* and *in vivo*. These findings propose HDSP as a potential therapeutic target or neoantigen candidate for gastric cancer treatment.

The worldwide incidence of gastric cancer (GC) ranks fifth among all malignancies and fourth in terms of mortality¹. According to the latest cancer statistics from the National Cancer Center in China, GC has the third highest incidence and mortality rates in the country². Despite significant advancements in molecular targeting therapy and immune checkpoint inhibitors, the 5-year overall survival rate of GC patients remains below 30%, which is primarily due to the lack of effective early diagnostic methods and limited therapeutic targets^{3,4}. Hence, it is

crucial to identify and characterize novel biomarkers and understand the regulatory mechanisms that govern GC cell proliferation and metastasis to improve diagnosis and facilitate the development of innovative therapeutic approaches.

Multi-omics studies, including RNA sequencing (RNA-seq), ribosome profiling, and bioinformatics analysis, have revealed numerous noncanonical open reading frames (ORFs) in the human genome^{5–7}. In contrast to classical coding sequences in mRNAs, noncanonical ORFs

¹Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Nanjing Medical University, Nanjing 211166, China. ²Suzhou Cancer Center Core Laboratory, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital, Gusu School, Nanjing Medical University, Suzhou 215001, China. ³Department of General Surgery, The Affiliated Yantai Yuhuangding Hospital of Qingdao University, Yantai 264000, China. ⁴The First Clinical Medical College of Nanjing Medical University, Nanjing 211166, China. ⁵Asset Management Company, Nanjing Medical University, Nanjing 211166, China. ⁶Department of Laboratory, Nanjing Chest Hospital, Nanjing 210029, China. ⁷Department of Cardiovascular Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing 210029, China. ⁸Department of Oncology, Zhongda Hospital, Medical School of Southeast University, Nanjing 210009, China. ⁹These authors contributed equally: Yuli Chen, Qiuhui Li, Xiang Yu. ¹⁰These authors jointly supervised this work: Xianghua Liu, Ming Sun.

✉ e-mail: liuxianghua@njmu.edu.cn; sunming348@hotmail.com

are often derived from the 5'untranslated regions of mRNAs, long non-coding RNAs (lncRNAs), and circular RNAs (circRNAs)^{8,9}. Previously, most of these noncanonical ORFs were unannotated, and their biological significance was unclear. However, Ji et al. recently identified high-priority ORFs in lncRNAs and determined a significant number of peptides/proteins encoded by these ORFs through mass-spectrometry-based proteomics¹⁰. In a previous study, we used ribosome profiling and bioinformatics analysis to identify hundreds of cryptic lncRNA-encoded ORFs. CRISPR-based screens revealed that disrupting these noncanonical ORFs resulted in diverse cellular effects in human estrogen receptor α -positive (ER+) breast cancer cells¹¹. Prensner et al. confirmed experimentally that 257 noncanonical ORFs out of 553 candidates were translatable, and that their translation was linked to biological activity based on CRISPR tiling¹². These findings indicate that systematic exploration of "hidden ORFs" in cancer can reveal an abundant source of previously undiscovered peptides and proteins.

Currently, only a few peptides/proteins translated from circRNAs have been identified in GC. CircPGD encodes a PGD-219aa protein that promotes malignant behavior in GC cells by regulating epithelial-mesenchymal transition through the SMAD2/3 and YAP signaling pathways¹³. AXINI-295aa, which is encoded by circAXINI, stimulates the progression of GC by competitively interacting with APC and activating the β -catenin signaling pathway¹⁴. GSPT1-238aa, which is encoded by circGSPT1, inhibits tumorigenesis in GC by regulating autophagy via the PI3K/AKT/mTOR pathway through its interaction with the vimentin/Beclin1/14-3-3 complex¹⁵. DIDO1-529aa, a protein encoded by circDIDO1, interacts with PARP1 to impair its binding to damaged DNA and weaken its enzymatic activity¹⁶. Although the functional impact of peptides/proteins encoded by circRNAs in GC has been established, little is known about the coding potential and functional relevance of noncanonical ORFs within lncRNAs in GC. Investigating and characterizing these ORFs may provide insights into GC mechanisms and potential therapeutic targets.

In the present study, we aim to identify translatable lncRNAs associated with GC by analyzing RNA-seq and ribosome profiling data. We find that HDSP, a microprotein encoded by the lncRNA HOXA10-HOXA9 in GC, has an oncogenic role by stabilizing the MECOM protein and activating the SPINK1-EGFR signaling pathway. In vivo experiments using patient-derived xenograft (PDX) and hematopoietic stem cell (HSC) humanized mouse models evaluate the potential of HDSP as a therapeutic target or neoantigen for GC treatment.

Results

HOXA10-HOXA9 encodes a microprotein HDSP in gastric cancer

Emerging evidence has demonstrated that functional peptides/proteins can be translated from lncRNAs¹⁷. We identified five GC-associated lncRNAs (HOTTIP, UCA1, FEZF1-AS1, HOXA10-HOXA9, and LINC00668) with coding potential by analyzing ribosomal profiling data, GC-related lncRNA microarray data, and sequencing data from the GEO database and TCGA (Fig. 1a, Supplementary Fig. 1a and Supplementary Data 1). Next, we observed the highest enrichment of HOXA10-HOXA9 among the identified lncRNAs in polysome-bound RNA from SGC-7901 cells (Fig. 1b and Supplementary Fig. 1b). Although the functions of the other four identified lncRNAs have been characterized in GC, the biological role of HOXA10-HOXA9 (NR_037940.1) in GC remains to be elucidated.

We further confirmed the upregulated expression of HOXA10-HOXA9 in 70 paired GC tissues and 7 GC cell lines, as well as in the datasets mentioned above, which captured our attention (Fig. 1c, d and Supplementary Fig. 1c). After confirming the cytoplasmic localization of HOXA10-HOXA9 through RNA FISH and nuclear/cytoplasmic fractionation assays, we identified a noncanonical 339-nt ORF2 within the sequence of HOXA10-HOXA9. This ORF has the potential to encode a 112-amino acid microprotein, which we have named HDSP (HOXA10-

HOXA9-derived small protein) (Fig. 1e and Supplementary Fig. 1d–k). Mutation of the start codon resulted in the absence of the HDSP-GFP or HDSP-Flag fusion protein, indicating that the initiation codon was functional (Fig. 1f–i). In addition, we confirmed the translation of exogenous HDSP-Flag through MS (Supplementary Fig. 1l). In vitro translation experiments provided additional validation regarding the codifiability of the HOXA10-HOXA9 sequence (Fig. 1j).

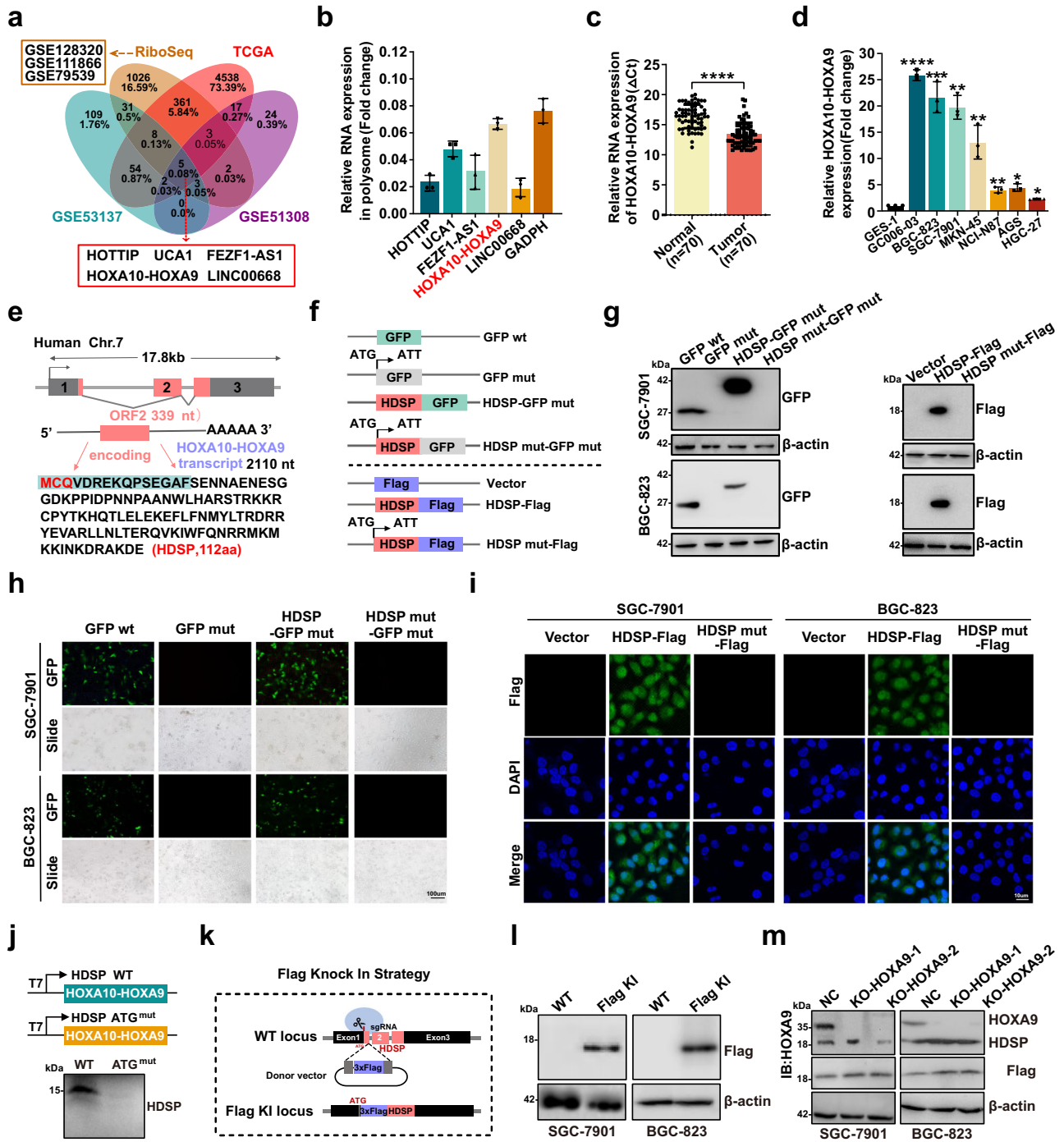
Moreover, we employed CRISPR/Cas9-mediated homologous recombination to introduce the Flag tag after the HDSP start codon in GC cells (Fig. 1k and Supplementary Fig. 1m). Western blot analysis further demonstrated the endogenous expression of HDSP in vivo (Fig. 1l). Due to the homologous nature of the HDSP sequence with that of HOXA9, we opted for a commercial anti-HOXA9 antibody capable of recognizing identical epitopes on both proteins for the detection of endogenous HDSP (Supplementary Fig. 2a, b). Western blot results revealed a significant upregulation of HDSP expression in both GC cells and tissues (Supplementary Fig. 2c, d). Importantly, no cross-reaction with HOXA9 was observed (Supplementary Fig. 2e, f). Concurrently, in Flag-KI GC cells, the expression of HDSP remained unaffected even upon knockout of HOXA9 (Fig. 1m). Collectively, these findings indicate that HOXA10-HOXA9 may encode a microprotein, HDSP, which is naturally expressed in GC tissues and cells, rather than being generated through shear of parental proteins.

High levels of endogenous HDSP correlate with poor prognosis for GC patients

To confirm the presence of endogenous HDSP in GC tissue and cells, we identified the unique peptide fragment of HDSP utilizing MS analysis of the whole-cell lysate, including the specific N-terminal amino acid sequence MCQ (Met-Cys-Gln), which is encoded by HDSP but not HOXA9. This further confirms the existence of endogenously expressed HDSP (Fig. 2b, d and Supplementary Fig. 2g). To discern the presence of endogenous HDSP in GC tissues and cells, we utilized the N-terminal amino acid sequence containing the initial three amino acids "MCQ" as the antigen for synthesizing HDSP-specific antibodies. Similar to the antibodies targeting HOXA9, those directed against HDSP exhibited no evidence of cross-reactivity (Supplementary Fig. 2e, f). Western blot analysis revealed a significant upregulation of HDSP expression in both GC cells and tissues (Fig. 2a, c). Notably, high expression of HDSP was significantly associated with larger tumor size, advanced clinical stage, and lymph node metastasis in GC patients (Fig. 2e, f and Supplementary Table 1). Collectively, these findings indicate that HOXA10-HOXA9 may encode a microprotein, HDSP, which is naturally expressed in GC tissues and cells. Kaplan–Meier survival analysis, receiver operating characteristic curve analysis, and receiver operating characteristic (ROC) proportional hazard model collectively identified high HDSP expression as an independent factor associated with poor prognosis in GC patients (Fig. 2g–i).

HDSP exerts oncogenic effects in gastric cancer

Currently, the biological functions of HOXA10-HOXA9 and HDSP remain elusive. However, in GC cells, silencing HOXA10-HOXA9 resulted in reduced proliferation, migration, and invasion. Conversely, overexpression of HOXA10-HOXA9 or HDSP led to the opposite effect, while overexpression of HOXA10-HOXA9 mut failed to significantly promote these cellular behaviors (Supplementary Fig. 3a–e). To confirm the functional role of HDSP, rescue experiments were conducted, and the efficiency of HDSP manipulation was assessed in the specified GC cells (Supplementary Fig. 3f). As shown in Fig. 3a, b, HDSP knockdown decreased cell proliferation and colony formation in vitro, which were restored by HDSP re-expression but not by a mutant form of HOXA10-HOXA9 (HOXA10-HOXA9 mut). Similarly, HOXA10-HOXA9 knockdown significantly suppressed tumor



growth, which was rescued by HDSP reintroduction but not by HOXA10-HOXA9 mut (Fig. 3c, d). Consistent results were obtained for tumor volume/weight and the expression of HDSP and Ki67 in the tumor tissues from mice (Fig. 3e, f and Supplementary Fig. 3g). Concurrently, we evaluated the impact of HDSP on the metastatic potential of GC cells. Both in vivo and in vitro functional assays demonstrated that silencing HOXA10-HOXA9 significantly suppressed the metastatic capacity of GC cells. Reexpression of HDSP effectively reversed this suppression, whereas overexpression of a mutant HOXA10-HOXA9 (HOXA10-HOXA9 mut) failed to replicate these effects (Fig. 3g–j). These results indicate that HOXA10-HOXA9 promotes the proliferation and metastasis of GC cells primarily by encoding the small protein HDSP.

HDSP interacts with MECOM and prevents it from ubiquitination-mediated degradation

To investigate the molecular mechanism behind HDSP-mediated GC progression, we co-immunoprecipitated HDSP interactors (Fig. 4a) and identified 413 proteins specifically bound to HDSP using LC/MS-MS analysis (Supplementary Data 2). GO analysis revealed significant enrichment of HDSP-interacting proteins involved in gene expression regulation and mRNA stability et al. (Supplementary Fig. 4a). Among the HDSP interacted candidates, MECOM (MDS1 and EVI1 complex locus) was selected particularly for further investigation, because it is a major member of transcriptional regulator family and has critical roles in gene expression regulation, and dysregulation of MECOM contributes to tumorigenesis and progression of various solid tumors and

Fig. 1 | HOXA10-HOXA9 encodes a microprotein HDSP in GC. **a** Analyzing GC datasets from TCGA, GC-related datasets from GEO, and ribosome sequencing datasets, the differentially expressed lncRNAs in various datasets were intersected to create a Venn diagram. **b** Polysome-bound RNA isolation and qRT-PCR were used to analyze the enrichment of the five candidate lncRNAs on polyribosomes in SGC-7901 cells. Mean \pm SD, $n = 3$ biologically independent experiments. **c** QRT-PCR analysis of HOXA10-HOXA9 expression in 70 paired GC tissues. Mean \pm SD, Paired samples, 2-sided Student's t test. **** $P < 0.0001$. **d** QRT-PCR analysis of HOXA10-HOXA9 expression levels in 7 GC cell lines and normal gastric mucosal epithelial cells (GES-1). Mean \pm SD, Unpaired two-tailed Student's t test. $n = 3$ biologically independent experiments. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. **e** A schematic diagram shows the HOXA10-HOXA9 ORF2 encoding a 112aa small protein. **f** Schematic representation of the structures of HDSP-GFP, HDSP-Flag, and their respective mutant variants. **g** The expression vector depicted in (f) was separately transfected into GC cells. After 48 h, the expression of the HDSP-GFP fusion protein (left) and HDSP-Flag fusion protein (right) were assessed using western blot analysis. The experiment was repeated 3 times independently with similar results. **h** Following transfection of GFP, HDSP-GFP, and their respective

mutant plasmids into GC cells, the green fluorescence emitted by GFP was visualized using a fluorescence microscope. The experiment was repeated 3 times independently with similar results. **i** After transfecting GC cells with HDSP-Flag and its mutant vector, the subcellular localization of the fusion protein was examined using immunofluorescence. The experiment was repeated 3 times independently with similar results. **j** The sequences of HOXA10-HOXA9 were cloned into over-expression vectors containing T7 promoters, and their coding potential was validated through in vitro translation experiments. The experiment was repeated 3 times independently with similar results. **k** The schematic diagram illustrates the Flag Knock-In Strategy. The insertion of the Flag tag after the start codon of HDSP represents a fusion expression. **l** Western blot was employed to ascertain the successful integration of the Flag tag into GC cells. The experiment was repeated 3 times independently with similar results. **m** After introducing Flag via knock-in in GC cells, HOXA9 was subsequently knocked out, followed by the assessment of HOXA9 and HDSP expression levels using western blot analysis. The experiment was repeated 3 times independently with similar results. Source data are provided as a Source data file.

hematopoietic cancers (Supplementary Fig. 4b)^{18,19}. Kaplan–Meier analysis demonstrated that GC patients with higher MECOM expression had poorer overall survival (Supplementary Fig. 4c). In addition, the binding between MECOM and HDSP was confirmed by reciprocal Co-IP (Fig. 4b). The PLA revealed the co-localization of MECOM and HDSP in GC cells, and their spatially bound conformation was analyzed using swMATH software (Fig. 4c, d). MECOM consists of two DNA-binding domains with zinc finger motifs, a repression domain, and an acidic domain²⁰. Co-IP experiments revealed that the repression domain (427–912 aa) of MECOM is essential for the interaction with HDSP. In addition, the C-terminal region of HDSP, rather than the N-terminal region, retained the ability to interact with MECOM (Fig. 4e, f). The in vitro GST pull-down assay also confirmed the interaction between HDSP and MECOM (Fig. 4g).

Notably, we observed that manipulation of HDSP only affected the protein level of MECOM but not its mRNA level (Fig. 4h and Supplementary Fig. 4d). A CHX chase assay confirmed that HDSP knockdown decreased the half-life of MECOM protein, whereas HDSP overexpression had the opposite effect (Fig. 4i and Supplementary Fig. 4e). HDSP appears to regulate MECOM expression through posttranslational mechanisms. The degradation of proteins is primarily mediated by two pathways: the ubiquitin-proteasome system (UPS) and the lysosome pathway²¹. Interestingly, the decrease in MECOM protein caused by downregulation of HDSP could be saved by the proteasome inhibitor MG132 rather than the lysosome inhibitors Chloroquine (CQ). In addition, MG132 but not CQ enhanced MECOM protein stability in both HDSP-Flag-expressing and control cells (Fig. 4j). We investigated whether HDSP regulates MECOM expression via the UPS and found that ubiquitinated MECOM levels increased with the presence of MG132, as shown by Co-IP analysis (Fig. 4k). K48-linked polyubiquitination and K63-linked polyubiquitination are the two main types of protein ubiquitination; the former leads to proteasomal degradation and the latter is generally involved in protein translocation²². Importantly, HDSP overexpression decreased K48-linked ubiquitination of MECOM, whereas knockdown of HDSP increased MECOM ubiquitination. Both exogenous ubiquitination and endogenous ubiquitination of MECOM were reduced by HDSP ectopic expression in GC cells (Fig. 4l–n). These results suggest that HDSP interacts with MECOM and protects it from K48-linked ubiquitination-mediated proteasomal degradation, thereby maintaining its protein stability.

TRIM25 interacts with MECOM to mediate its ubiquitination which could be competitively blocked by HDSP

With the assistance of a series of enzymes, ubiquitination marks specific proteins for degradation, with E3 ubiquitin ligases

responsible for target specificity. Our previous results confirm that the MECOM protein can undergo ubiquitination and subsequent degradation via the UPS. Co-IP/MS analysis identified TRIM25 as an E3 ligase, which exhibited the highest number of peptide ions matched to MECOM among the 11 interacting E3 ubiquitin ligases (Fig. 5a, b, Supplementary Fig. 4f, and Supplementary Data 2). The reciprocal Co-IP and GST pull-down assays confirmed the exogenous interaction between TRIM25 and MECOM in GC cells (Fig. 5c, d). The deletion of TRIM25 resulted in a significant increase in MECOM protein abundance but did not affect its mRNA levels (Fig. 5e and Supplementary Fig. 4g). In addition, the deletion of TRIM25 also prolonged the half-life of endogenous MECOM protein (Fig. 5f). The level of ubiquitinated MECOM was markedly reduced when TRIM25 was silenced, particularly in the presence of MG132 (Fig. 5g). To investigate potential ubiquitination modification sites on MECOM, we introduced the K656R mutant predicted by PhosphoSitePlus software. Western blot analysis showed that the protein level of wild-type MECOM changed with TRIM25 alteration, whereas the level of the MECOM-K656R mutant remained unchanged (Supplementary Fig. 4h). Moreover, the K656R mutation significantly eliminated the ubiquitin smear of MECOM (Supplementary Fig. 4i). Consistent with this, TRIM25 could induce ubiquitination of wild-type MECOM but not of MECOM-K656R (Supplementary Fig. 4j). As determined by the E2Select Ubiquitin Conjugation Assay, in the presence of E2 ligase UBE2D1, TRIM25 showed superior efficacy in conjugating ubiquitin to MECOM (Fig. 5h).

Interestingly, further Co-IP experiments showed that the interaction between MECOM and TRIM25 depended on MECOM's repression domain (Fig. 5i), which is also responsible for the physical binding to HDSP. Therefore, HDSP may compete with TRIM25 for MECOM binding, as evidenced by the weakened TRIM25-MECOM association with increasing HDSP levels and the strengthened association upon HDSP knockdown (Fig. 5j, k). The PLA demonstrated that TRIM25 interacts with MECOM subsequent to interference with HDSP (Fig. 5l). Interference with TRIM25 reversed the shortened half-life of MECOM protein induced by the knockdown of HDSP expression (Supplementary Fig. 4k). Collectively, these results suggest that TRIM25 is a genuine E3 ligase of MECOM in GC cells, and HDSP disrupts TRIM25-MECOM binding, which ultimately affects MECOM ubiquitination and abundance.

HDSP triggers the malignant phenotype of gastric cancer cells partially via MECOM

Previous studies have shown that MECOM plays an oncogenic role in various cancers, including in lung and ovarian cancers^{19,23}. We investigated whether MECOM is involved in GC progression by examining its

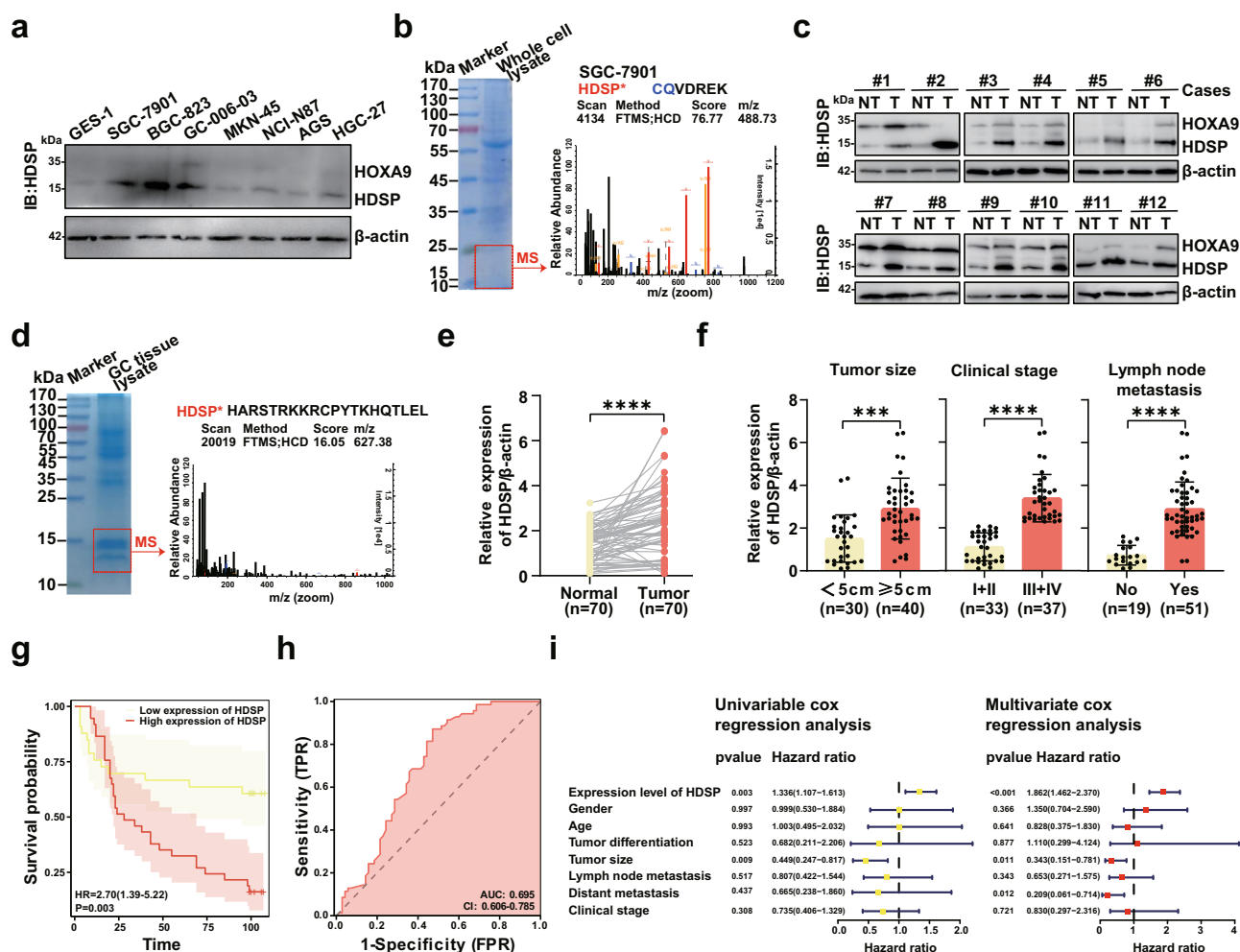


Fig. 2 | High levels of endogenous HDSP correlate with poor prognosis for GC patients. **a** Western blot detected endogenous HDSP expression in seven cell lines and normal gastric mucosal epithelial cells (GES-1). The experiment was repeated 3 times independently with similar results. **b** SGC-7901 cell protein was separated using SDS-PAGE and stained with Coomassie Blue. Specific bands corresponding to HDSP were identified by MS. **c** Western blot detection of HDSP expression in 70 pairs of GC patients (12 pairs were shown). The experiment was repeated 3 times independently with similar results. **d** Endogenous HDSP in GC tissue was detected by MS after Coomassie brilliant blue staining. **e** The grayscale values of HDSP/ β -actin in 70 pairs of GC tissues were quantitatively analyzed. Paired samples, 2-sided Student's *t* test. *****P* < 0.0001. **f** Combined with the clinicopathological data

analysis of 70 patients with GC, the expression of HDSP was examined in relation to tumor size, stage, and lymph node metastasis. Mean \pm SD, 2-sided Student's *t* test. ****P* < 0.001, *****P* < 0.0001. **g** 70 GC patients were stratified into two groups based on the expression levels of HDSP, n(HDSP Low) = 33, n(HDSP High) = 37, and Kaplan-Meier analysis was employed to plot the overall survival curve. A log-rank test. **h** Plotting the GC patients' Receiver Operating Characteristic (ROC) curve according to the expression of HDSP. Data are presented as AUC \pm 95% CI. AUC area under the curve, CI confidence interval. **i** HDSP expression correlated with the prognosis of GC patients, univariate and multivariate COX analysis, *n* = 70. NT non-tumor, T tumor. Source data are provided as a Source data file.

expression in GC specimens. IHC analysis revealed that MECOM is highly expressed in cancerous tissues and is positively correlated with tumor size, advanced stage, lymph node metastasis, and poor overall survival in GC patients (Supplementary Fig. 5a-c). Furthermore, the depletion of MECOM impaired cell viability, clonogenicity, migration, and invasion in GC cells (Supplementary Fig. 5d-g). In addition, we performed rescue assays to investigate whether HDSP exerts its oncogenic effects via MECOM, which confirmed that the effects of HDSP on malignant GC cell phenotypes are, to some extent, reliant on MECOM (Supplementary Fig. 6a-d).

SPINK1 is a transcriptional target co-regulated by the HDSP/MECOM axis

Although MECOM has been implicated in various human malignancies as a prominent transcription factor²⁴, its downstream targets in GC are still to be identified. We conducted RNA-seq analysis and identified 2947 upregulated and 2937 downregulated genes upon MECOM knockdown (Supplementary Fig. 6e and Supplementary

Data 2). GO and KEGG pathway analysis showed that these genes were enriched in the MAPK, PI3K-AKT, and Ras signaling pathways et al. (Fig. 6a). As HDSP is responsible for MECOM protein abundance, we focused on the genes co-regulated by the HDSP/MECOM axis at the transcriptional level. Analysis of RNA-seq data from HDSP and MECOM knockdown conditions, along with MECOM ChIP-seq data from the Cistrome Data Browser database, revealed 17 target genes (Fig. 6b and Supplementary Data 2), among which eight candidates were validated, with SPINK1 being the top-ranked downregulated gene in HDSP or MECOM-silenced GC cells (Fig. 6c, d and Supplementary Fig. 6f). Serine protease inhibitor Kazal type 1 (SPINK1), also known as pancreatic secreted trypsin inhibitor (PSTI) or tumor-associated trypsin inhibitor (TATI), inhibits pancreatic trypsin activation, preventing organ damage and regulating inflammation and cell proliferation. Its overexpression promotes tumor progression in various cancers like prostate, colon, pancreatic, and ovarian cancers, correlating with poorer prognoses²⁵. Moreover, SPINK1 promotes cancer by activating EGFR signaling due to its

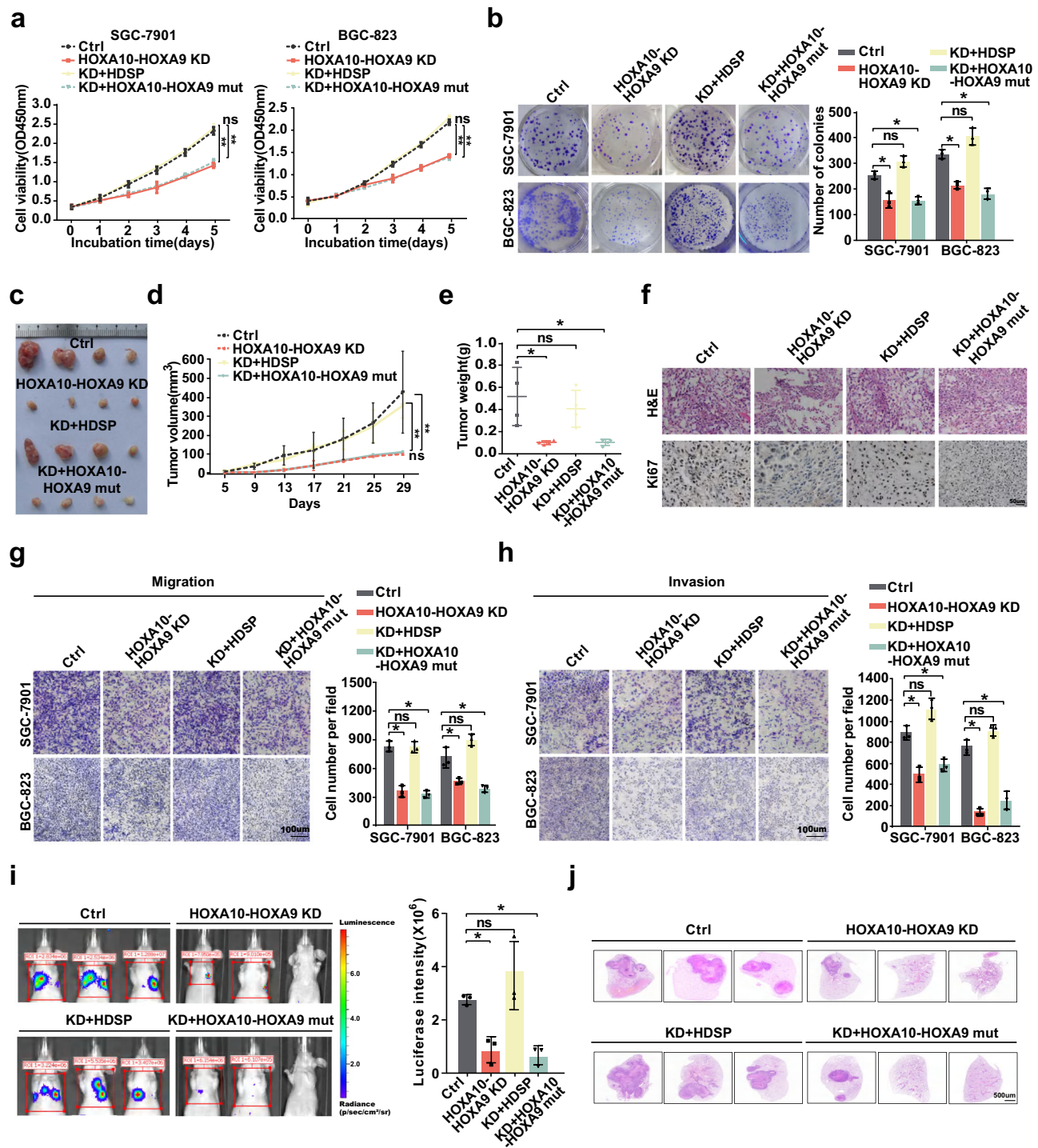
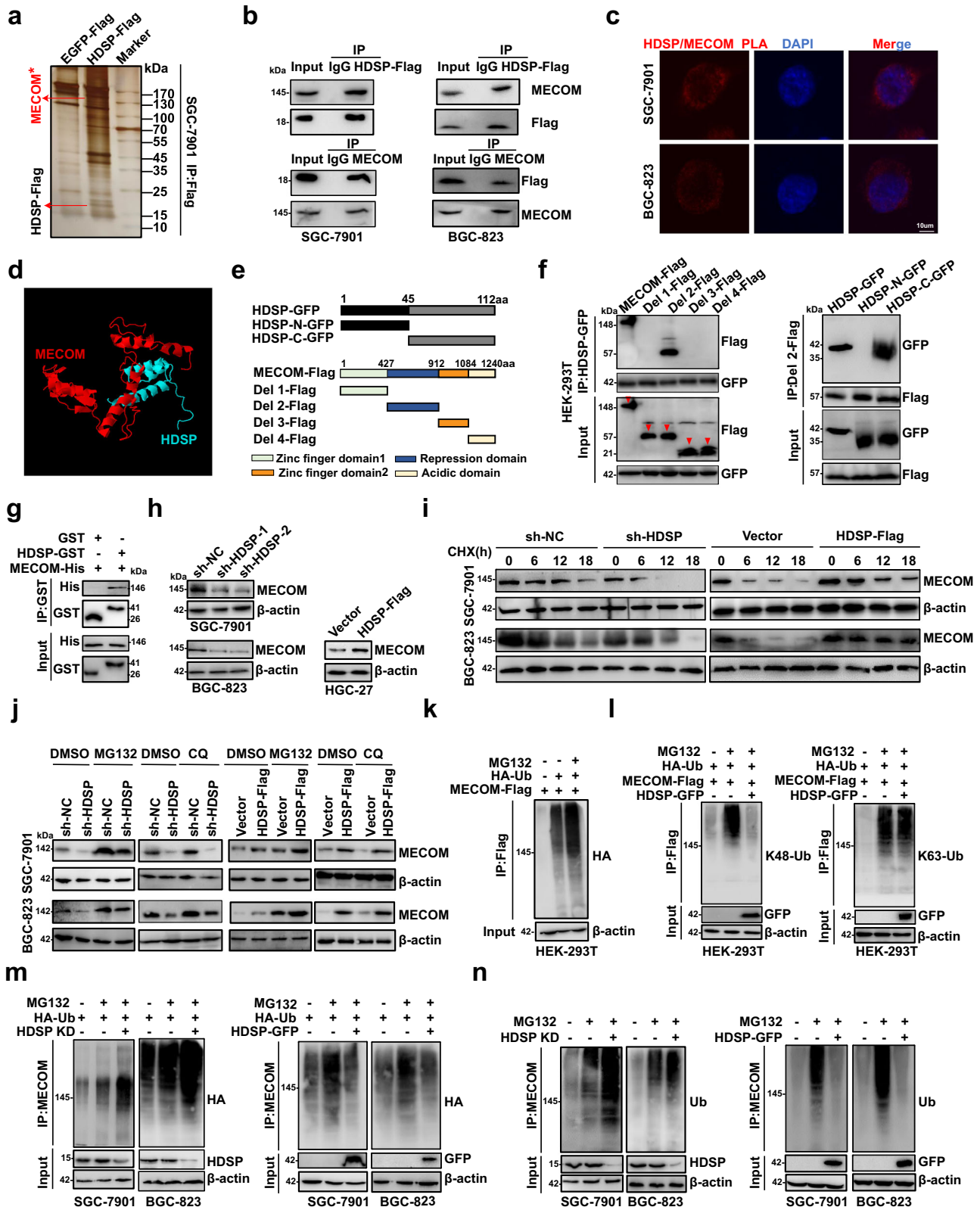


Fig. 3 | HDSP exerts oncogenic effects in gastric cancer. a, b The HDSP or HOXA10-HOXA9 mut expression plasmids were reintroduced into GC cells with stable low HDSP levels. Thus, the GC cells were divided into four groups: Ctrl, HOXA10-HOX9KD, KD + HDSP, and KD + HOXA10-HOXA9 mut. Proliferative activity of above four groups of GC cells were evaluated using CCK-8 (a) and colony formation (b) assays. Mean \pm SD, two-tailed Student's *t* test, *n* = 3 biologically independent experiments. ns, not significant; **P* < 0.05, ***P* < 0.01. **c** Subcutaneous injection of above four groups GC of cells into nude mice was performed and representative images of tumors in xenograft model were shown. **d** Absolute tumor volume presented as a mean \pm SD, *n* = 4 biologically independent animals. ns, not significant; ***P* < 0.01. **e** Tumor weight was measured after 29 days.

Mean \pm SD, two-tailed Student's *t* test, *n* = 4. ns, not significant; **P* < 0.05. **f** Representative H&E staining of xenograft tissues and IHC staining of Ki67 are shown. Scale bar: 50 μ m. **g, h** Transwell assays were conducted to analyze the migration and invasion ability of above four groups of GC cells. Scale bar: 100 μ m. Mean \pm SD, two-tailed Student's *t* test. *n* = 3 biologically independent experiments. ns, not significant; **P* < 0.05. **i** BLI showing lung metastasis in vivo. Mean \pm SD, two-tailed Student's *t* test. *n* = 3 biologically independent animals. ns, not significant; **P* < 0.05. **j** H&E staining to detect the number and size of nodule production in the lung metastasis model. Scale bar: 500 μ m. NC negative control, KD knock down. Source data are provided as a Source data file.



structural similarity with EGF^{26,27}, which is consistent with the downstream signaling pathways regulated by MECCOM. Therefore, we selected SPINK1 for further investigation.

To confirm the regulatory effects of MECCOM on SPINK1, a putative MECCOM-binding motif was predicted in the *SPINK1* promoter region (-981 to -966) (Fig. 6e). Luciferase reporter assays indicated that forced MECCOM expression significantly increased the activity of the

wild-type *SPINK1* promoter but not the binding site mutant promoter in GC cells (Fig. 6f). ChIP-PCR experiments further confirmed that MECCOM is physically bound to the *SPINK1* promoter in GC cells, with binding efficacy reduced in MECCOM knockdown cells and enhanced in MECCOM-overexpressing cells (Fig. 6g, h). MECCOM overexpression restored the decreased SPINK1 protein levels induced by HDSP knockdown (Supplementary Fig. 6a). MECCOM is a transcription factor

Fig. 4 | HDSP interacts with MECOM and prevents it from ubiquitination-mediated degradation. **a** In SGC-7901 cells with stable overexpression of HDSP-Flag, the IP enrichment of the target protein interacting with HDSP was assessed by silver staining. **b** Bidirectional Co-IP was employed to investigate the interaction between HDSP and MECOM in GC cells. **c** The interaction between HDSP and MECOM in GC cells was detected by PLA. The experiment was repeated 3 times independently with similar results. Scale bar: 10 μm . **d** The swMATH software provides the most likely docking method for HDSP and MECOM. **e** Schematic diagram of HDSP and MECOM truncation. HDSP is evenly divided into N-terminal and C-terminal segments, while MECOM is divided into four segments corresponding to its four domains. **f** The binding sites between HDSP and MECOM were verified through Co-IP experiments using their different truncated forms. **g** After the expression and protein purification of MECOM-His and HDSP-GST, the

interaction between HDSP and MECOM in vitro was detected using a GST pull-down assay. **h** Western blot analysis was conducted to assess MECOM expression levels in GC cells with stable low or overexpression of HDSP. **i** GC cells were treated with cycloheximide (CHX, 10 μM), and MECOM protein levels were assessed by western blot at 0, 6, 12, and 18 h. **j** GC cells were treated with either MG132 (10 mM) or CQ (20 μM) for 6 h, following which the expression levels of MECOM were evaluated using western blot. **k** After co-transfection of HA-Ub and MECOM-Flag, HEK-293T cells were treated with MG132, and the ubiquitination level of MECOM was assessed using the Co-IP. **l** Co-IP was employed to identify the type of ubiquitination modification of MECOM mediated by HDSP. **m, n** Co-IP was utilized to assess the ubiquitination modification level of MECOM following knockdown or overexpression of HDSP. All western blot assay was repeated 3 times independently with similar results. Source data are provided as a Source data file.

known to interact with Activator protein-1 (AP-1), comprising c-FOS and c-JUN, to activate downstream target genes^{28,29}. Our MS results indicated that MECOM is associated with c-JUN in GC cells (Supplementary Data 2), and c-FOS/c-JUN silencing also downregulated SPINK1 expression (Fig. 6i). ChIP-PCR analysis showed that c-FOS and c-JUN, along with MECOM, are enriched in the *SPINK1* promoter region (Fig. 6j). Subsequently, double luciferase reporter gene assays confirmed that c-FOS, c-JUN, and MECOM enhanced *SPINK1* promoter activity (Fig. 6k), suggesting that MECOM interacts with the AP-1 complex to activate *SPINK1* transcription. Spearman analysis revealed a significant correlation between SPINK1 expression and MECOM or HDSP levels (Fig. 5l). Collectively, these results indicate that SPINK1 is a downstream target of the HDSP/MECOM axis, showing that the oncogenic effects of the HDSP/MECOM axis in GC may be mediated by the transcriptional activation of SPINK1.

SPINK1 promotes the progression of gastric cancer by activating EGFR signaling

The roles of SPINK1 in various types of cancers have been well documented³⁰. However, whether SPINK1 is involved in GC remains unclear. Here, we evaluated SPINK1 protein levels in a tissue microarray comprising 90 pairs of GC and normal tissues. IHC analysis showed that SPINK1 was upregulated in 81% (73/90) of primary GC tissues (Fig. 6m and Supplementary Fig. 6g). Furthermore, high SPINK1 expression was significantly associated with a higher probability of cancer progression (Supplementary Fig. 6h). Kaplan-Meier analysis demonstrated that patients with low SPINK1 expression (based on a median cut-off score) had better overall survival and disease-free survival than those with high SPINK1 expression (Fig. 6n and Supplementary Fig. 6i). Subsequently, the functional significance of SPINK1 was further investigated. SPINK1 deletion was found to repress cell proliferation, colony formation, migration, and invasion in vitro (Supplementary Fig. 7a–d). To explore the potential regulatory mechanism of SPINK1, Co-IP assays were conducted to confirm that SPINK1 interacts with EGFR in GC cells (Fig. 6o). In addition, HDSP knockdown reduced the protein levels of p-EGFR, p-STAT3, p-AKT, and p-ERK, which were rescued by SPINK1 overexpression (Fig. 6p and Supplementary Fig. 7e). Mutations within the ubiquitination site of MECOM serve to enhance the activation of the EGFR signaling pathway (Supplementary Fig. 7f). A similar trend was observed for p-EGFR in xenograft tumor tissues from mice (Supplementary Fig. 7g). Subsequent cellular function experiments revealed that the employment of the EGFR inhibitor AG490 effectively suppressed cell proliferation, migration, and invasion induced by the overexpression of HDSP or MECOM (Supplementary Fig. 7h–j). The targeted inhibition of EGFR demonstrates a similar impact on the proliferation and metastasis of gastric cancer cells as the targeted knockdown of HDSP (Supplementary Fig. 8a–c). Overall, these findings suggest that HDSP promotes GC progression mainly through SPINK1-mediated activation of the EGFR signaling pathway.

MECOM/HOXA10-HOXA9-HDSP/MECOM form a positive feedback regulatory circuit

As indicated above, we demonstrated the oncogenic role of HDSP in GC and aimed to determine the factors responsible for its abnormal expression. Analysis of the *HOXA10-HOXA9* promoter region revealed a putative MECOM-binding site upstream of the transcription start site at –165 to –150 bp, suggesting that MECOM may contribute to the transcriptional activation of HOXA10-HOXA9 (Fig. 7a). Interestingly, MECOM and AP-1 deficiency downregulated HOXA10-HOXA9 and HDSP protein levels (Fig. 7b, c). Furthermore, ChIP-PCR assays demonstrated that MECOM was enriched in the binding region of the *HOXA10-HOXA9* promoter, and its binding efficiency was attenuated in MECOM knockdown cells, whereas it was enhanced in MECOM-overexpressing cells (Fig. 7d, e). Similarly, c-FOS and c-JUN were also enriched in the *HOXA10-HOXA9* promoter (Fig. 7f). Furthermore, luciferase reporter assays confirmed that MECOM, c-FOS, and c-JUN could bind to the *HOXA10-HOXA9* promoter and initiate luciferase transcription, as expected (Fig. 7g, h). Collectively, our findings suggest that the MECOM protein, stabilized by HDSP, interacts with AP-1 to activate transcription of the *HOXA10-HOXA9* gene, which encodes HDSP, thereby forming a regulatory feedback circuit to activate downstream SPINK1-EGFR signaling, which facilitates GC progression (Fig. 7i).

HDSP is a valuable therapeutic target and neoantigen for gastric cancer vaccine development

Recent studies identified the micropeptide ASAP and the protein circPLCE1-411 as potential therapeutic targets for cancer treatment using CRISPR-Cas9 in vivo^{31,32}. To investigate the potential of HDSP as a therapeutic target, we utilized a PDX model of GC. However, due to the high sequence homology between the coding sequences of HDSP and HOXA9, a unique sgRNA for HDSP could not be designed. Thus, we employed an shRNA-based strategy to target HDSP in the PDX model (Fig. 8a). After 4 weeks, we observed a significant reduction in the volume and weight of subcutaneous tumors in mice treated with shRNA-HDSP compared to the control group (Fig. 8b–d). The successful establishment of a PDX model was confirmed by H&E staining, and the knockdown efficiency of HDSP was validated (Supplementary Fig. 8d, e).

Recent studies suggested that lncRNAs may encode immunogenic peptides that can serve as a cancer vaccine, as many MHC class I-associated peptides have been found to originate from small ORFs within these genes³³. Weissman et al. identified 240 noncanonical peptides through MS-based HLA-I peptidomics, suggesting their potential immunogenicity³⁴. In addition, Perreault et al. found that tumor-specific antigens were primarily derived from non-coding regions, supporting this notion³⁵. Furthermore, Song et al. discovered that micropeptides encoded by circFAM53B effectively activate initial CD4⁺ and CD8⁺ T cells in an antigen-specific manner, leading to the induction of anti-tumor immunity. To investigate the

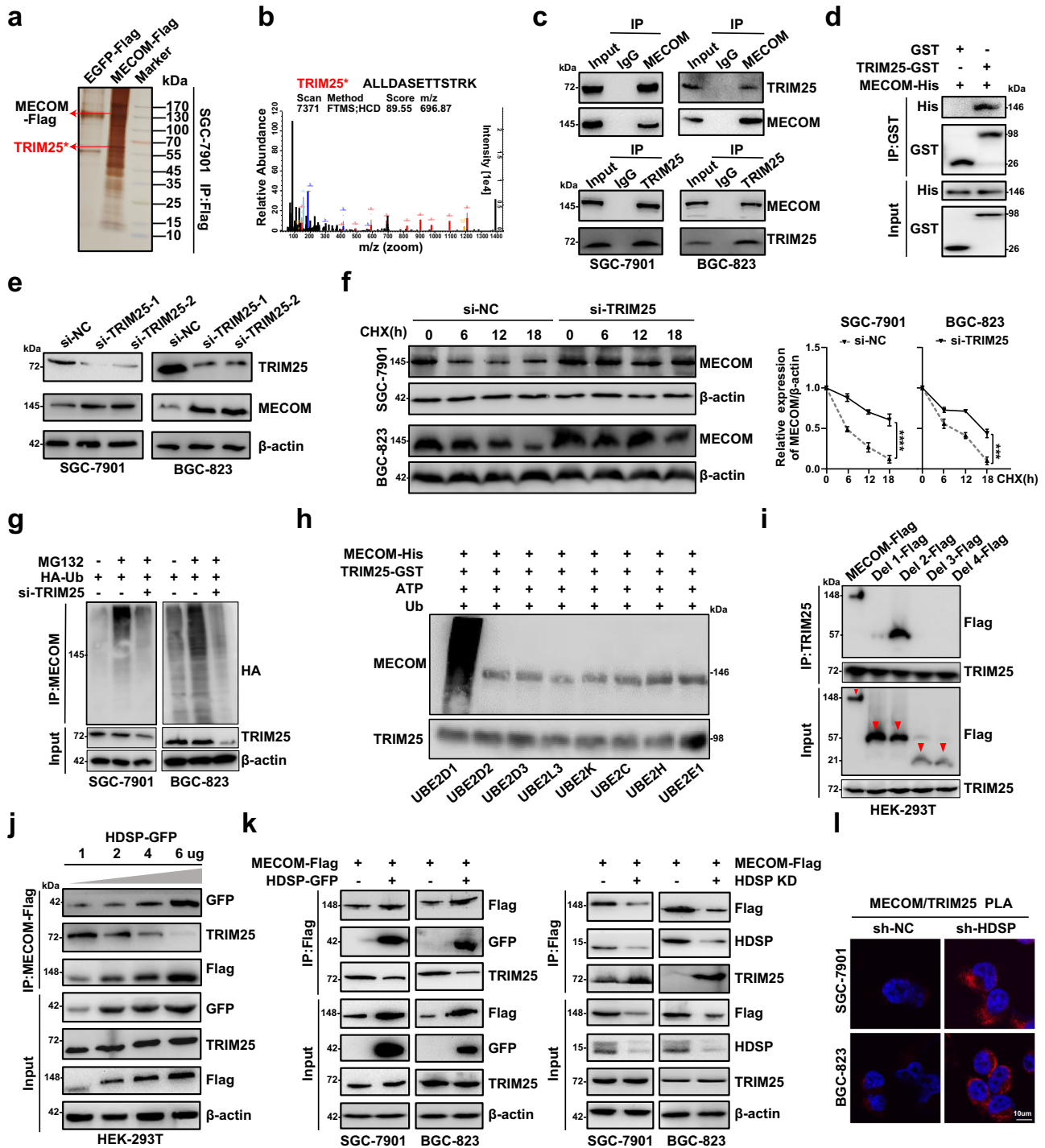
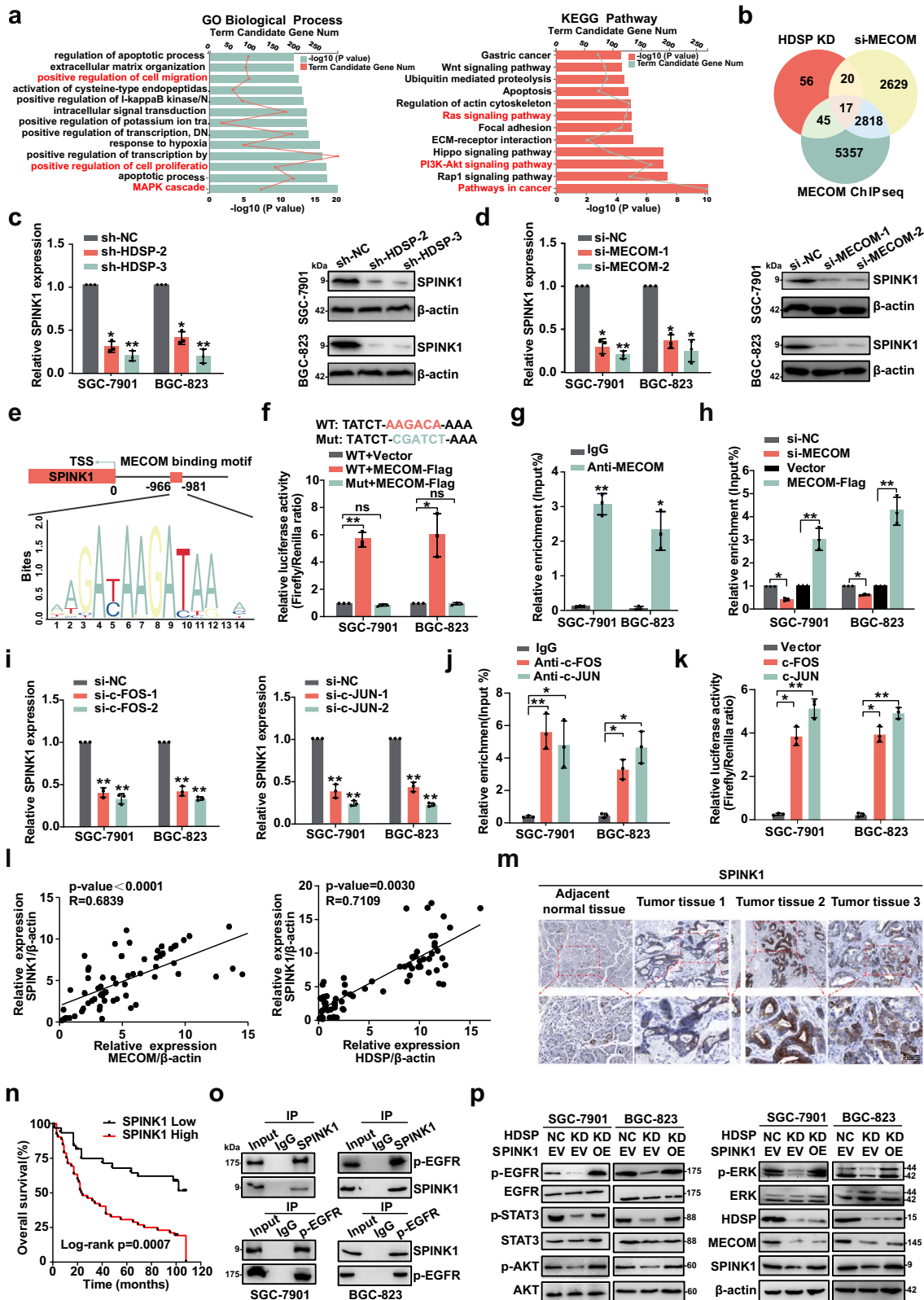


Fig. 5 | TRIM25 interacts with MECOM to mediate its ubiquitination which could be competitively blocked by HDSP. a In SGC-7901 cells with stable over-expression of MECOM-Flag, the proteins interacting with MECOM were enriched using IP, and the enrichment efficiency was evaluated by silver staining. **b** A mass spectrum displaying a unique peptide of TRIM25 identified in MECOM-IP/MS results. **c** Bidirectional Co-IP was performed to verify the interaction of MECOM with TRIM25 in GC cells. The experiment was repeated 3 times independently with similar results. **d** After the expression and protein purification of MECOM-His and TRIM25-GST, the interaction between TRIM25 and MECOM in vitro was detected using a GST pull-down assay. **e** Western blot analysis was performed to assess the levels of MECOM protein expression in two different TRIM25-depleted GC cells. **f** GC cells with TRIM25 knockdown were treated with CHX (10 μ M), and protein samples were collected at 0, 6, 12, and 18 h for western blot analysis to determine MECOM expression levels. Mean \pm SD, two-tailed Student's *t* test. *n* = 3 biologically

independent experiments. ****P* < 0.001, *****P* < 0.0001. **g** Following the knockdown of TRIM25, Co-IP was utilized to assess the level of ubiquitination modification of MECOM in GC cells. **h** After expressing and purifying MECOM-Flag and TRIM25-GST, E2 binding enzymes potentially involved in MECOM ubiquitination were identified using the E2select Ubiquitin Conjugation Assay. **i** After co-transfecting HEK-293T cells with four truncated MECOM expression plasmids and TRIM25 expression plasmids, the regions where TRIM25 and MECOM interacted were detected using Co-IP. **j** Co-IP was utilized to determine whether the competitive binding of HDSP and TRIM25 to MECOM is dose-dependent on the quantity of HDSP. **k** Co-IP was utilized to ascertain the interaction between MECOM and TRIM25 following either the knockdown or overexpression of HDSP. **l** The interaction between TRIM25 and MECOM following HDSP knockdown in GC cells was assessed using a PLA. All experiments were repeated 3 times independently with similar results. Source data are provided as a Source data file.



immunogenicity of HDSP, we synthesized six 20-amino-acid peptides as vaccines (Fig. 8e) and co-cultured them with GC cells and patient-derived cells along with PBMCs. The LDH release assay and ELISpot assay showed that HDSP peptide-3 (41–60 aa) induced significant immunological activity and boosted the killing efficacy of PBMCs against GC cells (Fig. 8f–h). Furthermore, MS analysis of HLA peptidomics verified that peptide-3 can be presented on the cell membrane

by HLA-I, suggesting that HDSP peptide-3 could serve as a neoantigen for GC vaccine development (Fig. 8i). To confirm the influence of HDSP peptide-3 on in vivo immune responses, we administered the peptide to an HSC humanized mouse model via tail vein injection twice a week for 3 weeks. The results showed a significant suppression of tumor growth and weight in mice treated with HDSP peptide-3 (Fig. 8j–m and Supplementary Fig. 8f–i). Furthermore, administration of HDSP

Fig. 6 | SPINK1 is a transcriptional target co-regulated by HDSP/MECOM axis. **a** After knocking down MECOM in SGC-7901 cells, RNA-seq identified differentially expressed genes for GO and KEGG pathway analysis. **b** RNA sequencing in SGC-7901 cells, post HDSP knockdown or MECOM interference, intersected with MECOM's ChIP-seq data to generate a Venn diagram. **c, d** The expression of SPINK1 were examined by qRT-PCR and western blot upon HDSP or MECOM knockdown. Two-way ANOVA analysis. * $P < 0.05$, ** $P < 0.01$. **e** Prediction of MECOM binding site in the promoter region of SPINK1 by JASPAR. **f** The effect of MECOM on SPINK1 promoter activity was tested using a dual luciferase reporter assay with wild-type and mutant plasmids containing predicted binding sites. Two-way ANOVA analysis. ns, not significant; * $P < 0.05$, ** $P < 0.01$. **g, h** The enrichment of MECOM binding at the SPINK1 promoter was assessed using ChIP-PCR. Two-sided Student's t test. * $P < 0.05$, ** $P < 0.01$. **i** The RNA expression of SPINK1 protein was measured after the knockdown of AP-1 complex. Two-way ANOVA analysis. * $P < 0.05$, ** $P < 0.01$. **j** ChIP-PCR assay detected the enrichment of c-FOS and c-JUN in the promoter region of SPINK1. Two-sided Student's t test. * $P < 0.05$, ** $P < 0.01$. **k** The transcriptional

activity of c-FOS and c-JUN on the SPINK1 promoter was evaluated using a dual luciferase reporter gene assay. One-way ANOVA analysis. * $P < 0.05$, ** $P < 0.01$. **l** A linear regression model was used to assess the correlation between MECOM abundance and SPINK1, HDSP, and SPINK1 proteins in 70 pairs of GC tissues. Two-sided chi-square test. **m** Representative image of SPINK1 immunohistochemical staining in 90 paired GC tissue microarrays. Scale bar: 50 μm . **n** GC patients were categorized into 2 groups based on SPINK1 expression, n(SPINK1 Low) = 31, n(SPINK1 High) = 53, and overall survival curves were generated using Kaplan-Meier analysis. **o** Bidirectional Co-IP verifies the interaction between SPINK1 and EGFR in GC cells. The experiment was repeated 3 times independently with similar results. **p** Western blot was used to measure EGFR phosphorylation levels and downstream signaling molecules after either HDSP knockdown or SPINK1 overexpression. The experiment was repeated 3 times independently with similar results. sh-NC, control shRNA; EV, empty vector; OE, overexpression. Data are means \pm SD, $n = 3$ biologically independent experiments. Source data are provided as a Source data file.

peptide-3 resulted in enhanced T cell activation and infiltration into tumor lesions (Fig. 8n–p and Supplementary Fig. 8j–l). In summary, our findings suggest that HDSP could serve as a valuable target for GC treatment and as a vaccine antigen resource.

Discussion

Multi-omics analyses have identified functional peptides/proteins encoded by non-coding RNA transcripts. An increasing body of evidence reveals that specific lncRNAs, such as APPLE, can promote cancer by serving as a component of the translation machinery¹⁷. The non-coding RNA CTD-2256P15.2 encodes the micropeptide PACMP, which sustains CtIP protein levels and activates PARYlation signaling during DNA repair³⁶. The micropeptide ASAP, which is derived from LINC00467 in colorectal cancer, promotes ATP production, oxygen consumption rate, and cell proliferation³¹. Although there have been no previous reports of lncRNA-encoded peptides/proteins in GC, our study identified a noncanonical ORF with coding ability in the lncRNA HOXA10-HOXA9, which is endogenously translated into a micro-protein, HDSP, in GC cells and tissues. Using RNA-seq, Ribo-seq, and MS data and biological approaches, we demonstrated that HDSP is highly expressed in GC and promotes its progression by preventing the oncogenic transcription factor MECOM from undergoing TRIM25-mediated ubiquitination and degradation.

MECOM has been extensively studied in acute myeloid leukemia, and its dysregulation has been shown to remodel the metabolic landscape of leukemogenesis²⁰. Recently, aberrant upregulation of MECOM has been observed in prostate cancer³⁷, lung squamous cell carcinoma, and ovarian cancer^{19,24}. The results of the present study suggest that high MECOM expression is associated with poor GC prognosis and could be a prognostic biomarker. However, the regulatory factors controlling MECOM abundance in GC pathogenesis are poorly understood. Although previous research has suggested that MECOM is a potential target of ubiquitin-proteasomal degradation, the specific E3 ligase involved in this process has yet to be identified³⁸. Through Co-IP/MS analysis, we discovered that TRIM25 is a critical E3 ligase that interacts with MECOM and catalyzes ubiquitination on MECOM's K656 site. Our data suggest that HDSP competes with TRIM25 to block its interaction with MECOM, preventing ubiquitination-mediated degradation of MECOM protein. Our results demonstrate that HDSP/MECOM are primarily responsible for triggering the malignant phenotype of GC cells. These findings elucidate the mechanism behind the oncogenic effects of HDSP and its role in the pathogenesis of GC.

Recent evidence has confirmed MECOM's role as a driver of malignant transformation; however, its downstream targets and regulatory mechanisms remain unclear. MECOM contains two DNA-binding zinc finger domains, which it uses to modulate the transcription of target genes through cis-binding. For example, MECOM

mediates the transcriptional repression of tumor-suppressive genes, such as *PTEN*, by recruiting polycomb repressor complex 2 (PRC2)³⁹, the H3K9me3 methyltransferase SUV39H1⁴⁰, and the de novo DNA methyltransferases DNMT3A and 3B⁴¹. Interestingly, MECOM's interaction with histone acetyltransferases (CBP and P/CAF) leads to its transition from a transcription repressor to an activator⁴². Furthermore, MECOM physically works in concert with AP-1 subunits identified by SILAC-MS screening, and 65% of their target genes are transcriptionally activated^{28,29}. This highlights that MECOM is an essential transcriptional activator. In our study, we found that MECOM primarily functions as a transcriptional activator in GC cells, as most of the HDSP/MECOM co-regulated genes are downregulated upon silencing of MECOM or HDSP. We speculate that HDSP resides temporarily in the repressive domain of MECOM, which may restrict its repressive transcriptional function. However, this hypothesis requires further experimental verification.

Among their target genes, SPINK1 is a prominent downstream target of MECOM and HDSP among these genes. SPINK1 is upregulated in various cancers and promotes tumor occurrence and progression by functioning as a ligand that interacts with EGFR, which leads to receptor dimerization and phosphorylation due to its structural similarity to EGF^{30,43}. As a downstream target of MECOM and HDSP, SPINK1 has an oncogenic impact on GC cells. Highly expressed SPINK1 is associated with a shorter progression-free survival of GC patients; SPINK1 activates the PI3K/AKT and MAPK/ERK signaling pathways to promote a tumorigenic phenotype. We also discovered a positive feedback loop involving HOXA10-HOXA9/HDSP/MECOM, in which MECOM acts downstream of HDSP and upstream of HOXA10-HOXA9, continuously activating SPINK1-EGFR signaling to accelerate GC progression.

ncRNA-derived proteins or peptides, including ASAP³¹ and circPCE1-411³², have been identified as potential therapeutic targets for cancer treatment. We discovered that in vivo knockdown of HDSP could induce tumor remission in a PDX model, which suggests that HDSP may be a valuable treatment target for GC. In addition, Thanugue et al. recently reported that peptides derived from lncRNAs can have immunogenic effects and can be used as tumor-specific antigens to boost the anti-tumor immune response³³. For example, MELOE-1, a particular melanoma antigen, is produced by internal ribosome entry site (IRES)-dependent translation of lncRNA and plays a crucial role in T cell immunosurveillance, which makes it a promising target for melanoma immunotherapy⁴⁴. Remarkably, immunogenic HDSP peptide-3 elicited a robust anti-tumor response. We further demonstrated that HDSP peptide-3 effectively induces tumor regression and CD3⁺ T cell infiltration in a GC PDX model through co-infusion with PBMCs. These results underscore the potential of HDSP as a therapeutic target and a neoantigen candidate for GC vaccine development.

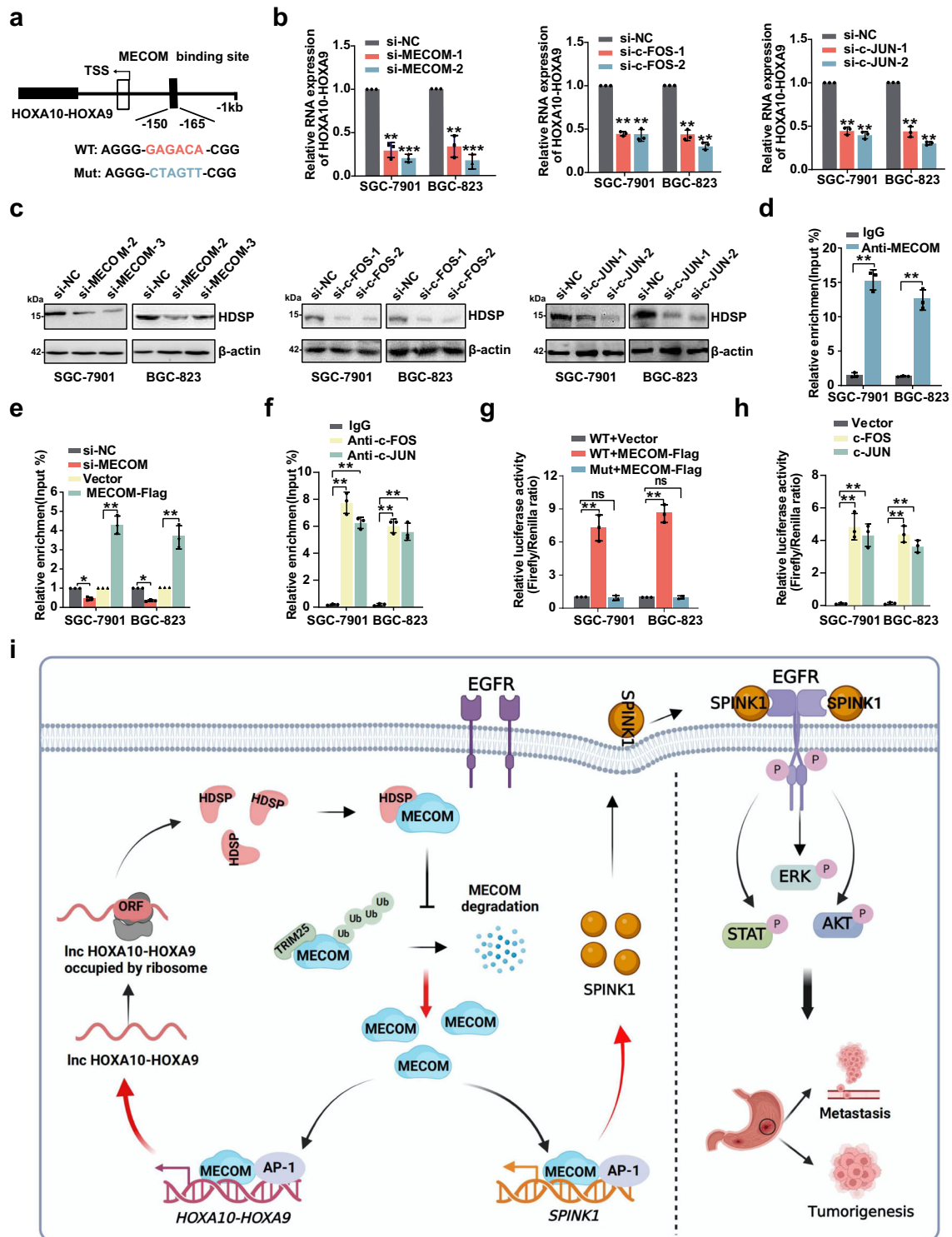
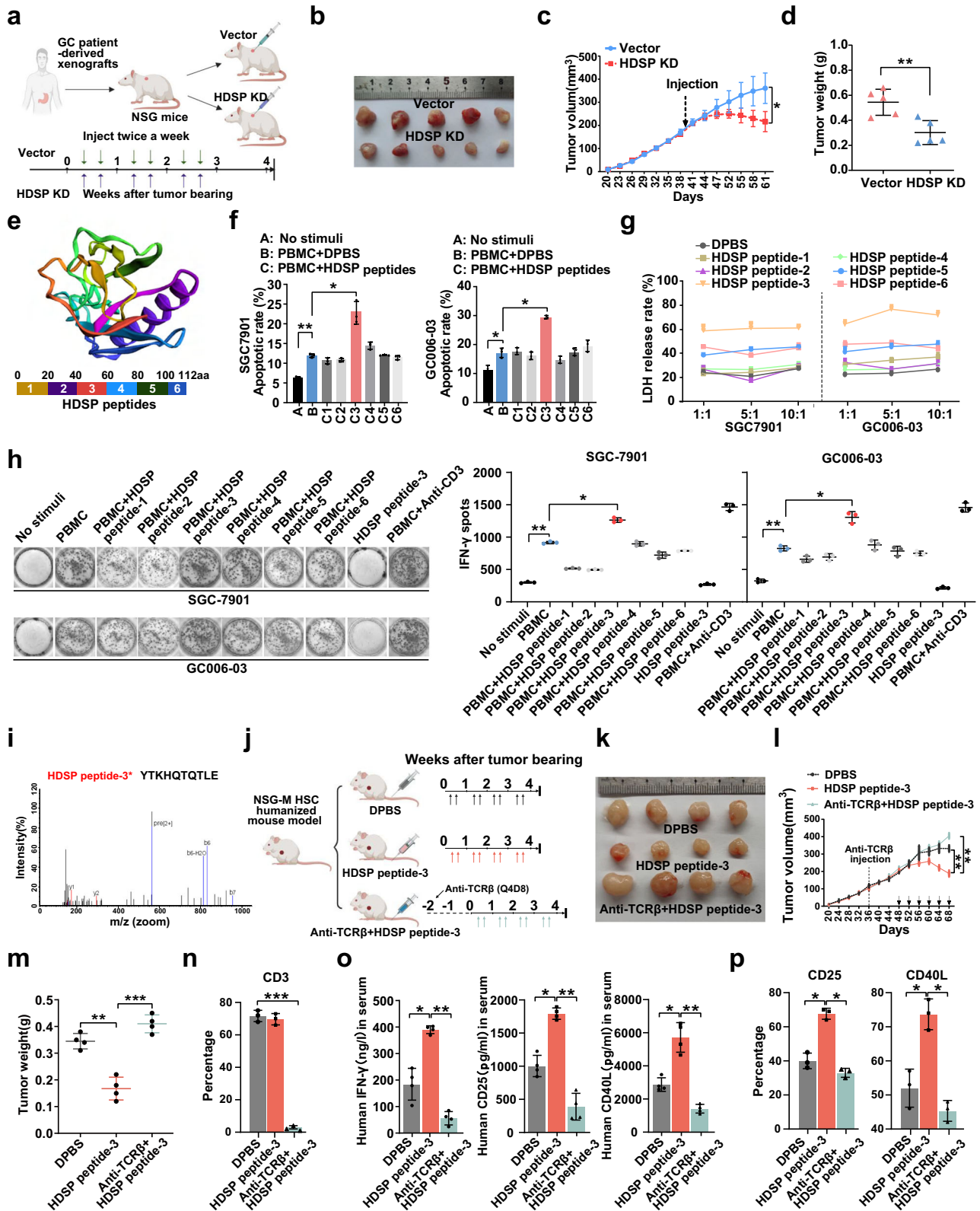


Fig. 7 | MECOM/HOXA10-HOXA9-HDSP/MECOM form a positive feedback regulatory circuit. **a** Diagram illustrating the binding of MECOM to the HOXA10-HOXA9 promoter. **b, c** qRT-PCR and western blot analyses were conducted to detect the effect of MECOM or AP-1 complex interference on HOXA10-HOXA9 in GC cells. The western blot assay was repeated 3 times independently with similar results. Mean \pm SD, one-way ANOVA analysis. $n = 3$ biologically independent experiments. ** $P < 0.01$, *** $P < 0.001$. **d-f** ChIP-PCR assay detected the enrichment of MECOM or AP-1 complex in the promoter region of HOXA10-HOXA9. Mean \pm SD, one-way ANOVA analysis. $n = 3$ biologically independent experiments. * $P < 0.05$, ** $P < 0.01$. **g, h** Transcriptional activity of MECOM or AP-1 complex on the HOXA10-

HOXA9 promoter was assessed via dual luciferase reporter gene assay. Mean \pm SD, one-way ANOVA analysis. $n = 3$ biologically independent experiments. ns, not significant; * $P < 0.05$, ** $P < 0.01$. **i** The mechanism diagram illustrates a positive feedback regulatory loop involving MECOM/HOXA10-HOXA9-HDSP, which continuously promotes the proliferation and metastasis of GC cells. This mechanism diagram was created using BioRender.com and is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>). Source data are provided as a Source data file.



In conclusion, this study unveils a translatable lncRNA, HOXA10-HOXA9, encoding the small protein HDSP, which drives GC progression by protecting MECOM from TRIM25-mediated ubiquitination and degradation. The study also reveals that the HOXA10-HOXA9/HDSP/MECOM module forms a positive feedback regulatory circuit that activates downstream SPINK1-EGFR signaling in GC cells. Targeting HDSP or using HDSP as a neoantigen for vaccine

development may have potential clinical applications in GC therapy. As HDSP is an identified lncRNA-derived small protein in GC, further exploration of lncRNAs with translational capacity and their peptides/protein products may shed light on the mechanisms underlying GC tumorigenesis and progression. These findings offer opportunities for developing effective therapeutic strategies to improve patient prognosis.

Fig. 8 | HDSP is a valuable therapeutic target and neoantigen for gastric cancer vaccine development. **a** A schematic of the PDX GC model construction is depicted. Mice were intratumorally injected with either shRNA-HDSP vector (purple arrows) or empty pLV6 vector (green arrows) with delivery buffer for 3 weeks. This schematic was created using BioRender.com and is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. **b–d** Representative tumor tissue image post-HDSP knockdown in the PDX model (**b**). Following intratumoral sh-HDSP plasmid injection, tumor volume (**c**) and weight (**d**) were assessed. Two-way/One-way ANOVA analysis. $n = 5$ biologically independent animals. $*P < 0.05$, $**P < 0.01$. **e** The conformation of HDSP, as predicted by trRosetta, is shown along with the bisection of HDSP into 6 peptides. **f, g** GC cells were treated with PBMC and 6 peptides, respectively. The apoptosis rate of GC cells was assessed using flow cytometry (**f**), and the LDH release rate was measured with an LDH release assay (**g**). One-way ANOVA analysis. $n = 3$ biologically independent experiments. $*P < 0.05$, $**P < 0.01$. **h** After co-incubating GC cells with PBMCs and different HDSP peptides, the ELISpot assay quantified IFN- γ secretion by PBMCs stimulated with these peptides. One-way ANOVA analysis. $n = 3$ biologically independent experiments. **i** HLA immunopeptidomic analysis revealed the

presentation of HDSP peptide-3 on cell membranes by HLA-1. **j** Schematic of HSC-humanized model construction: PDX model established, HSC-humanized mice divided into three groups. Group 1: DPBS injections; Group 2: HDSP peptide-3 injections for 4 weeks; Group 3: T cell depletion with TCR β monoclonal antibody before HDSP peptide-3 administration. This schematic was created using BioRender.com and is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. **k** Representative tumor image from euthanized HSC-humanized mice. **l, m** The tumor volume and weight of the above HSC-humanized model mice were measured. Two-way ANOVA analysis. $n = 4$ biologically independent animals. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. **n** Flow cytometry evaluated CD3-positive T cell percentage in HSC-humanized mice plasma. One-way ANOVA analysis. $n = 3$ biologically independent animals. $***P < 0.001$. **o** The serum levels of IFN- γ , CD25, and CD40L in the aforementioned mice were measured using ELISA. One-way ANOVA analysis. $n = 3$ biologically independent animals. $*P < 0.05$, $**P < 0.01$. **p** The levels of CD25 and CD69 expression in the serum of the mentioned mice were analyzed using flow cytometry. One-way ANOVA analysis. $n = 3$ biologically independent animals. $*P < 0.05$. Data are means \pm SD. Source data are provided as a Source data file.

Methods

GEO and TCGA database processing

The lncRNA expression profiles in GC were sourced from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>) and normalized and processed using TCGA biolinks package. The independent GC gene expression data of GSE51308⁴⁵/GSE53137⁴⁶ (lncRNA/mRNA expression profiling) and GSE128320⁴⁷/GSE111866⁴⁸/GSE79539⁴⁹ (ribosome profiling) were retrieved from the Gene Expression Omnibus (GEO) database. The lncRNAs were annotated using the human genome (GRCh38) GENCODE version 33. To find lncRNAs with coding potential, the trimmed Ribo-seq data were aligned with the UCSC HG38 reference genome for STAR alignment. Differentially expressed lncRNAs were identified by the DESeq2 package with the following threshold values: $|\log_2(\text{fold change})| \geq 1$ and adjusted P -value < 0.05 . The intersection of the results from the above databases was obtained.

Cell culture

The human gastric mucosal cell line GES-1, the GC lines NCI-N87, HGC-27, and AGS, and the human embryonic kidney epithelial cell line HEK-293T were purchased from the National Collection of Authenticated Cell Cultures (Shanghai, China). GC lines MKN-45, SGC-7901, and BGC-823 were obtained from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China). The signet ring cell GC cell line GC006-03 was provided by Dr. Jinfei's research group (Department of Oncology, Nanjing First Hospital)⁵⁰. Peripheral blood mononuclear cells (PBMCs) were acquired from OriBiotech (Shanghai, China). SGC-7901, BGC-823, and HEK-293T cells were cultured in DMEM (Thermo Fisher Scientific, MA, USA), whereas GES-1, MKN-45, NCI-N87, HGC-27, AGS, GC006-03, and PBMCs were cultured in RPMI 1640 medium (Thermo Fisher Scientific, MA, USA). The media were supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin. All cell lines were maintained at 37 °C in a humidified incubator with 5% CO₂ and regularly tested negative for mycoplasma contamination.

Tissue samples and microarray chips

Matched primary GC tissues and adjacent non-tumor tissues were collected from 70 GC patients at the First Affiliated Hospital of Nanjing Medical University. The cases were selected based on a clear pathological diagnosis, and the patients were not preoperatively treated with anticancer agents. Tissue array chips comprising 90 pairs of GC tissues and adjacent normal tissues were purchased from SuperChips (FM-S4006-1). Patient demographic information, including gender, age, tumor size, TNM stage, survival, and cause of death, with the corresponding tissues and microarrays is presented in Supplementary Table 2. This study was approved by the Ethics Committee on Human

Research of Nanjing Medical University, and informed written consent was obtained from each patient.

Polysome-bound RNA analysis

SGC-7901 cells were treated with cycloheximide (CHX; Sigma-Aldrich), followed by lysis with MCB polysome lysis buffer on ice. After centrifugation at 16,200 $\times g$ for 10 min, the supernatant was loaded onto a 15–50% sucrose gradient and centrifuged at 180,000 $\times g$ for 240 min using an SW 41Ti rotor (Beckman). The resulting gradient was separated into 30 samples, and the RNA in each sample were analyzed.

Flag knock-in experiment

The 3 \times Flag sequence was inserted following the HDSP start codon ATG using the CRISPR/Cas9 genome editing system. Guide RNAs (gRNAs) were initially designed using the online CRISPR design tool CRISPOR to target the genomic locus of the HDSP start codon ATG. These gRNA sequences were then cloned into CRISPR vectors. Gastric cancer cells were cultured in 6-well plates until reaching ~70% confluence on the day of transfection. Lipofectamine 3000 (Thermo Fisher Scientific, MA, USA) was used according to the manufacturer's instructions to co-transfect the CRISPR/Cas9 plasmid into the cells along with a homologous directed repair (HDR) template containing a 3 \times Flag sequence and flanking homologous arms. After a 48 h incubation period, the cells were subjected to monoclonal culture. Multiple single-cell clones were chosen for Western blot analysis to detect Flag expression. The genome of the positively selected cloned cells was then isolated, and the region encompassing the HDSP initiation codon ATG was amplified via PCR to confirm successful insertion of the 3 \times Flag sequence.

Construction of cell lines with stable expression

Lentiviral vectors expressing HDSP-shRNA, HDSP-Flag, and MECOM-Flag were co-transfected into HEK-293T cells with lentiviral packaging constructs psPAX2 and pMD2.G (VSV-G) at a 4:3:2 ratio to produce lentiviruses, which were subsequently collected and concentrated. Lentiviruses expressing the indicated gene constructs were used to infect SGC-7901 and BGC-823 cells, which were then selected using puromycin. Stable cell lines expressing HDSP-shRNA, HDSP-Flag, MECOM-Flag, and others were established for subsequent studies.

RNA-seq experiment

Total RNA was extracted from SGC-7901 cells with HDSP or MECOM knockdown, and its purity and concentration were assessed. RNA samples were denatured to disrupt secondary structures, and mRNA was enriched using oligo (dT) magnetic beads. The mRNA was then fragmented by adding fragmentation buffer and incubating at an appropriate temperature. cDNA was synthesized using a two-step

method. The ends of the double-stranded cDNA were repaired, and an A base was added to the 3' end for adapter ligation. The resulting product was PCR-amplified to construct the library, which was quality-controlled. PCR products were denatured into single strands and circularized. These single-stranded circular DNA molecules underwent rolling circle replication to form DNA nanoballs (DNBs) containing multiple DNA copies. The DNBs were loaded onto a high-density DNA nanochip and sequenced using combinatorial probe-anchor synthesis (cPAS) technology. Raw sequencing data were filtered using SOAPnuke (v1.5.6) to remove adapter-contaminated reads, reads with more than 5% unknown bases (N), and low-quality reads (with more than 20% of bases having a quality score below 15), resulting in clean data. Clean data were aligned to the reference genome using HISAT2 (v2.1.0) and to the reference gene set using Bowtie2 (v2.3.4.3). Gene expression levels were quantified using RSEM (v1.3.1), and pheatmap (v1.0.8) was used to generate clustering heatmaps of gene expression across samples. Differentially expressed genes were identified using DESeq2 (v1.4.5), DEGseq, or PoissonDis, with thresholds set at Q value ≤ 0.05 or FDR ≤ 0.001 . Enrichment analysis of differentially expressed genes was performed using the Phyper function for GO (<http://www.geneontology.org/>) and KEGG, with a significance threshold of Q value ≤ 0.05 . Genes meeting this criterion were considered significantly enriched in the candidate gene set.

Co-IP and mass spectrometry analysis

Co-immunoprecipitation (Co-IP) experiments were performed using the Dynabeads Protein A/G Immunoprecipitation Kit (Thermo Fisher Scientific, MA, USA) as per the manufacturer's instructions. SGC-7901 cells expressing EGFP-Flag, HDSP-Flag, or MECOM-Flag were collected and lysed in IP lysis buffer. The lysates were incubated with anti-Flag or anti-MECOM antibodies, followed by the addition of Protein A/G magnetic beads and a further 2-h incubation. After three washes, proteins were eluted with elution buffer. The immunoprecipitated proteins were separated via SDS-PAGE and visualized using silver staining. Differential protein bands were analyzed by mass spectrometry (MS) (Biotech-pack, Beijing, China) and western blot. Antibody details are provided in Supplementary Table 5.

To identify HDSP- or MECOM-interacting proteins, SGC-7901 cells were stably transfected with HDSP-Flag or MECOM-Flag plasmids, using the EGFP-Flag plasmid as a negative control. After a 48-h incubation, total proteins were extracted and immunoprecipitated using an anti-Flag antibody. The enriched samples were separated by SDS-PAGE and silver-stained to visualize differential bands and assess enrichment efficiency. Peptides were eluted using a 60-min linear gradient of 3% to 25% acetonitrile in 5% DMSO, 0.1% formic acid at a flow rate of 250 nl/min and 40 °C, and introduced into a mass spectrometer using a nano EASY-Spray source at 2000 V (Thermo Scientific). These proteins were analyzed using a Q-Exactive mass spectrometer with a nano-electrospray ion source. Online peptide analysis against the UniProt human protein database was performed using Mascot (v2.3.02) with default parameters (false discovery rate $\leq 1\%$ and score $R \geq 40$), no enzyme specificity, no peptide modifications, peptide tolerance: ± 5 ppm, and fragment tolerance: ± 0.03 Da. Differentially expressed proteins were subjected to Gene Ontology (GO) functional enrichment analysis.

For characterizing endogenous HDSP properties, total proteins from SGC-7901 cells or GC specimens were separated and stained with Coomassie brilliant blue. The HDSP molecular weight band was excised and subjected to in-gel trypsin digestion for MS analysis. To identify HLA-bound peptides, lysates from HDSP-expressing HEK-293T cells were loaded onto a cross-linked mAb column after generating an MHC immunoaffinity column. The MHC eluate was separated by RP-HPLC, and LC-MS/MS analyses of HLA-bound peptides were conducted using a Q-Exactive-Plus mass spectrometer equipped with an Easy nLC 1000 nano-capillary UHPLC system (Thermo Fisher Scientific).

Proximity ligation assay (PLA)

The PLA was conducted using a Duolink® In situ Detection Reagents Red kit (Merck Millipore, MA, USA, C6628, DUO92008) as per the vendor's protocol. GC cells were cultured in chambered slides, fixed, and permeabilized. Non-specific binding was blocked using Duolink® Blocking Solution, followed by incubation with primary antibodies (rabbit anti-TRIM25/Flag and mouse anti-MECOM) overnight. After washing, PLA probes were applied and incubated, followed by ligation and amplification steps. Finally, slides were washed, mounted with coverslips using Duolink® In situ mounting medium with DAPI, and analyzed using a confocal microscope (Olympus, FV1200).

In vitro translation

The HOXA10-HOXA9 (HOXA10-HOXA9 mut) and HOXA10-HOXA9 ORF2-Flag (HOXA10-HOXA9 ORF4-Flag) sequence was subcloned into a pcDNA-3.1 vector downstream of the T7 promoter respectively. The resulting construct was then subjected to in vitro transcription/translation using a rabbit reticulocyte-coupled system (Promega, WI, USA, L4540) supplemented with [³⁵S] methionine for a duration of 1 h. Following the incubation period, the translated product was separated by 15% SDS-PAGE and visualized by western blot.

E2Select ubiquitin conjugation assay

The E2Select ubiquitin conjugation assay was conducted using the E2Select Ubiquitin Conjugation Assay Kit in accordance with the manufacturer's instructions (Ubibotech, Changchun, China). Recombinant proteins MECOM-His and TRIM25-GST at a concentration of 2 μ M were added to the reaction buffer and incubated at 37 °C for 2 h. Subsequently, 5 μ l of 5 \times SDS-PAGE buffer containing 1 μ l of freshly prepared 1 M DTT was added to 20 μ l of the sample, and the mixture was heated at 95 °C for 5 min. Finally, western blot analysis was performed using the anti-MECOM and anti-TRIM25 antibodies.

ChIP assay

Analysis of the potential MECOM binding site in the *SPINK1* and *HOXA10-HOXA9* promoter regions was performed using JASPAR software and MECOM chromatin immunoprecipitation-sequencing (ChIP-seq) data. Specific primers for the indicated promoter regions were designed. ChIP assays were conducted using the EZ-Magna ChIP™ G-Chromatin Immunoprecipitation Kit (Millipore, MA, USA, 17-10085) following the manufacturer's protocol. Briefly, SGC-7901 and BGC-823 cells were cross-linked with 1% formaldehyde and quenched by glycine. The cross-linked lysates were sonicated to obtain DNA fragments with an average size of 100–500 bp and immunoprecipitated using antibodies (anti-MECOM, anti-c-FOS, and anti-c-JUN). Finally, the abundance of enriched fragments of the *SPINK1* or *HOXA10-HOXA9* promoter was determined using qRT-PCR assay.

LDH release assay

Lactate dehydrogenase (LDH) release in the culture medium was measured using the CytoTox 96 Non-Radioactive Cytotoxicity Assay kit (Promega, WI, USA, G1780), following the manufacturer's protocol. PBMCs from healthy donors (effector cells) were pretreated with six HDSP peptides (5 μ g/well) for 12 h, and then, 1×10^4 SGC-7901 and GC006-03 target cells were added to the 96-well U-shaped plate. PBMCs were added to each well at an effector:target ratio of 1:1–15:1 and co-cultured with GC cells (target cells) for an additional 24 h. The target cells were treated with isopyknic culture medium in the natural release group and with 2% Triton X-100 in the maximum release group. The OD value of each well was measured at 490 nm using a microplate reader (Bio-Rad). The LDH release rate (%) was calculated as [(OD value of experimental group-OD value of target cell natural release group)/(OD value of target cell maximum release group-OD value of target cell natural release group)] $\times 100\%$.

ELISpot assay

Human PBMCs were incubated overnight in complete medium before being seeded at a density of 2×10^5 cells per well in MultiScreen HTS Filter Plates (Merck Millipore) precoated with anti-IFN- γ (15 mg/ml; clone 1-D1K; Mabtech, Stockholm, Sweden). Test wells were supplemented with HDSP peptides 1–6 (each at 2 mg/ml). Negative control wells lacked peptides, whereas positive control wells included the mAb CD3 (dilution 1:1000; Mabtech). The plates were then incubated for 24 h at 37 °C. The plates were washed six times with PBS and then incubated for 2 h at room temperature with biotinylated anti-IFN- γ (1 mg/ml; clone mAb-7B6-1; Mabtech). After six further washes, alkaline phosphatase-conjugated streptavidin (dilution 1:1000, Mabtech) was added for 1 h at room temperature. The plates were then washed a further six times and developed for 20 min with BCIP/NBT Substrate (Mabtech). Finally, the spots were counted using an automated enzyme-linked immunospot (ELISpot) Reader System (Mabtech IRIS-FluoroSpot).

Enzyme-linked immunosorbent assay (ELISA)

ELISA assay was used to measure the cytokine expression in plasma from HSC-humanized mice. Initially, 96-well plates were coated overnight with specific antibodies at 4 °C. The next day, the plates were blocked with a blocking buffer for 1 h. Following this, samples or standards were added to each well and incubated for 2 h at room temperature. After incubation, the plates were washed three times with wash buffer. A biotinylated detection antibody specific to the target protein was then added and incubated for 1 h. After another washing step, streptavidin conjugated with HRP was added and incubated for 30 min. The final step involved adding a substrate solution and incubating for 20 min in the dark. The reaction was stopped by adding a sulfuric acid stop solution, and the absorbance was measured at 450 nm.

Flow cytometry

Flow cytometry was used to measure cytokine levels in the plasma of HSC-humanized mice. Initially, 100 μ l of whole blood was mixed with 30 μ l of heparin sodium solution for anticoagulation. The anticoagulated blood was then incubated with anti-CD3, anti-CD25, and anti-CD69 antibodies in darkness for 30 min. Afterward, 2 ml of red blood cell lysate was gently mixed with the sample and allowed to react for 10 min at room temperature, shielded from light. Following centrifugation at 1500 rpm for 5 min and discarding of the supernatant, 2 ml of PBS was added to the pellet, mixed, and centrifuged again at 1500 rpm for 5 min, with the supernatant being discarded. Finally, the cellular precipitate was resuspended in 500 μ l of PBS and analyzed using a flow cytometer (BD, FACSymphony A5 SORP).

Animal experiments

All animal procedures were approved by the Institutional Animal Care and Use Committee of Nanjing Medical University (IACUC) (permit number IACUC-2112030-1). The maximal tumor size permitted by IACUC of Nanjing Medical University is 2000 mm³. In the experiments, the maximum permitted tumor burden (2000 mm³) was not exceeded in the mice at any time point.

Xenograft model

Subcutaneous xenografts were established by injecting SGC-7901 cells suspended in 100 μ l PBS subcutaneously into 4-week-old male BALB/c-nude mice (Gempharmatech, Nanjing, China). Tumor volumes were measured twice a week using the following formula: volume (mm³) = length \times width²/2. Mice were sacrificed 5 weeks after inoculation, and tumors were resected, weighed, and further analyzed by H&E staining and IHC. For the metastasis model, SGC-7901-luci cells were injected into the lateral tail veins of 4-week-old male BALB/c-nude mice. The development of pulmonary metastases was monitored using a bioluminescence imaging system (Xenogen IVIS 200). Mice were sacrificed 9 weeks after implantation, and their lungs were dissected for H&E

staining. The luciferase signal obtained from the pulmonary metastatic nodules was detected using a small-animal in vivo imaging system (Nikon, Japan).

PDX model

PDX models were established by mincing GC surgical specimens obtained from the First Affiliated Hospital of Nanjing Medical University into small pieces (\sim 8 mm³) and subcutaneously transplanting them into 3-week-old male NOD/ShiLtJGpt-Prkdc^{em26Cd52}Il2rg^{em26Cd22}/Gpt (NSG) mice (Gempharmatech, Nanjing, China). When the tumor burden reached 1 cm in diameter, sections of these tumors were collected and transplanted into new nude mice to establish the PDX model. To knock down HDSP, the mice were randomly divided into two groups (Vector and HDSP KD), which were subsequently treated with HDSP shRNAs or empty pLV6 vector with PNT200 in vivo transfection reagent (ProteinBio, Nanjing, China) via intratumoral injection twice per week. Tumor volume was monitored every 3 days, and the mice were sacrificed when they met institutional euthanasia criteria based on tumor size and their overall health condition. The tumors were weighed and collected for histological and protein analyses.

Hematopoietic stem cell humanized model

To test the potential antigenicity, HDSP was divided into six peptides of 20 amino acids in length, which were chemically synthesized by Genscript (Nanjing, China). For the in vitro cell assay, PBMCs from healthy donors were pretreated with corresponding HDSP peptide (100 μ g) for 12 h and then co-cultured with GC cells (SGC-7901 and GC006-03) for an additional 24 h. The apoptosis rate of GC cells was determined by flow cytometry adopting FITC-Annexin V double-staining on a BD FACSCalibur Flow Cytometer (BD Biosciences, NJ, USA). For in vivo experiments, 3-week-old male NOD/ShiLtJGpt-Prkdc^{em26Cd52}Il2rg^{em26Cd22}Rosa26^{em1Cin(hCSF2&IL3&KITLG)}/Gpt (NSG-M) mice obtained from Gempharmatech (Nanjing, China) were randomly allocated into 3 groups. The first group received isopycnic DPBS (control), the second group received HDSP peptide 3 (400 μ g, twice per week), while the third group of mice were given HDSP peptide-3 before being cleared of T cells by TCR β monoclonal antibodies. Tumor volume was measured bi-weekly, and after 4 weeks, blood samples were collected from the mice via eyeball extraction to assess the expression of T-cell-activating molecules using ELISA and flow cytometry. Concurrently, tumors were weighed, and their morphology was examined through H&E staining. In addition, CD3 expression in the tumors was assessed via immunohistochemistry.

Bioinformatics analysis

The coding potential of HOXA10-HOXA9 was evaluated using CPC2 and PhyloCSF (<http://genome.ucsc.edu/>), while the candidate ORF was identified using ORF Finder (<https://www.ncbi.nlm.nih.gov/orffinder/>). The spatial structure model of HDSP was analyzed using SWISS-MODEL (<https://swissmodel.expasy.org/>) and trRosetta (<https://yanglab.nankai.edu.cn/trRosetta/>), and the protein–protein interaction network between HDSP and MECOM was assessed using swMATH (<https://www.swmath.org/software/17207>). The correlation between MECOM RNA expression and overall survival of GC patients was determined using Kaplan–Meier Plotter (<http://kmplot.com/analysis/>), while the correlation between SPINK1 RNA expression and disease-free survival of GC patients was analyzed using GEPIA. The RNA expression correlation between MECOM and SPINK1 was investigated using ChipBase (<https://rna.sysu.edu.cn/chipbase/>), and the ubiquitination modification site was analyzed using PhosphoSitePlus (<https://www.phosphosite.org/homeAction.action/>).

Statistics and reproducibility

The experimental data were analyzed using SPSS 20.0, GraphPad Prism 7.0, and ImageJ software. Data are expressed as mean \pm standard

deviation from three independent experiments, unless otherwise specified. Parametric data were analyzed using unpaired, two-tailed Student's *t*-tests, and nonparametric data were analyzed using a two-sided Mann–Whitney U test. Corresponding figure legends indicate sample sizes and significance levels. A value of $P < 0.05$ was considered statistically significant.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The RNA-seq data have been archived in the Genome Sequence Archive (GSA) database under the accession codes [HRA005065](#) and [HRA005068](#). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier [PXD043737](#) and [PXD053200](#). All lncRNA-derived peptide sequences were cross-referenced with the human SwissProt protein database. MECOM ChIP-seq data is downloaded and analyzed from the Cistrome Data Browser database (<http://cistrome.org/db/#/>). The remaining data are available within the article, Supplementary Information or Source data file. Source data are provided with this paper.

References

- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **70**, 7–30 (2020).
- Xia, C. et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chin. Med. J.* **135**, 584–590 (2022).
- Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin.* **72**, 7–33 (2022).
- Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
- Mudge, J. M. et al. Standardized annotation of translated open reading frames. *Nat. Biotechnol.* **40**, 994–999 (2022).
- Martinez, T. F. et al. Accurate annotation of human protein-coding small open reading frames. *Nat. Chem. Biol.* **16**, 458–468 (2020).
- Couso, J. P. & Patraquim, P. Classification and function of small open reading frames. *Nat. Rev. Mol. Cell Biol.* **18**, 575–589 (2017).
- Wright, B. W., Yi, Z., Weissman, J. S. & Chen, J. The dark proteome: translation from noncanonical open reading frames. *Trends Cell Biol.* **32**, 243–258 (2022).
- Ruiz-Orera, J. & Alba, M. M. Translation of small open reading frames: roles in regulation and evolutionary innovation. *Trends Genet.* **35**, 186–198 (2019).
- Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *Elife* **4**, e08890 (2015).
- Zheng, C. et al. CRISPR/Cas9 screen uncovers functional translation of cryptic lncRNA-encoded open reading frames in human cancer. *J. Clin. Invest.* **133**, e159940 (2023).
- Prensner, J. R. et al. Noncanonical open reading frames encode functional proteins essential for cancer cell survival. *Nat. Biotechnol.* **39**, 697–704 (2021).
- Liu, Y. et al. Circular RNA circPGD contributes to gastric cancer progression via the sponging miR-16-5p/ABL2 axis and encodes a novel PGD-219aa protein. *Cell Death Discov.* **8**, 384 (2022).
- Peng, Y. et al. A novel protein AXIN1-295aa encoded by circAXIN1 activates the Wnt/beta-catenin signaling pathway to promote gastric cancer progression. *Mol. Cancer* **20**, 158 (2021).
- Hu, F. et al. Vimentin binds to a novel tumor suppressor protein, GSPT1-238aa, encoded by circGSPT1 with a selective encoding priority to halt autophagy in gastric carcinoma. *Cancer Lett.* **545**, 215826 (2022).
- Zhang, Yu. et al. CircDIDO1 inhibits gastric cancer progression by encoding a novel DIDO1-529aa protein and regulating PRDX2 protein stability. *Mol. Cancer* **20**, 101 (2021).
- Sun, L. et al. The oncomicropeptide APPLE promotes hematopoietic malignancy by enhancing translation initiation. *Mol. Cell* **81**, 4493–4508 e4499 (2021).
- Backx, E. et al. MECOM permits pancreatic acinar cell dedifferentiation avoiding cell death under stress conditions. *Cell Death Differ.* **28**, 2601–2615 (2021).
- Bleu, M. et al. PAX8 and MECOM are interaction partners driving ovarian cancer. *Nat. Commun.* **12**, 2442 (2021).
- Birdwell, C. et al. EVI1 dysregulation: impact on biology and therapy of myeloid malignancies. *Blood Cancer J.* **11**, 64 (2021).
- Dikic, I. Proteasomal and autophagic degradation systems. *Annu. Rev. Biochem.* **86**, 193–224 (2017).
- Gallo, L. H., Ko, J. & Donoghue, D. J. The importance of regulatory ubiquitination in cancer and metastasis. *Cell Cycle* **16**, 634–648 (2017).
- Ma, Y. et al. CRISPR-mediated MECOM depletion retards tumor growth by reducing cancer stem cell properties in lung squamous cell carcinoma. *Mol. Ther.* **30**, 3341–3357 (2022).
- Liang, B. & Wang, J. EVI1 in leukemia and solid tumors. *Cancers* **12**, 2667 (2020).
- Räsänen, K., Itkonen, O., Koistinen, H. & Stenman, U. H. Emerging roles of SPINK1 in cancer. *Clin. Chem.* **62**, 449–457 (2016).
- Stenman, U. H. Role of the tumor-associated trypsin inhibitor SPINK1 in cancer development. *Asian J. Androl.* **13**, 628–629 (2011).
- Chen, F. et al. Targeting SPINK1 in the damaged tumour microenvironment alleviates therapeutic resistance. *Nat. Commun.* **9**, 4315 (2018).
- Bard-Chapeau, E. A. et al. EVI1 oncoprotein interacts with a large and complex network of proteins and integrates signals through protein phosphorylation. *Proc. Natl Acad. Sci. USA* **110**, E2885–E2894 (2013).
- Sayadi, A. et al. Functional features of EVI1 and EVI1Delta324 isoforms of MECOM gene in genome-wide transcription regulation and oncogenicity. *Oncogene* **35**, 2311–2321 (2016).
- Lin, T. C. Functional roles of SPINK1 in cancers. *Int. J. Mol. Sci.* **22**, 3814 (2021).
- Ge, Q. et al. Micropeptide ASAP encoded by LINC00467 promotes colorectal cancer progression by directly modulating ATP synthase activity. *J. Clin. Invest.* **131**, e152911 (2021).
- Liang, Z. X. et al. A novel NF-kappaB regulator encoded by circPLCE1 inhibits colorectal carcinoma progression by promoting RPS3 ubiquitin-dependent degradation. *Mol. Cancer* **20**, 103 (2021).
- Barczak, W. et al. Long non-coding RNA-derived peptides are immunogenic and drive a potent anti-tumour response. *Nat. Commun.* **14**, 1078 (2023).
- Chen, J. et al. Pervasive functional translation of noncanonical human open reading frames. *Science* **367**, 1140–1146 (2020).
- Laumont, C. M. et al. Noncoding regions are the main source of targetable tumor-specific antigens. *Sci. Transl. Med.* **10**, eaau5516 (2018).
- Zhang, C. et al. Micropeptide PACMP inhibition elicits synthetic lethal effects by decreasing CtIP and poly(ADP-ribosylation). *Mol. Cell* **82**, 1297–1312 e1298 (2022).
- Queisser, A. et al. Ecotropic viral integration site 1, a novel oncogene in prostate cancer. *Oncogene* **36**, 1573–1584 (2017).
- Pradhan, A. K., Kuila, N., Singh, S. & Chakraborty, S. EVI1 up-regulates the stress responsive gene SIRT1 which triggers deacetylation and degradation of EVI1. *Biochim. Biophys. Acta* **1809**, 269–275 (2011).

39. Yoshimi, A. et al. Evi1 represses PTEN expression and activates PI3K/AKT/mTOR via interactions with polycomb proteins. *Blood* **117**, 3617–3628 (2011).
40. Cattaneo, F. & Nucifora, G. EVI1 recruits the histone methyltransferase SUV39H1 for transcription repression. *J. Cell Biochem.* **105**, 344–352 (2008).
41. Senyuk, V., Premanand, K., Xu, P., Qian, Z. & Nucifora, G. The oncoprotein EVI1 and the DNA methyltransferase Dnmt3 co-operate in binding and de novo methylation of target DNA. *PLoS ONE* **6**, e20793 (2011).
42. Chakraborty, S., Senyuk, V., Sitailo, S., Chi, Y. & Nucifora, G. Interaction of EVI1 with cAMP-responsive element-binding protein-binding protein (CBP) and p300/CBP-associated factor (P/CAF) results in reversible acetylation of EVI1 and in co-localization in nuclear speckles. *J. Biol. Chem.* **276**, 44936–44943 (2001).
43. Liao, C. et al. SPINKs in tumors: potential therapeutic targets. *Front Oncol.* **12**, 833741 (2022).
44. Charpentier, M. et al. IRES-dependent translation of the long non coding RNA meloe in melanoma cells produces the most immunogenic MELOE antigens. *Oncotarget* **7**, 59704–59713 (2016).
45. Zhang, Z. Z. et al. TEM7 (PLXDC1), a key prognostic predictor for resectable gastric cancer, promotes cancer cell migration and invasion. *Am. J. Cancer Res.* **5**, 772–781 (2015).
46. Gu, W., Ren, J. H., Zheng, X., Hu, X. Y. & Hu, M. J. Comprehensive analysis of expression profiles of long non-coding RNAs with associated ceRNA network involved in gastric cancer progression. *Mol. Med. Rep.* **20**, 2209–2218 (2019).
47. Xu, W. et al. Ribosome profiling analysis identified a KRAS-interacting microprotein that represses oncogenic signaling in hepatocellular carcinoma cells. *Sci. China Life Sci.* **63**, 529–542 (2020).
48. Lauria, F. et al. riboWaltz: optimization of ribosome P-site positioning in ribosome profiling data. *PLoS Comput. Biol.* **14**, e1006169 (2018).
49. Lu, S. et al. A hidden human proteome encoded by ‘non-coding’ genes. *Nucleic Acids Res.* **47**, 8111–8125 (2019).
50. Su, X. et al. Establishment and characterization of gc-006-03, a novel human signet ring cell gastric cancer cell line derived from metastatic ascites. *J. Cancer* **9**, 3236–3246 (2018).

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (82272972 and 81872269 to L.X.H., 82372625 to S.M., 81773284 to L.Y.W.); Natural Science Foundation of Jiangsu Province (BK20220259 to S.M.); The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital, Gusu School (GSRCKY20210102 to S.M.); The College Students’ Innovative Entrepreneurial Training Plan Program of China (202310312089Y to L.Q.N.). The authors thank Prof. Jinfei Chen from Nanjing First Hospital for presenting Signet ring cell gastric cancer (SRCGC) cell line GC006-03, and Prof. Zejian Liu from Nanjing Medical University providing helpful advice on the modification of Methods.

Author contributions

Y.C., Qiuhui L., and X.Y. performed cell and mouse experiment, analyzed the data, and prepared the manuscript. Y.C., Qiuhui L., X.G., and Y.H. constructed tumor model. L.L., Z.Z., and H.L. performed PCR validation and western blot. G.G. and J.G. performed western blot and polysome profiling analysis. X.Y. and Y.L. provided the source of GC tissue and study supervision. Y.C., Qiuhui L., and M.L. prepared the figures. R.X. and Qiu Luo L. analyzed the TCGA and GEO data. M.S. and X.L. supervised the project, participated in experimental design and provided laboratory resources. Y.C., Qiuhui L., Z.Z., M.S., and X.L. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-50986-7>.

Correspondence and requests for materials should be addressed to Xianghua Liu or Ming Sun.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions 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-NoDerivatives 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) 2024