BMC Cancer

RESEARCH

Open Access

Identification of endometrial cancer biomarkers using weighted gene coexpression network analysis and machine learning

Hui Wang[1†], Yanwen Gong[2†], Xiu Han[3*†] and Jiawei Zeng[2,4*]

Abstract

Background Endometrial carcinoma (UCEC) exhibits a rising incidence in China, imposing a substantial burden on both women and society. Identifying biomarkers for UCEC is critical for precise diagnosis and treatment.

Methods RNA-seq data for UCEC and normal endometrial tissues were sourced from the GEO and TCGA databases. The Limma package was used to analyze differentially expressed genes (DEGs) from GEO datasets, while weighted gene co-expression network analysis (WGCNA) was employed to identify gene modules associated with UCEC. Machine learning algorithms (GBM, KNN, LASSO, SVM, NNET, RF, DT, GLM) were subsequently applied for further gene screening. Key biomarkers were identified by analyzing overall survival (OS) and progression-free survival (PFS) using TCGA-UCEC data on the Xiantao Academic online analysis platform.Functional enrichment analysis, clinical significance assessment, COX regression analysis, prognostic nomogram construction, single-gene logistic regression analysis, potential mechanism exploration, and immune characterization were all conducted via the Xiantao Academic online analysis platform. Drug sensitivity profiling was performed using the CPADS database, followed by molecular docking validation. In vitro functional assays included the EdU assay for evaluating cell proliferation and the Transwell assay for assessing cell invasion.

Results Kaplan–Meier survival analysis showed that both overall survival (OS) and progression-free survival (PFS) were significantly shorter in the high-MAL-expression group than in the low-expression group, suggesting a close association between high MAL expression and poor prognosis. Further multivariate Cox regression analysis, which included clinical stage and tumor grade, revealed that the independent predictive value of MAL for OS was attenuated, indicating that its prognostic significance may be partially confounded by traditional clinicopathological factors. MAL may serve as a potential biomarker in UCEC patients, with its high expression promoting tumor progression, correlating negatively with prognosis, and modulating immune characteristics. In vitro experiments

†Hui Wang and Yanwen Gong contributed equally to this work.

*Correspondence:
Xiu Han
152884883665@163.com
Jiawei Zeng
zjweee@sina.cn

Full list of author information is available at the end of the article

demonstrated that elevated MAL expression promoted the proliferation and invasion of ECC-1 cells, while MAL knockdown suppressed these phenotypes.

Conclusion This study found that high MAL expression is significantly correlated with poor prognosis in endometrial cancer patients, which is consistent with its functional role in promoting tumor cell proliferation and invasion. However, after adjusting for clinical factors such as stage and grade, the independent prognostic effect of MAL on overall survival was not prominent, suggesting that it may more likely reflect the extent of tumor progression and adverse clinical features, rather than serving as a sole determinant of prognosis. Therefore, MAL can be regarded as a molecular marker closely associated with disease progression and survival outcomes, and it holds potential for providing auxiliary value in risk stratification and personalized management when combined with clinical parameters.

Keywords EC, WGCNA, Machine learning, Biomarkers, MAL

Introduction

Endometrial carcinoma (EC) is a common malignancy in women, with its global incidence steadily increasing [1]. Uterine corpus endometrial carcinoma (UCEC) leads to substantial annual mortality, creating major health burdens and socioeconomic strain on patients and society [2–4]. Although various treatments, such as surgical resection, radiotherapy, and chemotherapy, have improved survival rates, many patients experience poor responses or severe side effects [5–7]. Therefore, novel biomarkers and therapeutic targets are urgently needed to enhance EC prognosis.

The molecular mechanisms underlying EC are being elucidated, especially regarding gene mutations, epigenetic changes, and tumor microenvironment dynamics [8, 9]. Myelin and lymphocyte protein (MAL), a 17-kDa hydrophobic membrane protein, regulates apical trafficking of membrane and secretory proteins in polarized epithelial cells, DNA methylation in tumors, and axon–glia interactions, and shows elevated expression in multiple cell types [10–12]. Evidence indicates that MAL influences proliferation and metastasis in several cancers [13]; however, its functional role in EC progression remains unclear.

MAL has demonstrated significant biological importance in various cancers; however, its expression profile, prognostic value, and potential therapeutic implications in endometrial cancer remain systematically unelucidated. This study centers on the core research question: "Does MAL serve as a key biomarker and a potential therapeutic target in endometrial cancer?" Initially, we conducted integrated multi-cohort analysis, weighted gene co-expression network analysis (WGCNA), and machine learning screening [14] using publicly available transcriptomic cohorts to identify candidate genes closely associated with endometrial carcinogenesis and progression, with a specific focus on MAL. Subsequently, we systematically evaluated the potential role of MAL across molecular, clinical, and pharmacological dimensions by integrating analyses of its associations with clinical prognosis and pathological features, assessing its correlation with immune microenvironment characteristics, and performing drug sensitivity profiling and molecular docking predictions. Furthermore, we investigated the impact of MAL on the proliferation and invasion of endometrial cancer cells through in vitro functional experiments. This multi-faceted approach aims to provide mechanistic and translational evidence supporting the utility of MAL as a biomarker for the diagnosis and treatment of endometrial cancer.

Despite progress in EC research, MAL's mechanistic role remains underexplored. This study addresses that gap, proposing MAL as a novel biomarker and therapeutic target to improve clinical management and outcomes for patients with EC.

Materials and methods

Data acquisition and preprocessing

RNA sequencing (RNA-seq) profiles (Supplementary Tables S1 and S2) from 98 EC and 17 normal endometrial samples were obtained by batch-correcting and integrating datasets GSE17025 and GSE63678 from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) via principal component analysis in R (v4.5.0). Additionally, standardized transcripts per million–format TCGA-UCEC data (554 tumors and 35 adjacent tissues) were retrieved from the Xiantao Academic platform (v2.1; https://www.xiantaozi.com/) and the TCGA portal (https://portal.gdc.cancer.gov). The full dataset included 560 clinical records (some lacking RNA-seq matches), 577 RNA-seq profiles (some lacking clinical data), and 9 patient samples with duplicate RNA-seq entries.

Identification of DEGs

Using the Limma package in R (v4.5.0), we identified differentially expressed genes (DEGs) between UCEC and normal tissues in the integrated GSE17025/GSE63678 dataset. Genes with a false discovery rate (FDR) < 0.05 and an absolute $\log_2$ fold change ($|\log_2 FC|$) > 1 were considered significant DEGs ($\log_2 FC > 1$: upregulated; $\log_2 FC < -1$: downregulated), yielding 100 upregulated

and 67 downregulated genes. Results were visualized using a volcano plot generated with the ggplot2 package.

Functional enrichment analysis

Using the GO/KEGG Analysis module within the Clustering Analysis toolkit of the Xiantao Academic platform, we performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of DEGs. This identified significantly enriched biological processes, cellular components, and molecular functions as well as specific signaling pathways.

WGCNA

Using the WGCNA package in R (v4.5.0), we analyzed the top 25% most variably expressed genes from the GSE17025 and GSE63678 datasets. After hierarchical clustering to remove outliers and computing Pearson correlation coefficients, a soft-thresholding power of 6 was selected to meet scale-free topology criteria ($R^2 = 0.85$; slope = −1). This was used to construct an adjacency matrix, which was then transformed into a topological overlap matrix for network generation and module detection.

Hub gene identificationvia machine learning

To improve the robustness of candidate gene selection and mitigate potential biases inherent in any single algorithm, we adopted an integrated machine learning strategy. The 33 overlapping genes identified from both WGCNA and differential expression analysis were used as input features. Eight distinct supervised learning models—Gradient Boosting Machine (GBM), *k*-Nearest Neighbors (KNN), Least Absolute Shrinkage and Selection Operator (LASSO), Support Vector Machine (SVM), Neural Network (NNET), Random Forest (RF), Decision Tree (DT), and Generalized Linear Model (GLM)—were implemented using R (version 4.5.0). All input features were standardized, and each model was trained under a fivefold cross-validation framework to optimize hyperparameters and reduce overfitting risk.

To systematically integrate results across models, genes were first ranked within each model based on their feature importance (or coefficient magnitude in LASSO), and the top 12 candidate genes from each model were retained. Subsequently, the predictive performance of all models was evaluated based on cross-validation error and area under the curve (AUC). The three models exhibiting the lowest prediction error—GBM, KNN, and LASSO—were selected for further integration. Finally, the intersection of the candidate gene lists from these three models yielded six high-confidence hub genes: HP GD, MTCL1, MAL, GABRP, SCGB1D2, and MICA.

Survival analysis

We used the Kaplan–Meier survival analysis module in the Xiantao Academic platform to assess the impact of the hub genes on OS and PFS in the TCGA-UCEC cohort, with $*p* < 0.05$ considered statistically significant. MAL was associated with poor prognosis and SCGB1D2 with favorable outcomes; thus, MAL was selected for downstream investigation.

Clinical significance of MAL

Using the Xiantao Academic platform's TCGA-UCEC data module, we sequentially performed the following analyses: (1) differential expression analysis of MAL between tumor and normal tissues; (2) diagnostic accuracy assessment via ROC curves; (3) clinical correlation studies across primary therapy outcomes (PD, SD, PR, and CR), clinical grade, age, BMI, histologic type (endometrioid, mixed, or serous), tumor invasion percentage, histologic grade (G1: low; G2: intermediate; G3: high), and survival endpoints (OS, DFS, PFI); and (4) univariate logistic regression modeling for MAL. In all cases, $*p* < 0.05$ was considered statistically significant.

Clinical subgroup survival analysis of MAL

Survival analyses stratified by 12 clinical parameters [clinical grade I–IV, primary treatment outcome (PD, SD, PR, and CR), age (≤ 60 vs. > 60 years), BMI (≤ 30 vs. > 30 kg/m^2), histologic type (endometrioid, mixed, or serous), residual tumor (R0, R1, or R2), radiation therapy (yes/no), hormone therapy (yes/no), diabetes therapy (yes/no), tumor invasion (< 50% vs. ≥ 50%), menopause status (premenopausal, perimenopausal, or postmenopausal), and race (Asian, Black/African American, or White)] were conducted using the Clinical Subgroup module on the Xiantao Academic platform. Statistical significance was defined as $*p* < 0.05$.

Nomogram development

Using the Xiantao Academic platform, independent prognostic factors were identified via Cox regression and incorporated into a nomogram for predicting 1-, 3-, and 5-year OS. Nomogram predictive accuracy was assessed using calibration curves comparing predicted and observed survival probabilities.

Potential mechanism analysis of MAL

Based on the mRNA expression levels of MAL in TCGA-UCEC, samples were stratified into high-expression and low-expression groups. Differential analysis between these groups was conducted using the Limma package. Genes with an FDR < 0.05 and an absolute log2-fold change (|log2FC|) > 1 were defined as significantly differentially expressed genes (DEGs), where log2FC > 1 indicated upregulation and log2FC < −1 indicated

downregulation. This analysis identified 1,384 upregulated genes and 1,014 downregulated genes. The results were visualized using a volcano plot generated with the ggplot2 package.GO/KEGG analysis was performed on the identified DEGs to predict their cellular components, molecular functions, biological processes, and associated signaling pathways. Additionally, gene set enrichment analysis (GSEA) was conducted using all genes meeting the differential analysis threshold (FDR < 0.05) to explore potential biological pathways.We included RNA-seq data from 553 cases in TCGA-UCEC. Genes with minimal expression across most samples were initially filtered out. Outliers significantly deviating from the majority of samples were excluded based on principal component analysis. Gene expression data were log-transformed for subsequent analyses. The expression of the MAL gene and its homologs varied substantially across samples. Samples were categorized into low-, medium-, and high-expression groups based on MAL expression levels. Significant differences in pathological subtypes and clinical stages were observed among these groups. Moreover, MAL expression levels exhibited distinct patterns in certain more aggressive tumor subtypes.Finally, we identified genes whose expression patterns were highly synchronized with MAL and constructed a potential co-expression network to infer the biological processes in which MAL may be involved. Genes with an absolute correlation coefficient $|r| > 0.3$ and $p < 0.05$ were retained. Visualization of the results was performed using the Matplotlib and Seaborn libraries in Python.

Immune infiltration profiling

The ssGSEA algorithm in the Xiantao Academic platform, was employed to quantify immune infiltration profiles in the TCGA-UCEC cohort. Specifically we assessed the following: 1) differences in immune cell infiltration associated with high versus low *MAL* expression, and 2) expression variations across 23 immune cell types. Additional ssGSEA analysis of paired tumor and adjacent normal tissues further characterized tumor immune microenvironment alterations. Correlations between *MAL* expression and immune cell abundance were visualized via scatter plots.

Drug sensitivity profiling and molecular docking

To identify potential sensitive agents targeting MAL, we utilized the CADSP online database (https://smuonco.shinyapps.io/CADSP/) and identified MAL as exhibiting sensitivity to Navitoclax (abt-263) treatment.MAL-associated therapeutic agents were identified via the CPADS database (https://smuonco.shinyapps.io/CADSP/). MAL protein sequence data were retrieved from UniProt (https://www.uniprot.org/uniprotkb), and the corresponding crystal structure was downloaded from the RCSB PDB database (https://www.rcsb.org/). Structural optimization of the MAL protein using Discovery Studio 2019 included dehydration, hydrogenation, charge assignment, amino acid repair, and side-chain correction to generate an optimized PDB file. The navitoclax compound structure was obtained from PubChem and energy-minimized in Discovery Studio 2019. Protein–ligand docking was conducted as follows: 1) PDBQT file preparation via AutoDock 1.5.6., 2) molecular docking using AutoDock Vina 1.2.6., and 3) binding visualization through PyMOL 3.1 and Discovery Studio 2019.

Molecular dynamics analysis

This study employed Gromacs 2022 for molecular dynamics (MD) simulations, wherein force field parameters were obtained using the pdb2gmx tool and AutoFF web server. The CHARMM36 force field [15] was applied to the receptor protein, while ligand parameters were derived from the CGenff force field. The system was solvated within a cubic TIP3P water box extending 1 nm from the protein surface [16], with ionic species added via the gmx genion tool to achieve charge neutrality. Long-range electrostatic interactions were treated using the Particle Mesh Ewald (PME) method with a 1 nm cutoff distance. All bond constraints were implemented through the SHAKE algorithm, and the Verlet leapfrog integrator was utilized with a 1 fs timestep. Prior to production MD, the system underwent energy minimization involving 3000 steps of steepest descent followed by 2000 steps of conjugate gradient optimization, performed sequentially under three conditions: (1) solute-constrained water minimization, (2) counterion-constrained minimization, and (3) full-system minimization. The NPT ensemble simulations were conducted at 310 K for 100 ns, during which structural properties including root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), hydrogen bonds (HBonds), radius of gyration (Rg), and solvent-accessible surface area (SASA) were calculated using respective Gromacs analysis tools (g-rmsd, g-rmsf, g-hbond, g-Rg, and g-sasa).

Cell culture

EEC and ECC-1 cells were obtained from Procell Life Science and cultured in Dulbecco's modified Eagle medium (DMEM; Gibco) supplemented with 10% fetal bovine serum and 1% penicillin–streptomycin at 37 °C in a humidified 5% CO_2 incubator.

Lentiviral transfection

Short hairpin RNA (shRNA) targeting MAL (MAL-ShRNA) and MAL overexpression (MAL-OE) lentiviruses were synthesized by Genepharma (Shanghai, China). The target sequences were as follows:

MAL-shRNA(Homo,puro):
5′-TGGGTGATGTTCGTGTCTGTG-3′;
Sh-negative control(Homo,puro):
5′-UUCUCCGAACGUGUCACGUTT-3′;

ECC-1 cells were trypsinized and seeded in T25 flasks at 80% confluency 24 h prior to transfection. For transduction, 60 μL of viral supernatant was mixed with 5 μg/mL polybrene, and 400 μL of fresh DMEM. This mixture was applied to cells after culture medium replacement. Following 16–24 h of incubation at 37 °C under 5% CO_2, cells were returned to complete medium and allowed to recover for 24–48 h. Stable transfectants were selected with 2 μg/mL puromycin (Beyotime Biotechnology, ST551-10 mg) for 48 h.

Western blotting analysis

Proteins were extracted using the Beyotime Extraction Kit (P0033) and quantified with a BCA Assay Kit (Beyotime, P0010). Lysates mixed with loading buffer underwent 5–10 min of denaturation at 95 °C. Proteins were electrophoresed in 12% precast gels (BOSTER AR1203) and transferred to polyvinylidene difluoride membranes. Membranes were blocked with 5% bovine serum albumin for 1–2 h, followed by overnight incubation in primary antibodies at 4 °C (1:1000 dilution). After washing in TBST, membranes were incubated with secondary antibodies for 1–2 h at room temperature. Bands were visualized using enhanced chemiluminescence reagent (Beyotime, P0018S). GAPDH served as the internal control. Experiments were conducted independently in triplicate ($n = 3$), and protein expression levels were quantified using ImageJ. Statistical significance was defined as $p < 0.05$.

EdU cell proliferation assay

Cell proliferation was measured using the EdU Cell Proliferation Detection Kit (Beyotime Biotechnology, C0071S) following the manufacturer's protocol. Fluorescence images were acquired at 100 × magnification. Proliferating cell counts were quantified using ImageJ across three independent biological replicates. $P < 0.05$ was defined as statistically significant.

Matrigel invasion assay

Transwell invasion assays were performed using 8-μm pore, 6.5-mm polycarbonate Transwell inserts (Corning, NY, USA) precoated with Matrigel (BD Biosciences, MA, USA) and preincubated for 6 h at 37 °C. Specifically, 5×10^4 ECC-1 cells were seeded into the upper chambers and cultured for 48 h. Post-incubation, invaded cells were fixed with 4% paraformaldehyde and stained with 0.5% crystal violet (Beyotime). Cells were imaged under a light microscope (100 ×) across five randomly selected fields per chamber. Invasion was quantified using ImageJ across three independent experiments. Statistical significance was set at $p < 0.05$).

Statistical analysis

Data that conformed to a normal distribution are presented as mean ± standard deviation (SD). Statistical analyses were performed using GraphPad Prism version 9.5 software. Multiple testing corrections were applied to the following analyses: differential expression analysis, GO/KEGG enrichment analysis, GSEA enrichment analysis, and survival analysis. For group comparisons, the unpaired two-tailed Student's t-test was used for comparisons between two groups, and two-way analysis of variance (ANOVA) was employed for comparisons among multiple groups. p-value < 0.05 was considered statistically significant. Significance levels are denoted as follows: "*" *for* $p < 0.05$, "**" for $p < 0.01$, "***" for $p < 0.001$, and "****" for $p < 0.0001$.

Results

Identification and functional enrichment analysis of differentially expressed genes

Batch correction of the integrated GSE17025/GSE63678 datasets (Fig. 1A, B) identified 167 differentially expressed genes (DEGs; 100 upregulated, 67 downregulated;), which are summarized in a volcano plot (Fig. 1C). Gene Ontology (GO) enrichment analysis revealed that the DEGs were enriched in several categories. For biological processes, the top enrichments included sensory organ morphogenesis, pattern specification, embryonic organ development, nuclear division, and organelle fission. For cellular components, enriched terms comprised microtubule-associated complex, axonemal dynein complex, outer dynein arm, lysosomal lumen, and dynein complex. For molecular functions, significant enrichments included cytoskeletal motor activity, microtubule motor activity, DNA-binding transcription activator activity (RNA polymerase II-specific), and serine-type endopeptidase activity. No significant enrichments were found in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Fig. 1D–F; Table 1). A co-expression heatmap (Fig. 1G) illustrates the relationships between the top five enriched terms from each of the three GO categories and the differentially expressed genes (DEGs).

Identification of related WGCNA modules

WGCNA was performed on integrated UCEC and normal endometrial datasets from GSE17025/GSE63678, establishing a scale-free network at soft threshold power 6 (Fig. 2A, B) and constructing a clustering dendrogram (Fig. 2C) to identify disease-stage–associated gene modules. Sixteen modules were identified (Fig. 2D), with nonclustered genes excluded (gray); modules exhibiting

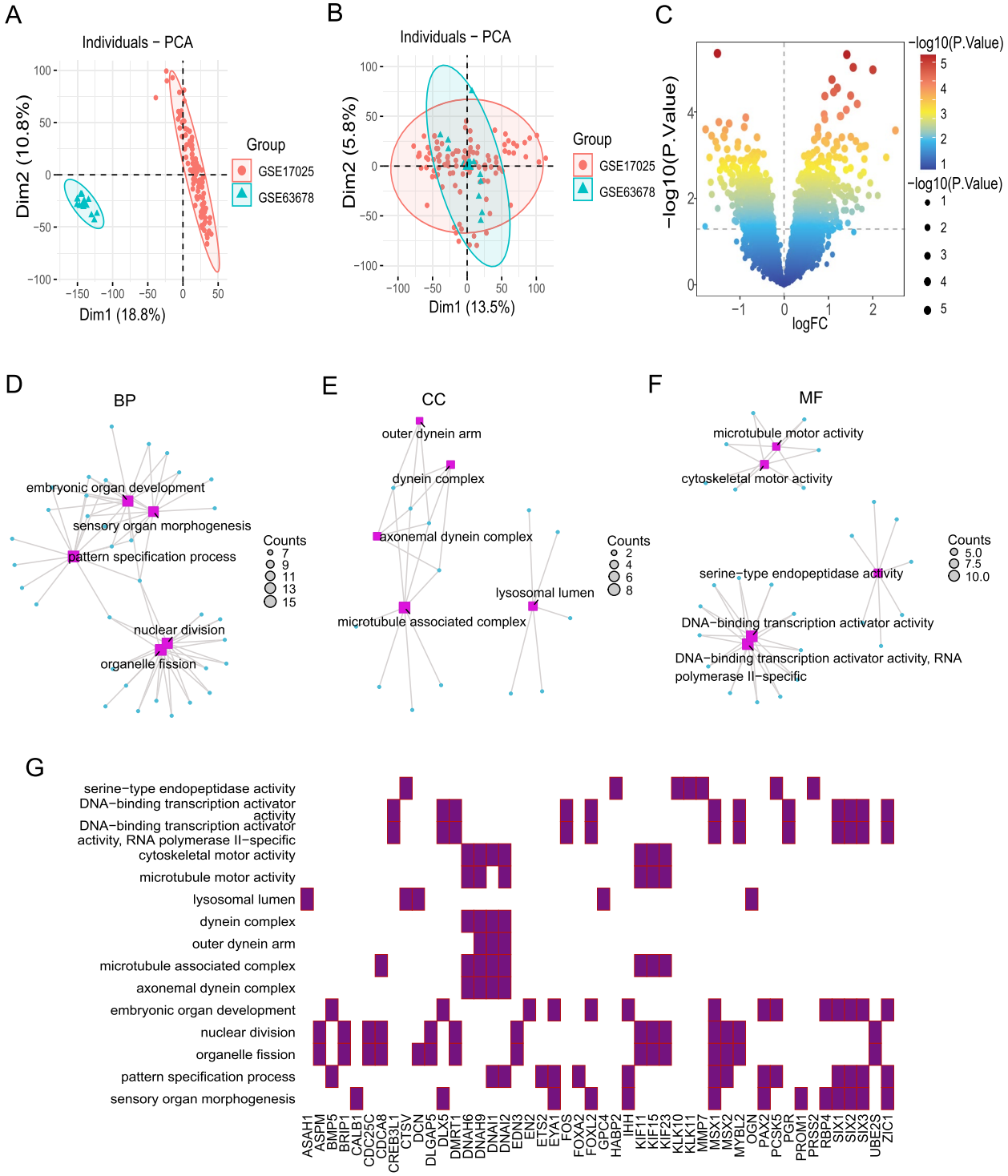

Fig. 1 Identification and functional enrichment analysis of DEGs. **A**, **B** Principal component analysis of GSE17025 and GSE63678 datasets after batch effect correction. **C** Volcano plot of DEGs. **D**–**F** GO functional enrichment bubble plots. **G** Co-expression heatmap of GO functional enrichment top 5 and related genes

Table 1 GO enrichment analysis of DEGs (TOP5)

Ontology	ID	Description	GeneRatio	BgRatio	*p* value	p.adjust
BP	GO:0090596	sensory organ morphogenesis	13/140	266/18800	9.87e-08	0.0002
BP	GO:0007389	pattern specification process	15/140	463/18800	1.94e-06	0.0024
BP	GO:0048285	organelle fission	15/140	493/18800	4.18e-06	0.0034
BP	GO:0000280	nuclear division	14/140	446/18800	6.25e-06	0.0034
BP	GO:0048568	embryonic organ development	14/140	449/18800	6.75e-06	0.0034
CC	GO:0005858	axonemal dynein complex	4/146	23/19594	2.35e-05	0.0032
CC	GO:0005875	microtubule associated complex	8/146	160/19594	2.69e-05	0.0032
CC	GO:0036157	outer dynein arm	3/146	11/19594	6.4e-05	0.0051
CC	GO:0030286	dynein complex	4/146	54/19594	0.0007	0.0381
CC	GO:0043202	lysosomal lumen	5/146	97/19594	0.0008	0.0381
MF	GO:0003777	microtubule motor activity	6/144	67/18410	1.39e-05	0.0048
MF	GO:0003774	cytoskeletal motor activity	7/144	111/18410	2.67e-05	0.0048
MF	GO:0001228	DNA-binding transcription activator activity, RNA polymerase II-specific	12/144	462/18410	0.0003	0.0273
MF	GO:0001216	DNA-binding transcription activator activity	12/144	466/18410	0.0003	0.0273
MF	GO:0004252	serine-type endopeptidase activity	7/144	174/18410	0.0004	0.0318

positive tumor–tissue correlations ($R > 0.23$) were prioritized, including the green (779 genes, $R = 0.22$, $p < 0.02$) and pink (270 genes, $R = 0.21$, $p < 0.03$) modules, collectively yielding 1,049 candidate genes visualized in module–trait correlation scatterplots (Fig. 2E, F).

Machine learning-based biomarker discovery

The Venn diagram illustrates the 33 overlapping genes identified from WGCNA and DEGs (Fig. 3A). After analyzing these 33 genes using eight machine learning models (GBM, KNN, LASSO, SVM, NNET, RF, DT, and GLM), the absolute residual plots (Fig. 3B, C) display the residual values for each model, with the GBM, KNN, and LASSO models showing the smallest residuals. The ROC curves (Fig. 3D) present the AUC values of the eight different models (RF: 0.897, SVM: 0.959, GLM: 0.883, GBM: 0.993, KNN: 0.990, NNET: 0.972, LASSO: 0.98, DT: 0.683), indicating that RF, SVM, GLM, GBM, KNN, NNET, and LASSO have substantial discriminatory power. The feature importance plot (Fig. 3E) highlights GBM, KNN, and LASSO as the most influential models. Subsequently, we extracted the top 12 genes from each of the GBM, KNN, and LASSO models and visualized their intersection using a Venn diagram (Fig. 3F). This resulted in the identification of six shared genes: MAL, SCGB1D2, MICA, GABRP, HPGD, and MTCL1.

Survival analysis of key genes

Kaplan–Meier survival analysis using data from The Cancer Genome Atlas Uterine Corpus Endometrial Carcinoma (TCGA-UCEC) cohort revealed distinct prognostic associations. Elevated *MAL* expression was associated with significantly reduced overall survival [OS; hazard ratio (HR) = 2.34, 95% confidence interval (CI) = 1.51–3.62, $^*p^* < 0.001$] and progression-free survival [PFS; HR = 1.85, 95% CI = 1.29–2.65, $^*p^* < 0.001$] (Fig. 4A, B). Conversely, *SCGB1D2* overexpression was linked to improved OS [HR = 0.41, 95% CI = 0.26–0.63, $^*p^* < 0.001$] and PFS [HR = 0.57, 95% CI = 0.40–0.81, $^*p^* = 0.002$] (Fig. 4C, D). No significant survival correlations were observed for the remaining biomarkers: *MICA* (OS: HR = 0.89, $^*p^* = 0.563$; PFS: HR = 1.11, $^*p^* = 0.559$), *GABRP* (OS: HR = 0.83, $^*p^* = 0.371$; PFS: HR = 0.79, $^*p^* = 0.177$), *HPGD* (OS: HR = 0.89, $^*p^* = 0.578$; PFS: HR = 0.92, $^*p^* = 0.626$), and *MTCL1* (OS: HR = 1.13, $^*p^* = 0.553$; PFS: HR = 0.92, $^*p^* = 0.650$) (Fig. 4E–L). These findings implicate *MAL* and *SCGB1D2* as a potential tumor promoter and suppressor, respectively, in UCEC pathogenesis; thus, *MAL* was prioritized for further mechanistic analysis.

Correlation between MAL expression and clinicopathological features in patients with UCEC

Using the Xiantao Academic online analytics platform, we assessed the relationship between MAL expression levels and clinicopathological traits in the TCGA-UCEC cohort. MAL was significantly overexpressed in tumor tissues versus controls ($p < 0.001$; Fig. 5A). Diagnostic ROC analysis showed moderate discriminative power (AUC = 0.732, 95%CI = 0.661–0.803; Fig. 5B), although time-dependent AUC curves indicated decreasing prognostic utility (AUC < 0.7; Fig. 5C), with specific 1-, 3-, and 5-year ROC AUCs of 0.596, 0.650, and 0.649, respectively (Fig. 5D). MAL overexpression correlated significantly ($p < 0.01$) with eight UCEC clinical features: poor primary therapy response [progressive disease (PD)/stable disease (SD) vs. complete response (CR)/partial response (PR), $p < 0.01$], advanced stage (III–IV vs. I–II, $p < 0.001$), age (≤ 60 vs. > 60, $p < 0.001$), body mass index (BMI; ≤ 30 vs. > 30, $p < 0.001$), nonendometrioid histology (mixed/serous vs. endometrioid, $p < 0.001$), extensive tumor

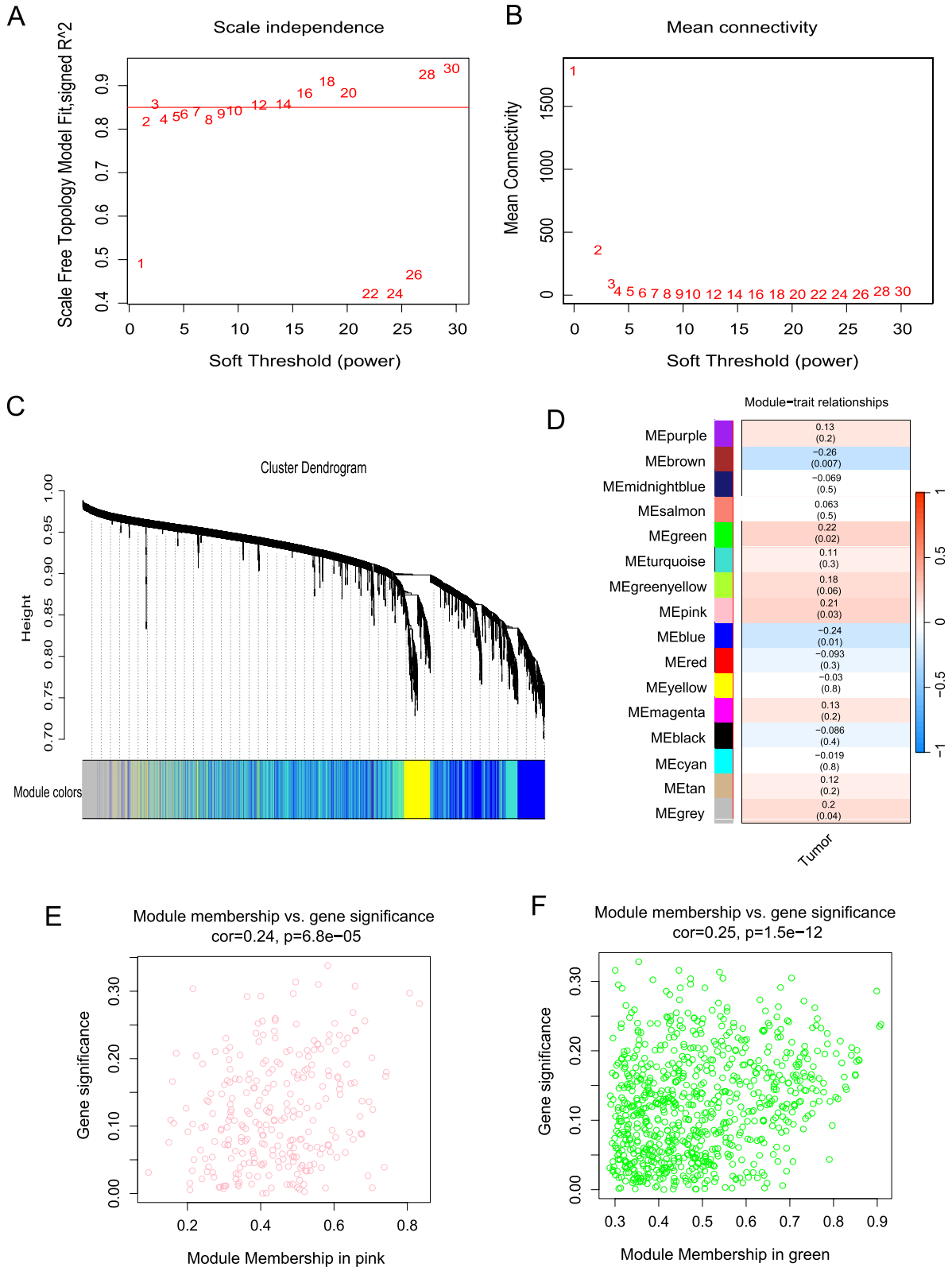

Fig. 2 Identification of gene modules via WGCNA. **A**, **B** Soft threshold power selection in WGCNA. **C** Module identification via dynamic tree cutting. **D** Heatmap of module–trait correlations. **E** Module-related scatter plot

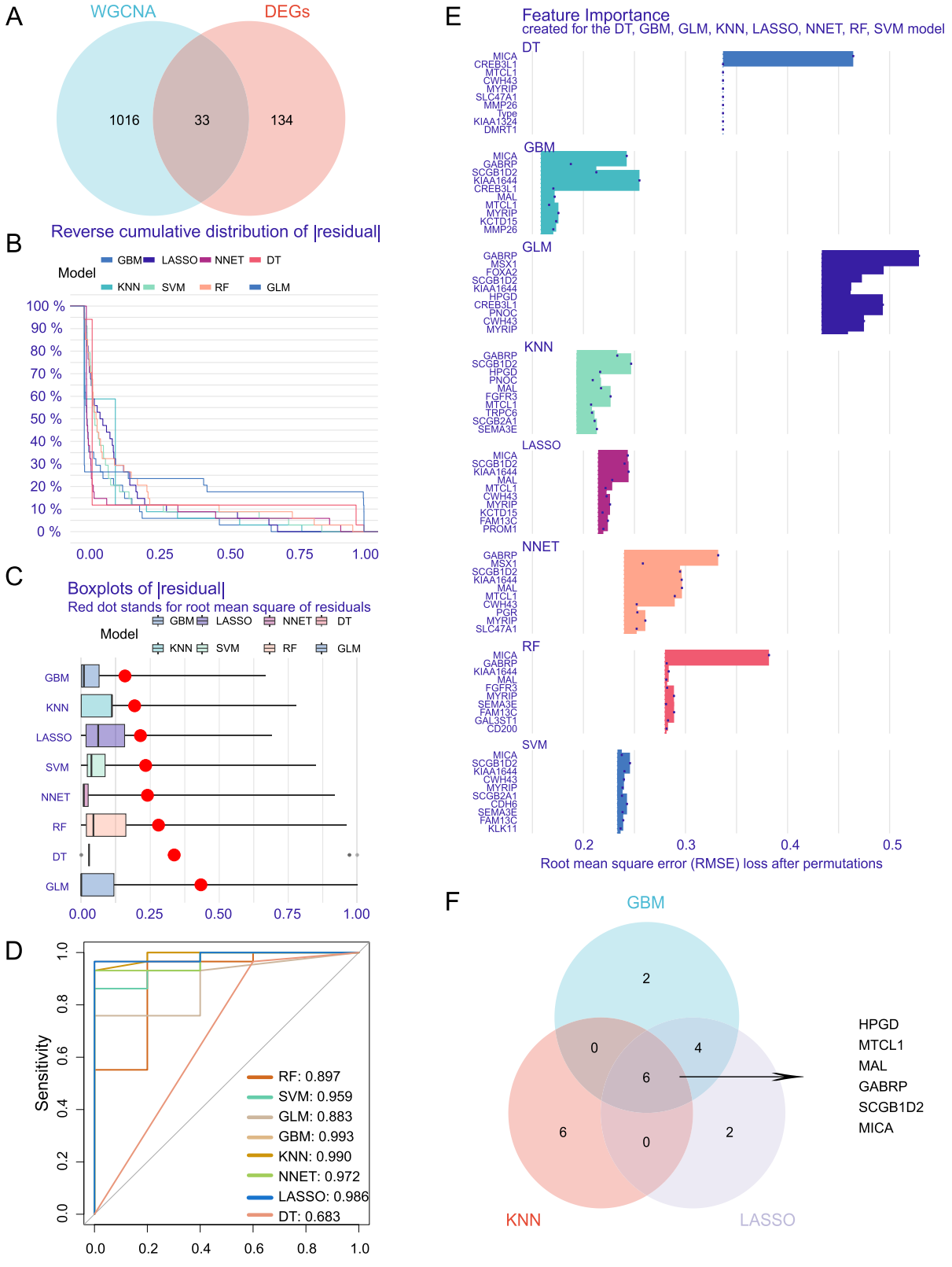

Fig. 3 Biomarker discovery using machine learning models. **A** Venn diagram of WGCNA-derived genes and DEGs. **B**, **C** Absolute residual error values across the eight machine learning models. **D** ROC curves evaluating diagnostic performance across eight machine learning models: GBM, KNN, LASSO, SVM, NNET, RF, DT, GLM. **E** Feature importance ranking across models. **F** Venn diagram of shared features among the three top-performing models: GBM, KNN, and LASSO

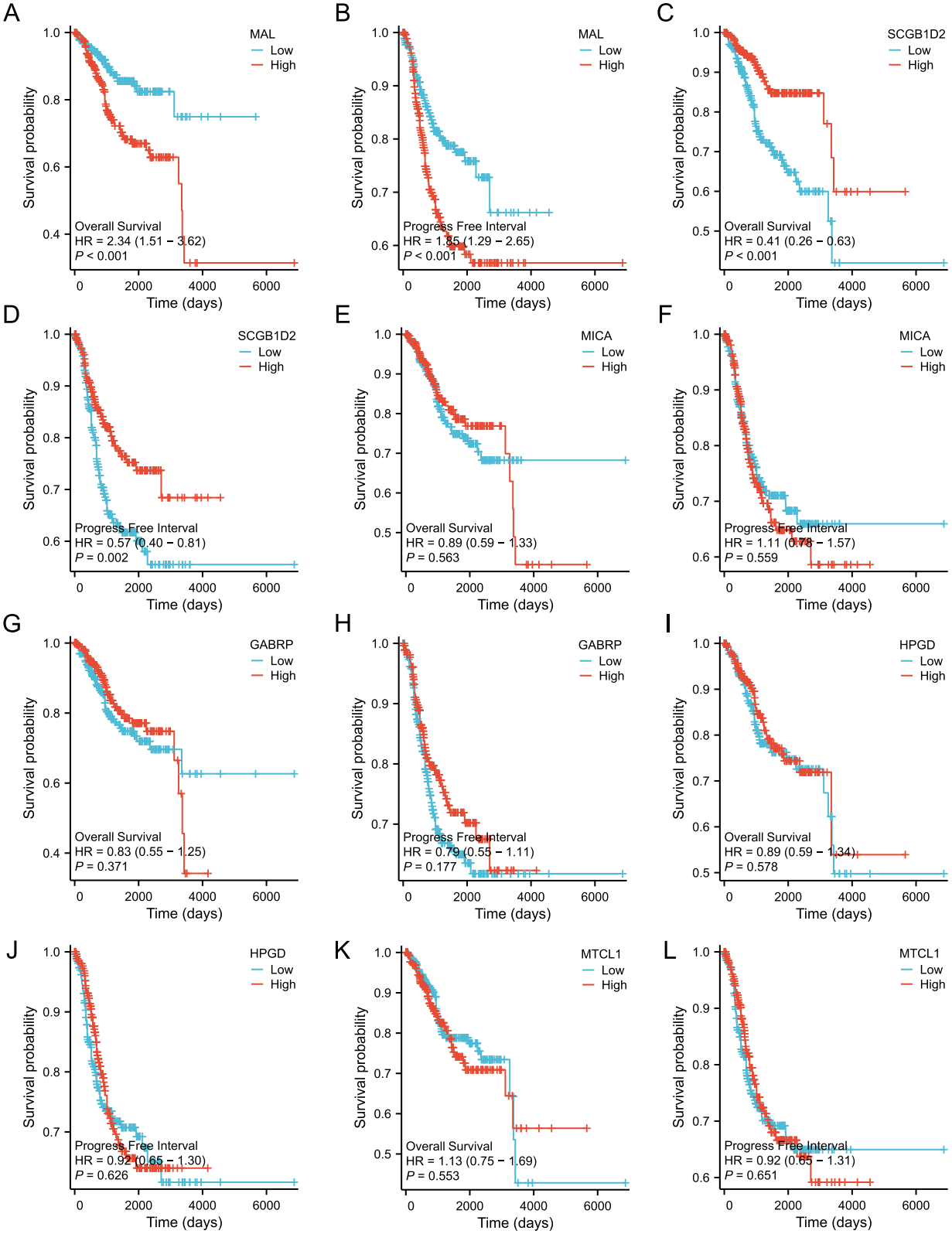

Fig. 4 Survival analysis of key genes. OS and PFS curves for (**A**, **B**) MAL, (**C**, **D**) SCGB1D2, (**E**, **F**) MICA, (**G**, **H**) GABRP, (**I**, **J**) HPGD, and (**K**, **L**) MTCL1.

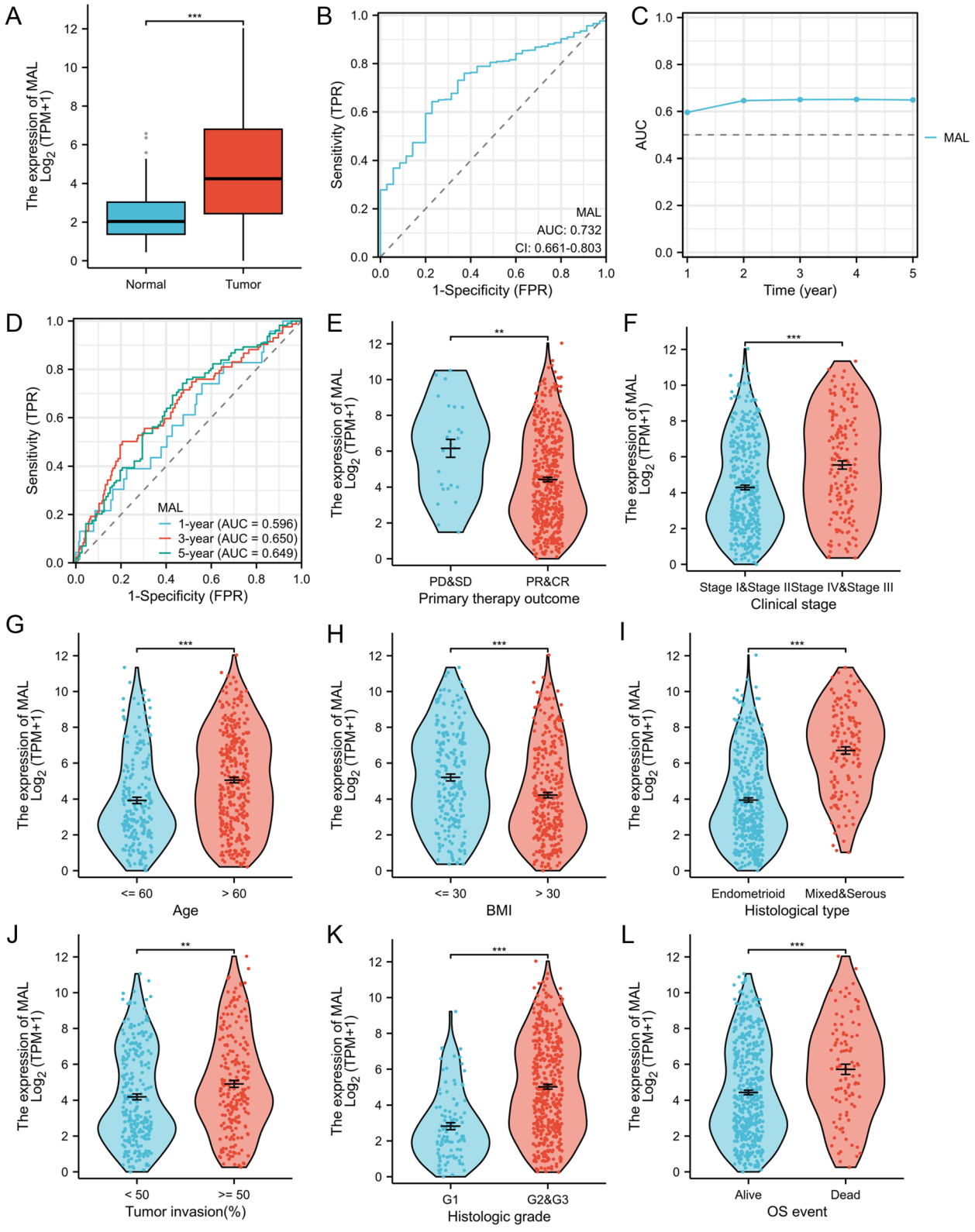

Fig. 5 Correlation between MAL expression and clinicopathological features in UCEC. **A** Analysis of MAL mRNA expression across 35 normal and 554 UCEC tumor samples from TCGA using paired and unpaired methods after TPM normalization. **B** ROC curve evaluating MAL diagnostic performance. **C**, **D** Time-dependent AUC and ROC curves for MAL. **E**–**L** MAL expression across clinical subgroups: initial treatment outcome, clinical stage, age, BMI, histologic type, tumor infiltration depth, tumor grade, and OS event status. *, **, and ***: $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively

invasion (≥ 50% vs. < 50%, $p < 0.01$), high tumor grade (G2/G3 vs. G1, $p < 0.001$), and poor survival outcomes [OS: alive vs. dead, $p < 0.001$; disease-specific survival: yes vs. no, $p < 0.001$; progression-free interval (PFI): yes vs. no, $p < 0.001$] (Figs. 5E–L and S1A, B). Under varied pathological conditions, high MAL expression adversely impacted patient prognosis (Fig. 6A–L). Across clinical stages I to IV, elevated MAL levels were associated with reduced OS (HR = 2.43, 95%CI = 1.55–3.82, $p < 0.001$). Similarly, OS was poorer in patients with elevated MAL expression and endometrioid, mixed, and serous histologies (HR = 2.34, 95%CI = 1.51–3.62, $p < 0.001$), different residual tumor statuses (R0: complete resection; R1: microscopic residues; R2: gross residue; HR = 2.29, 95%CI = 1.37–3.82, $p = 0.002$), with or without radiotherapy (HR = 2.51, 95%CI = 1.39–3.96, $p < 0.001$), and with or without hormone therapy (HR = 3.17, 95%CI = 1.74–5.76, $p < 0.001$). With high MAL expression, OS also declined in both diabetic and nondiabetic patients (HR = 2.71, 95%CI = 1.66–4.42, $p < 0.001$), cases with tumor infiltration < 50% or ≥ 50% (HR = 2.45, 95%CI = 1.50–3.97, $p < 0.001$), among premenopausal perimenopausal, and postmenopausal women (HR = 2.44, 95%CI = 1.57–3.81, $p < 0.001$), and across ethnic groups (Asian, black, African American, and White: HR = 2.26, 95%CI = 1.45–3.54, $p < 0.001$). Single-gene logistic regression (Table 2) showed that MAL was positively correlated with clinical stage [odds ratio (OR) = 2.012, 95%CI = 1.380–2.933, $p < 0.001$], age (OR = 2.496, 95% CI = 1.749–3.563, $p < 0.001$), histological type (OR = 5.892, 95%CI = 3.745–9.270, $p < 0.001$), tumor invasion (%; OR = 1.502, 95%CI = 1.044–2.159, $p = 0.028$), and histologic grade (OR = 4.700, 95%CI = 2.803–7.881, $p < 0.001$) but negatively associated with BMI (OR = 0.506, 95%CI = 0.355–0.721, $p < 0.001$).

High MAL expression is not an independent prognostic factor for OS

Comprehensive univariate and multivariate Cox regression analyses identified clinical stage (HR = 2.87, 95%CI = 1.243–6.657, $p = 0.014$), primary therapy outcome (HR = 0.271, 95%CI = 0.084–0.872, $p = 0.028$), histologic grade (HR = 9.232, 95%CI = 1.203–70.853, $p = 0.033$), and radiation therapy (HR = 0.357, 95%CI = 0.185–0.692, $p = 0.002$) as independent prognostic factors for UCEC (Table 3). Despite prior strong associations, MAL expression was not independently predictive in multivariate analysis (HR = 0.955, 95%CI = 0.474–1.924, $p = 0.899$). These four validated predictors were used to construct a nomogram estimating 1-, 3-, and 5-year OS probabilities (Fig. 7A). Calibration plots showed strong agreement between predicted and actual outcomes across timepoints (Fig. 7B), supporting the nomogram's accuracy for UCEC mortality risk stratification.

Potential mechanism analysis of MAL

Based on the expression levels of MAL in TCGA-UCEC, all samples were divided into high and low expression groups for differential analysis. The differential expression profile is visualized in a volcano plot (Fig. 8A), and a co-expression heatmap (Fig. 8B) illustrates the distribution of the top 15 up-regulated and top 15 down-regulated genes relative to MAL expression. Functional enrichment analysis was performed on all differentially expressed genes. The results indicated that these genes are associated with 309 biological processes (BP), 64 cellular components (CC), 92 molecular functions (MF), and 35 KEGG pathways (Supplementary Table S5). Bubble plots were used to display the top five enriched categories for each of the BP, CC, MF, and KEGG domains (Figs. 8C-F), with significance assessed based on the adjusted Q-value.The GO terms revealed that biological processes are mainly involved in cilium movement, cilium-dependent cell motility, cilium or flagellum-dependent cell motility, pattern specification process, and axoneme assembly. Cellular components were primarily associated with ciliary plasm, axoneme, ion channel complex, cation channel complex, and motile cilium. Molecular functions mainly included receptor ligand activity, signaling receptor activator activity, metal ion transmembrane transporter activity, voltage-gated channel activity, and voltage-gated ion channel activity. Key pathways identified were Neuroactive ligand-receptor interaction, cAMP signaling pathway, Thyroid hormone synthesis, Calcium signaling pathway, and Bile secretion. Gene Set Enrichment Analysis (GSEA) of the differentially expressed genes related to MAL identified a total of 23 significantly enriched pathways ($|NES| > 1$, $p < 0.05$), comprising 8 positively correlated pathways ($NES > 1$) and 15 negatively correlated pathways ($NES < -1$) (Supplementary Table S6). The top five pathways positively correlated with MAL were: glucuronidation, ascorbate and aldarate_metabolism, aspirin adme, pentose and glucuronate interconversions, and porphyrin and chlorophyll metabolism (Fig. 8G). The top five pathways negatively correlated with MAL were: FCGR activation, initial triggering of complement, creation of C4 and C2 activators, complement cascade, and scavenging of heme from plasma (Fig. 8H). These findings suggest that pathways regulating glucuronidation, ascorbate and aldarate metabolism, and pentose and glucuronate interconversions play significant roles in MAL-mediated progression of UCEC. Subsequent analysis revealed that MAL exhibits high expression heterogeneity in endometrial cancer (mean: 137.33, median: 18.01). Based on its expression levels, patients could be stratified into low (40.3%), intermediate (32.9%), and high (26.8%) expression subgroups, indicating its potential as a molecular subtype biomarker (Supplementary File, Figure A). Co-expression network

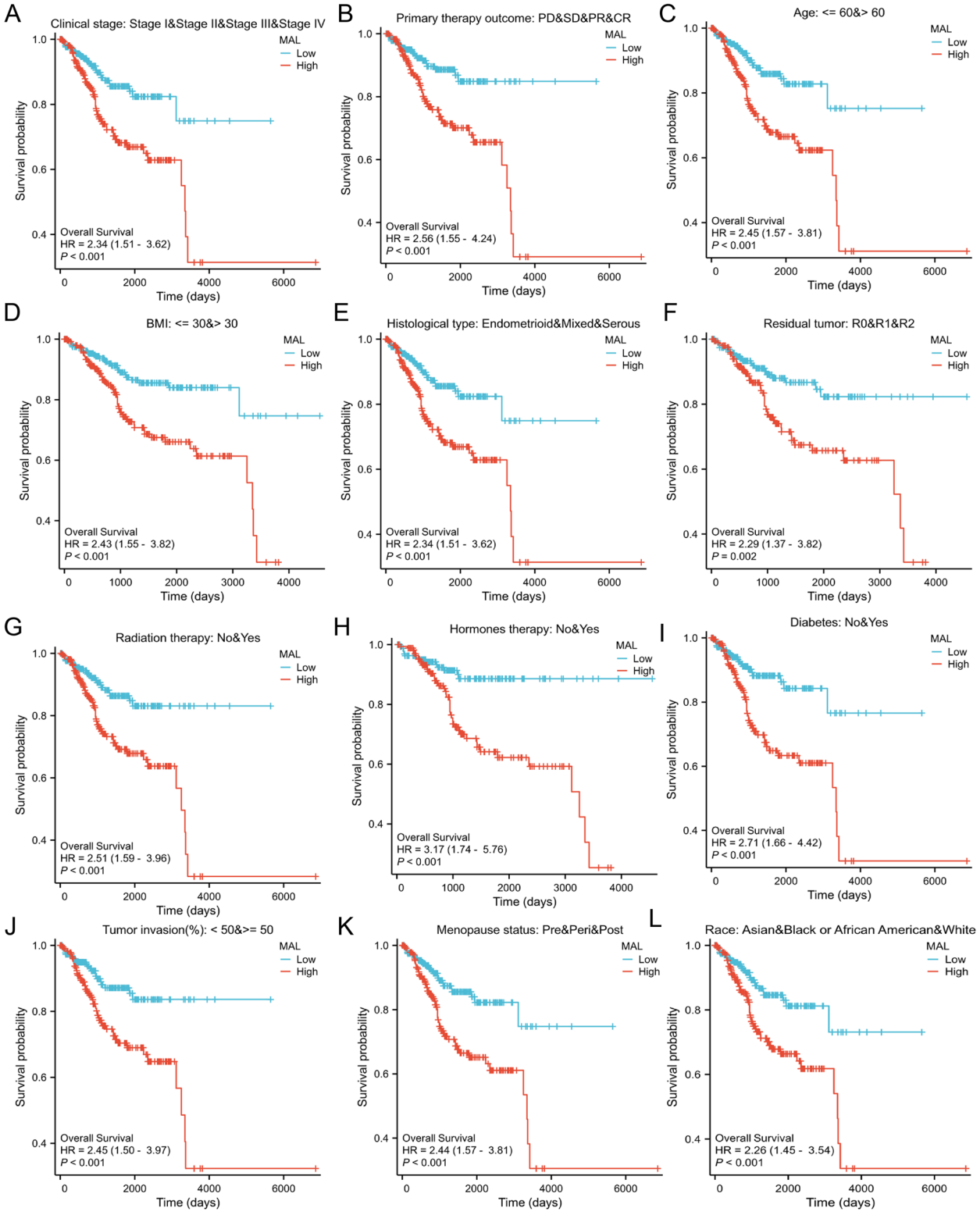

Fig. 6 Prognostic analysis of MAL expression and clinicopathological features. **A**–**L** Prognostic analysis of OS across diverse pathological features in UCEC

Table 2 Single gene logistic regression analysis of MAL in patients with UCEC

Characteristics	Total (N)	OR (95% CI)	*P* value
Clinical stage (Stage III&Stage IV vs. Stage I&Stage II)	554	2.012 (1.380–2.933)	< 0.001
Primary therapy outcome (CR&PR vs. PD&SD)	482	0.476 (0.208–1.091)	0.080
Age (> 60 vs. < = 60)	551	2.496 (1.749–3.563)	< 0.001
BMI (> 30 vs. < = 30)	521	0.506 (0.355–0.721)	< 0.001
Histological type (Mixed&Serous vs. Endometrioid)	554	5.892 (3.745–9.270)	< 0.001
Radiation therapy (Yes vs. No)	529	1.241 (0.882–1.747)	0.216
Hormones therapy (Yes vs. No)	346	0.973 (0.525–1.802)	0.931
Diabetes (Yes vs. No)	453	0.859 (0.568–1.299)	0.472
Tumor invasion(%) (> = 50 vs. < 50)	476	1.502 (1.044–2.159)	0.028
Histologic grade (G2&G3 vs. G1)	543	4.700 (2.803–7.881)	< 0.001
Residual tumor (R1&R2 vs. R0)	415	1.817 (0.912–3.621)	0.089

analysis demonstrated that MAL is significantly positively correlated with 251 genes. Key partners include the calcium channel gene CACNA2D2 ($R = 0.399$) and the cytoskeleton regulatory gene CAPG ($R = 0.340$). These co-expressed genes are collectively enriched in functions related to ion channel activity, cytoskeletal regulation, and signal transduction pathways (Supplementary File, Figure B-E). Mechanistically, its expression may be regulated by specific transcription factors (such as GATA/ETS families) or epigenetic modifications. The strong association with calcium channel genes suggests its potential integration into calcium signaling pathways. Downstream functional predictions indicate that MAL may regulate cytoskeletal remodeling (e.g., cooperating with CAPG to influence pseudopodia formation) and modulate ion channel signaling (e.g., cooperating with CACNA2D2 to affect calcium influx), thereby potentially influencing tumor cell migration, invasion, and epithelial-mesenchymal transition (EMT). This provides a theoretical foundation for further investigation of MAL as a prognostic biomarker or therapeutic target. However, these findings are primarily based on transcriptomic data from public databases and cellular-level experiments, lacking validation in animal models and at the histological level. Future studies will incorporate animal experiments, immunohistochemical analysis, and validation in larger clinical cohorts to further elucidate the role of MAL in the development and progression of endometrial cancer and to explore its potential clinical applications.

Correlation between MAL expression and immune characteristics

Pan-cancer single-sample gene set enrichment analysis (ssGSEA) established MAL as an immunoregulator (Fig. 9A). In UCEC, MAL expression was positively correlated with activated dendritic cells (aDCs; $R = 0.240$), macrophages ($R = 0.198$), T helper type 2 (Th2) cells ($R = 0.136$), B cells ($R = 0.124$), and Th1 cells ($R = 0.113$) but negatively associated with regulatory T cells (Tregs; $R = -0.117$), Th17 ($R = -0.099$), T follicular helper (TFH) cells ($R = -0.123$), cytotoxic effectors [central memory T (Tcm) cells/T cells; $R = -0.121$), and innate populations [CD56bright natural killer (NK) cells: $R = -0.157$; mast cells: $R = -0.156$; eosinophils: $R = -0.281$; immature dendritic cells (iDCs): $R = -0.143$], with significance at $p < 0.05$ in all cases (Figs. 9B, D–I and S1C–K). To further investigate immune infiltration, MAL-high and MAL-low expression groups were compared using ssGSEA. Immune cell types with significant differences were as follows (Fig. 9C): aDCs ($p < 0.001$), CD8 + T cells ($p < 0.05$), eosinophils ($p < 0.001$), iDCs ($p < 0.001$), macrophages ($p < 0.001$), mast cells ($p < 0.01$), CD56bright NK cells ($p < 0.01$), plasmacytoid dendritic cells ($p < 0.01$), T cells ($p < 0.001$), Th cells ($p < 0.01$), Tcm cells ($p < 0.05$), TFH cells ($p < 0.001$), and Tregs ($p < 0.01$). Scatterplots were generated to highlight significant correlations between MAL expression and immune cells (Figs. 9D–I and S1C–K): Tregs ($R = -0.117$, $p = 0.006$), Th2 cells ($R = 0.136$, $p = 0.001$), Th17 cells ($R = -0.099$, $p = 0.020$), Th1 cells ($R = 0.113$, $p = 0.008$), TFH cells ($R = -0.123$, $p = 0.004$), Tcm cells ($R = -0.120$, $p = 0.005$), aDCs ($R = 0.240$, $p < 0.001$), B cells ($R = 0.124$, $p = 0.004$), Th cells ($R = -0.123$, $p = 0.004$), T cells ($R = -0.122$, $p = 0.004$), CD56bright NK cells ($R = -0.157$, $p < 0.001$), mast cells ($R = -0.156$, $p < 0.001$), eosinophils ($R = -0.281$, $p < 0.001$), macrophages ($R = 0.198$, $p < 0.001$), and iDCs ($R = -0.1431$, $p < 0.001$). These results indicate that MAL expression drives an immunosuppressive microenvironment, promoting Th2/Treg bias while suppressing cytotoxic and innate responses, thereby positioning MAL as a viable immunotherapeutic target in UCEC.

Docking analysis and molecular dynamics analysis of MAL and navitoclax

Navitoclax (abt-263) is an effective oral active Bcl-2 family protein inhibitor, which can bind to a variety of anti apoptotic Bcl-2 family proteins and inhibit tumor progression. Navitoclax treatment efficacy differed significantly between the MAL-high and MAL-low expression groups ($p = 0.016$; Fig. 10A), prompting molecular docking analysis to elucidate potential targeting mechanisms. The ligand–protein interface revealed critical interactions: Van der Waals forces stabilized the binding conformation while four hydrogen bonds formed between

Table 3 Univariate and multivariate Cox regression analyses of clinical characteristics associated with OS of UCEC in TCGA

	Total(N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value
Clinical stage	553				
Stage I	342	Reference		Reference	
Stage II	52	1.738 (0.833—3.626)	0.140	0.815 (0.183—3.630)	0.789
Stage III	130	3.084 (1.911—4.977)	< 0.001	2.876 (1.243—6.657)	0.014
Stage IV	29	8.082 (4.497—14.524)	< 0.001	2.432 (0.715—8.267)	0.155
Primary therapy outcome	482				
PD	20	Reference		Reference	
SD	6	0.583 (0.166—2.052)	0.401	0.429 (0.049—3.791)	0.446
PR	12	0.790 (0.297—2.096)	0.635	1.767 (0.370—8.435)	0.475
CR	444	0.110 (0.060—0.204)	< 0.001	0.271 (0.084—0.872)	0.028
Age	551				
< = 60	207	Reference		Reference	
> 60	344	1.850 (1.162—2.944)	0.009	1.784 (0.826—3.856)	0.141
BMI	520				
< = 30	211	Reference			
> 30	309	0.963 (0.634—1.464)	0.861		
Histological type	553				
Endometrioid	411	Reference		Reference	
Mixed & Serous	142	2.636 (1.751—3.968)	< 0.001	1.329 (0.647—2.728)	0.439
Histologic grade	542				
G1	99	Reference		Reference	
G2 & G3	443	11.641 (2.864—47.318)	< 0.001	9.232 (1.203—70.853)	0.033
Radiation therapy	529				
No	281	Reference		Reference	
Yes	248	0.596 (0.387—0.919)	0.019	0.357 (0.185—0.692)	0.002
Diabetes	453				
No	329	Reference			
Yes	124	1.167 (0.728—1.871)	0.520		
Hormones therapy	346				
No	299	Reference			
Yes	47	0.804 (0.381—1.695)	0.566		
Tumor invasion (%)	475				
< 50	261	Reference		Reference	
> = 50	214	2.825 (1.752—4.554)	< 0.001	1.219 (0.602—2.470)	0.582
Residual tumor	414				
R0	376	Reference		Reference	
R1 & R2	38	3.112 (1.774—5.459)	< 0.001	2.255 (0.911—5.580)	0.078
MAL	553				
Low	277	Reference		Reference	
High	276	2.337 (1.508—3.621)	< 0.001	0.955 (0.474—1.924)	0.899

navitoclax and THR11/SER14 residues in MAL's catalytic domain (Fig. 10B). Notably, the THR11 hydrogen bond exhibited a biologically optimal length (2.2 Å), indicating structural indispensability. Thermodynamic quantification confirmed strong binding affinity ($\Delta G = -7.5$ kcal/mol), and surface topology mapping localized navitoclax in a complementary hydrophobic cleft on MAL (Fig. 10C). The ligand steric conformation demonstrated near-perfect cavity occupancy, a feature conferring binding specificity. To further validate the binding stability of the complex system, molecular dynamics simulations were performed. In the initial molecular dynamics simulation, the equilibrium of the simulated system was evaluated using root-mean-square deviation (RMSD), which revealed that the complex system reached equilibrium after 80 ns and subsequently fluctuated within a range of approximately 8 Å (Fig. 10D), indicating high stability during small molecule-target protein binding. Radius of gyration (Rg) analysis demonstrated slight fluctuations during the dynamic process (Fig. 10E), suggesting conformational changes in the small molecule-protein complex. Solvent-accessible surface area (SASA) calculations

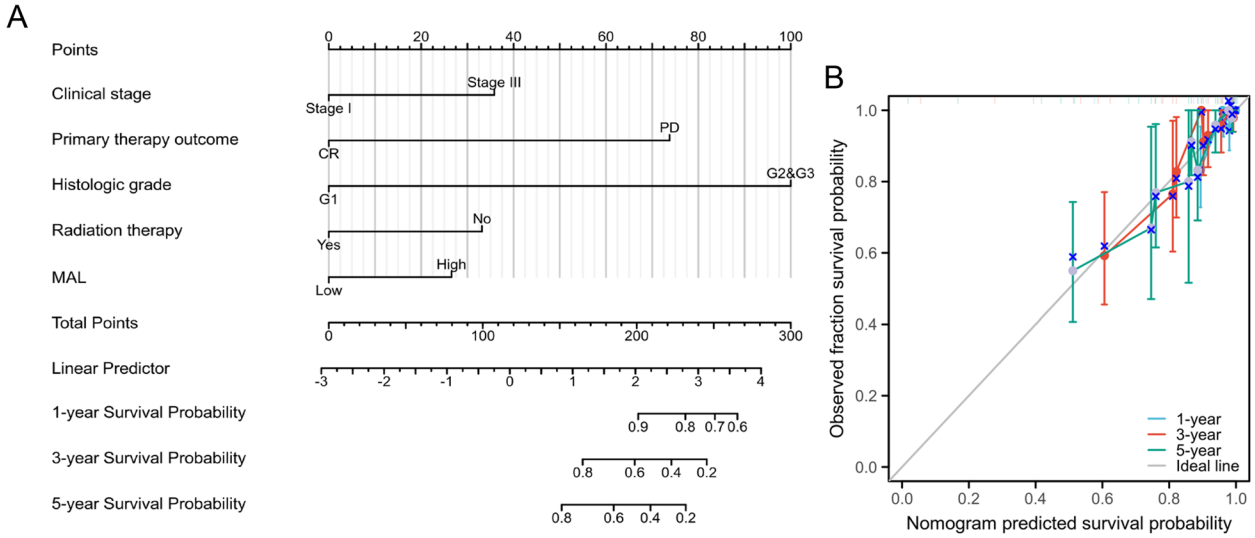

Fig. 7 Prognostic modeling of MAL expression and clinical variables. **A** Prognostic nomogram integrating MAL and independent prognostic factors to predict survival. **B** Calibration plot for nomogram performance validation

(Fig. 10F) further confirmed these conformational alterations through observed minor fluctuations. Hydrogen bond analysis (Fig. 10G) revealed that the number of intermolecular hydrogen bonds ranged from 0 to 6, with an average of 3 bonds maintained throughout most of the simulation, indicating favorable hydrogen bonding interactions between the ligand and target protein. Root-mean-square fluctuation (RMSF) analysis of amino acid residues showed relatively low values (mostly below 3 Å), reflecting limited flexibility and enhanced stability of the complex (Fig. 10H). The second molecular dynamics simulation demonstrated that the root-mean-square deviation (RMSD) analysis indicated the complex system reached equilibrium after 60 ns, with subsequent fluctuations stabilizing around 3.9 Å (Fig. 10I). Radius of gyration (Rg) measurements revealed minor variations throughout the simulation trajectory (Fig. 10J). Solvent-accessible surface area (SASA) evaluation showed no significant alterations in the complex's SASA following ligand-receptor binding (Fig. 10K). Hydrogen bond analysis between the small molecule and target protein (Fig. 10L) demonstrated a dynamic range of 0–7 intermolecular hydrogen bonds, with the complex predominantly maintaining approximately 4 hydrogen bonds. Root-mean-square fluctuation (RMSF) assessment of amino acid residues indicated that most fluctuations remained below 3 Å (Fig. 10M). These results demonstrate that the complex system exhibits stable binding characteristics with robust hydrogen bonding interactions, confirming favorable binding between Navitoclax and MAL. Collectively, our research supporting its role in modulating UCEC progression. Based on drug sensitivity analysis using the CPADS database, samples with high MAL expression exhibited distinct pharmacological responses to Navitoclax. Molecular docking and molecular dynamics simulations further demonstrated that Navitoclax can form a relatively stable binding conformation with MAL, suggesting that Navitoclax may serve as a candidate small molecule for targeting MAL-related pathways. These findings provide structural-level theoretical support for the future exploration of targeted therapeutic strategies.

MAL expression in endometrial carcinoma cells vs. Normal endometrial cells and its effect on cell behavior

Western blotting analysis revealed significantly elevated MAL expression in endometrial carcinoma cells (ECC-1) versus normal endometrial cells (EEC) ($p < 0.05$; Fig. 11A, B). Genetic manipulation via MAL knockdown (Sh-MAL) and overexpression (OE-MAL) confirmed dose-dependent MAL modulation relative to respective controls (NC-Sh-MAL and NC-OE-MAL; $p < 0.01$; Fig. 11C, D). 5-Ethynyl-2'-deoxyuridine (EdU) assays revealed that MAL knockdown significantly inhibited proliferation rates ($p < 0.01$), whereas MAL overexpression significantly enhanced them ($p < 0.01$). Transwell invasion analyses showed that MAL knockdown significantly reduced invasion ($p < 0.01$), whereas MAL overexpression significantly increased it ($p < 0.0001$) (Fig. 11E–H). These results affirm MAL's key role in promoting UCEC cell proliferation and invasion.

Discussion

UCEC remains one of the most prevalent gynecological malignancies worldwide, with incidence rates increasing particularly among high-risk populations [17–19]. Its subtle early-stage presentation often results in delayed diagnosis, contributing to limited treatment efficacy and increased mortality [20–22]. Given the insufficient

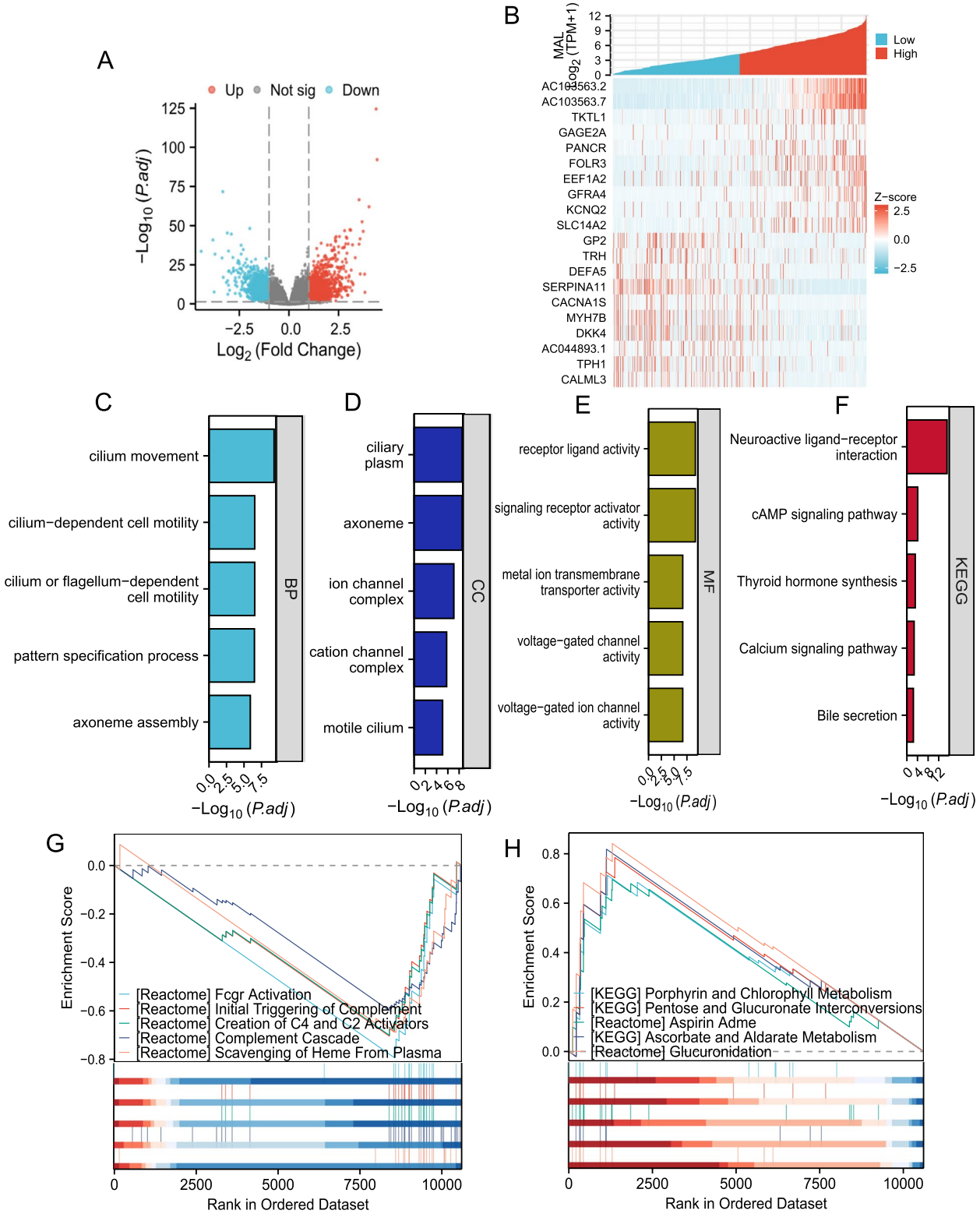

Fig. 8 Potential Mechanism Analysis of MAL. **A** Volcanic diagram of differences between high/low expression groups. **B** Co-expression heatmap between MAL and the top 10 upregulated as well as top 10 downregulated genes. **C**–**F** The top 5 results of GO/KEGG analysis on differentially expressed genes between high and low expression groups. **G** GSEA analysis of the top 5 negatively correlated pathways. **H** GSEA analysis of the top 5 positively correlated pathways

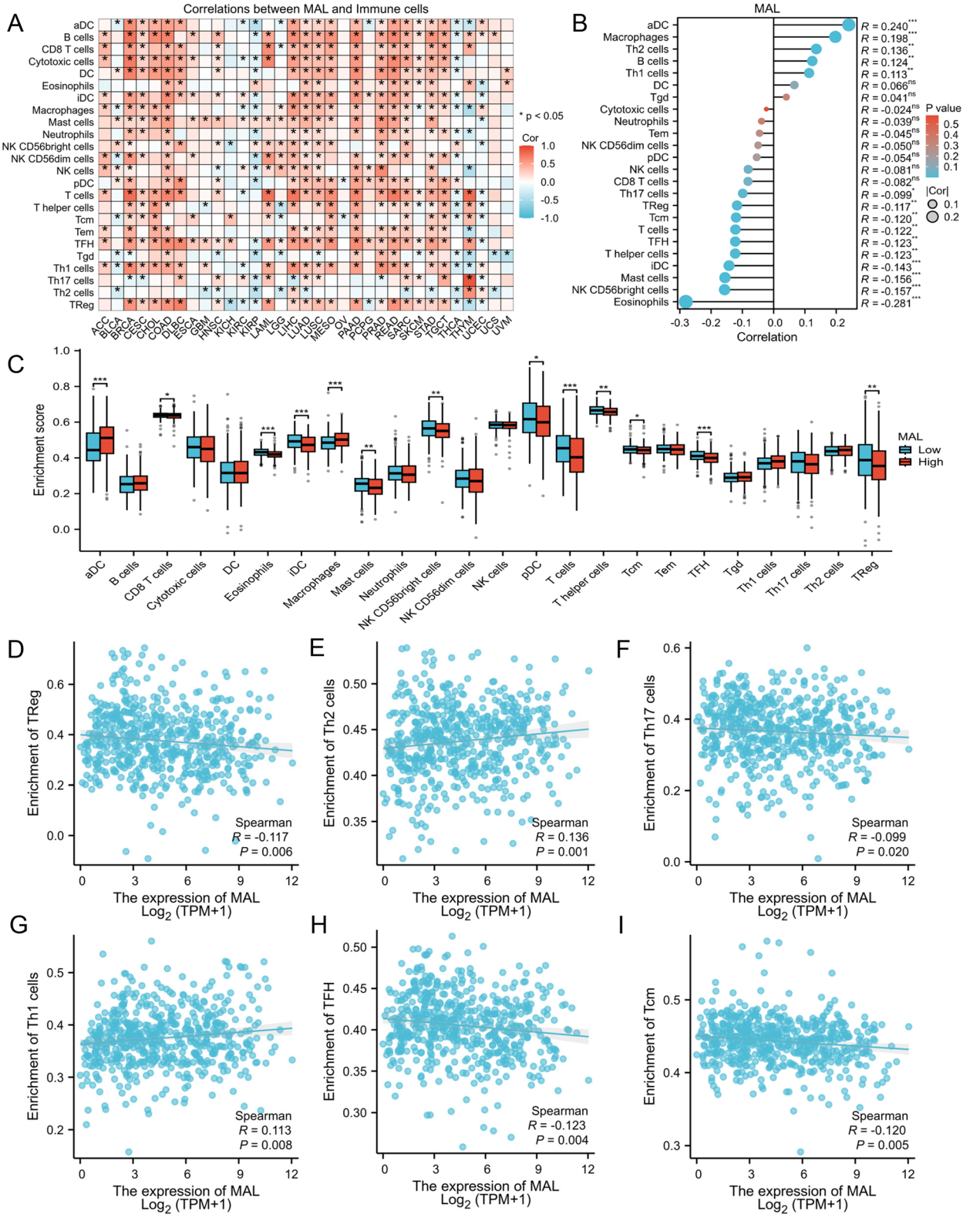

Fig. 9 Correlation between MAL expression and immune characteristics. **A** Heatmap of MAL-regulated immune cells abundance across multiple cancer types. **B** Correlation heatmap between MAL and immune cell infiltrates in UCEC. **C** Immune cell differential analysis based on high vs. low MAL expression. **D**–**I** Scatter plots showing correlations between MAL expression and specific immune subsets: TReg, Th2, Th17, Th1, TFH, and Tcm. *, **, and ***: $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively

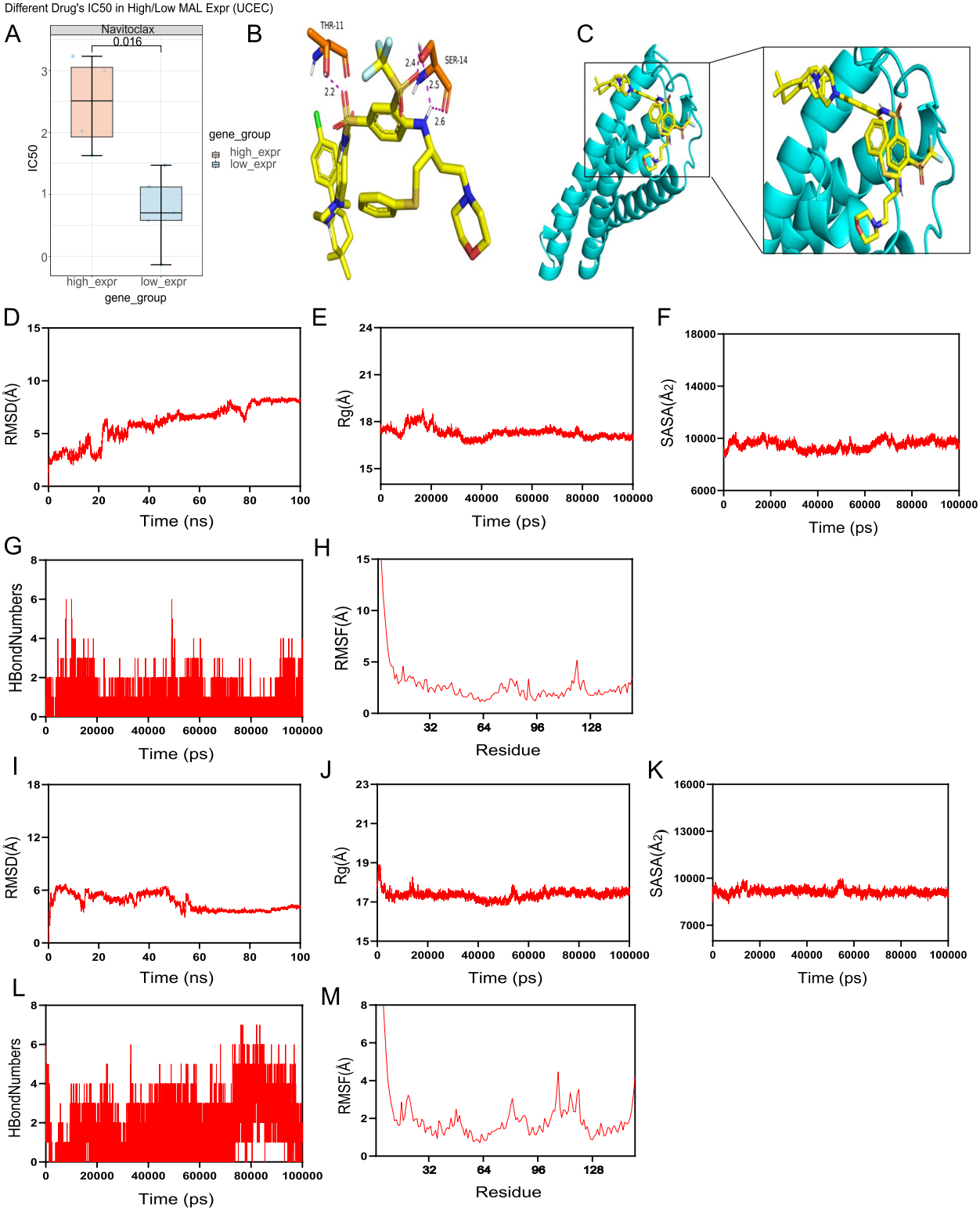

Fig. 10 (See legend on next page.)

(See figure on previous page.)

Fig. 10 Docking Analysis and *Molecular Dynamics analysis* of MAL and Navitoclax. **A** Navitoclax IC_{50} in high/low MAL expression groups. **B** Three-dimensional model of MAL. **C** Structural docking interactions between MAL and navitoclax. **D** Time-dependent root-mean-square deviation (RMSD) values of protein–ligand complexes. **E** Time-dependent radius of gyration (Rg) values of protein–ligand complexes. **F** Time-dependent solvent-accessible surface area (SASA) values of protein–ligand complexes. **G** Time-dependent hydrogen bond (HBond) formation in protein–ligand complexes. **H** Root-mean-square fluctuation (RMSF) values of protein–ligand complexes. **I** Time-dependent root mean square deviation (RMSD) values of the protein–ligand complex after the second molecular dynamics simulation. **J** Time-dependent radius of gyration (Rg) values of the protein–ligand complex after the second molecular dynamics simulation. **K** Time-dependent solvent accessible surface area (SASA) values of the protein–ligand complex. **L** Time-dependent hydrogen bond (HBond) formation in the protein–ligand complex after the second molecular dynamics simulation. **M** Root mean square fluctuation (RMSF) values of the protein–ligand complex after the second molecular dynamics simulation

sensitivity of current diagnostics, such as imaging and histopathology, there is a pressing need for novel biomarkers to improve detection and prognostic precision [23, 24].

MAL, a gene previously implicated in various malignancies, demonstrates significant overexpression in multiple cancer types and is associated with poor patient prognosis [13, 25, 26]. Functionally, it mediates intercellular substance exchange [27]. In this study, we implemented an integrated bioinformatics approach, combining WGCNA and machine learning algorithms to analyze transcriptomic data from the GEO database. This strategy enabled the identification of UCEC-associated biomarkers and facilitated subtype-specific survival stratification. Our findings support a new paradigm in UCEC research that prioritizes early detection and individualized treatment.

Through an integrated approach combining Weighted Gene Co-expression Network Analysis (WGCNA) and machine learning, we successfully identified potential biomarkers for endometrial carcinoma (UCEC), providing novel strategies for early detection and prognosis assessment of the disease. Compared to conventional single-method analyses, this study employed multiple data sources and computational tools, thereby enhancing the reliability of the results and their potential for clinical application.In the screening of differentially expressed genes (DEGs), analysis using the "limma" package identified a total of 167 DEGs, comprising 100 up-regulated and 67 down-regulated genes. These DEGs offer important clues to the pathogenesis of UCEC and may serve as new biomarkers. Further GO/KEGG analyses revealed multiple biological processes and pathways associated with UCEC. The identification of key pathways contributes to understanding the biological characteristics of UCEC and its potential therapeutic targets, providing valuable information for clinical management.WGCNA analysis constructed gene modules related to UCEC, identifying 16 key modules that may be closely linked to the biological features of UCEC, laying a foundation for further mechanistic investigation. In machine learning modeling, six key genes (HPGD, MTCL1, MAL, GABRP, SCGB1D2, MICA) were selected through multiple machine learning algorithms. The identification of these genes provides new evidence for UCEC biomarkers and may improve clinical prognostic evaluation.The MAL gene showed significant expression differences between normal and cancerous tissues, suggesting its potential role as an important oncogenic factor, with expression levels possibly influencing survival and disease progression in UCEC patients. Subsequently, multiple analytical methods were employed to enhance the reliability of the findings, including survival analysis (Kaplan–Meier and Cox regression), prognostic nomogram construction, gene functional enrichment analysis (GO and KEGG), immune infiltration analysis, and exploration of MAL-related mechanisms.

Although direct experimental validation of specific signaling pathways is still lacking, functional enrichment analysis in this study indicated that MAL-related gene sets were significantly enriched in pathways such as ciliary movement, ion channel complexes, and the cAMP signaling pathway. This is partially consistent with previous reports on the role of MAL in maintaining membrane integrity and epithelial polarity. Based on these findings, it can be hypothesized that MAL may promote tumor cell proliferation and invasion by influencing ciliary structure and function, fundamental signal transduction (e.g., cAMP/calcium signaling), and metabolic states (e.g., hormone/bile acid metabolism) of tumor cells.Drug sensitivity analysis combined with molecular docking and molecular dynamics simulations suggested a potentially stable interaction pattern between Navitoclax, a Bcl-2 family inhibitor, and MAL. This finding provides preliminary candidate compound clues for pharmacological intervention in MAL-related signaling pathways. Compared to EEC, MAL was highly expressed in ECC-1 cells. EdU and Transwell assays verified that high MAL expression promotes proliferation and invasion, whereas inhibition of MAL activity suppresses these phenotypes.

The ssGSEA results of this study reveal that MAL expression levels exhibit significant correlations with the infiltration degree of multiple immune cell subsets, suggesting that MAL may be involved in remodeling the immune microenvironment of endometrial cancer. Combined with the GO/KEGG enrichment analysis exploring MAL-associated mechanisms, MAL is predominantly enriched in pathways related to ciliary movement, ion channel complexes, and the cAMP signaling pathway, which provides clues for understanding

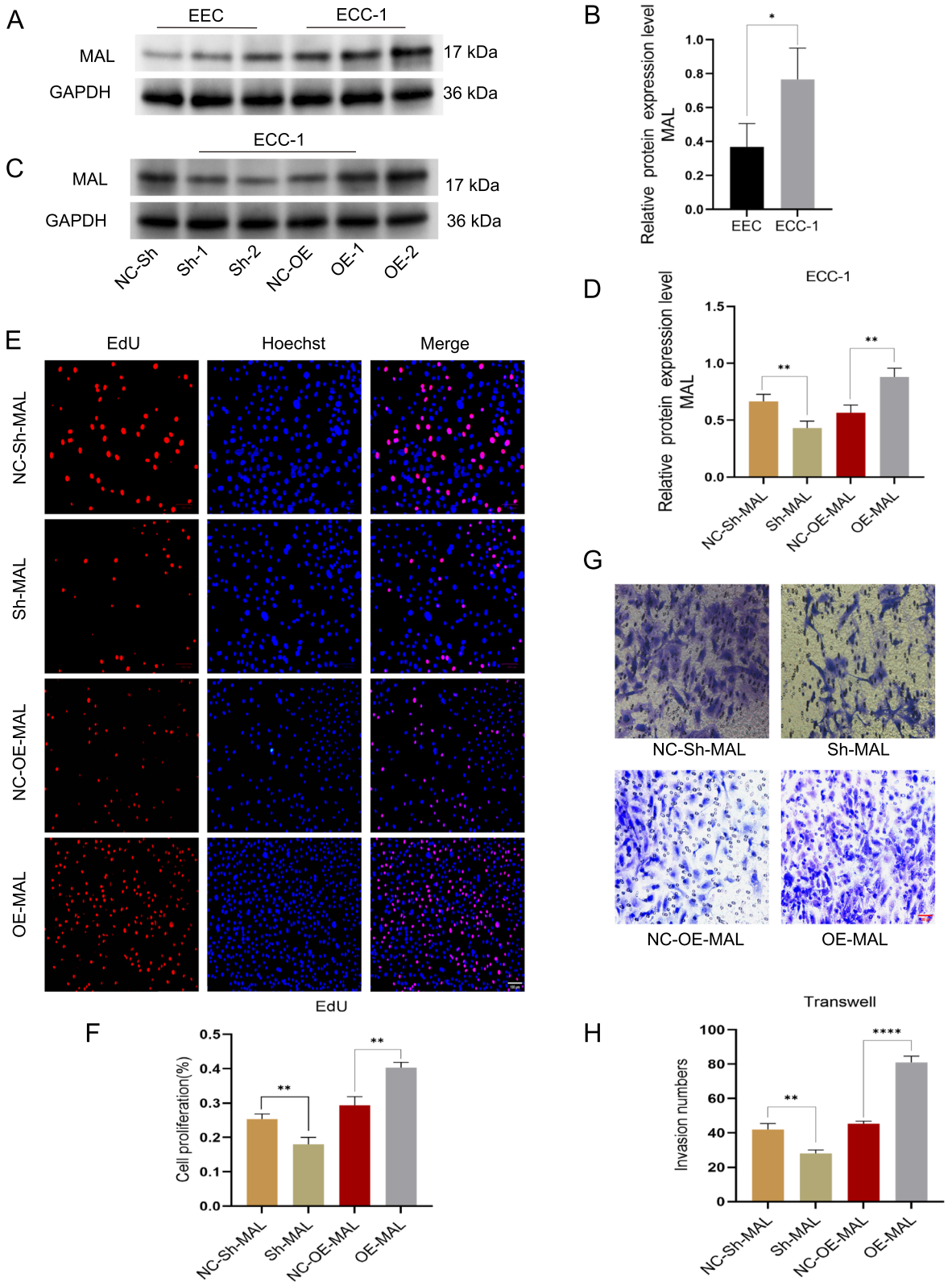

Fig. 11 MAL expression in EEC and ECC-1 cells. **A**, **B** Expression of MAL protein in EEC and ECC-1 cells. **C**, **D** Validation of MAL protein expression knockdown and overexpression in ECC-1 cells. **E**, **F** EdU assay results for ECC-1 cell proliferation after MAL silencing or overexpression. **G**, **H** Transwell invasion assay after MAL silencing or overexpression. *, **, and ***: $p < 0.05$, $p < 0.01$, and $p < 0.0001$, respectively

its immunomodulatory role. From a biological function perspective, it is hypothesized that MAL may alter the malignant phenotype of tumor cells by affecting ciliary structure and function, fundamental signal transduction (e.g., cAMP/calcium signaling), and metabolic states (e.g., hormone/bile acid metabolism). This altered malignant phenotype may subsequently release distinct signaling molecules, shaping an "immunosuppressive," "immunoactive," or "immune-excluded" tumor microenvironment, which ultimately determines the quantity, type, and functional status of immune cell infiltration. This hypothetical model helps contextualize the correlation findings of this study within a more comprehensive immunoregulatory framework and provides direction for future validation at the cellular and molecular levels to elucidate the mechanisms underlying MAL-mediated immune microenvironment reprogramming.

These findings not only provide a critical foundation for the early diagnosis and personalized treatment of UCEC but also establish a basis for future in-depth research in this field.The limitations of this study primarily include the relatively small sample size, which may constrain the generalizability and extrapolation of the results. Furthermore, the reliance on publicly available databases for data analysis could introduce batch-to-batch variability, thereby potentially affecting the accuracy and reliability of gene expression measurements. Although this study employed multiple statistical approaches, machine learning models, and preliminary experimental validation, clinical validation has not yet been conducted. This indicates that the practical utility and effectiveness of the identified biomarkers require further experimental studies for confirmation.It should be emphasized that the evidence regarding the relationship between MAL and Navitoclax in this work mainly originates from drug sensitivity database analyses and computational simulations, such as molecular docking and molecular dynamics. These are predictive and exploratory in nature and currently lack direct support from biochemical assays or cellular-level drug-response experiments. The current findings are more appropriately regarded as theoretical support and candidate clues suggesting a possible interaction and therapeutic potential of the MAL–Navitoclax axis, rather than as conclusive evidence for its causal relationship or clinical feasibility. Future studies are warranted to conduct Navitoclax intervention experiments in endometrial cancer cells and animal models, combined with manipulation of MAL expression, to further validate its effects on tumor biology and the underlying mechanisms.The mechanistic role of MAL in UCEC has not been fully elucidated. Building on this study, future research could employ MAL overexpression or knockdown models to systematically assess the activation status of classical oncogenic pathways, such as PI3K/AKT and Wnt/β-catenin. By examining downstream target gene expression profiles and phosphorylation states of key proteins, the MAL-mediated signal transduction network and its causal role in endometrial cancer progression could be clarified, thereby refining the proposed framework of molecular markers and potential therapeutic targets established here. Therefore, future investigations should consider expanding the sample size and integrating clinical data with fundamental experiments to enhance the credibility and translational relevance of the findings.

In conclusion, this study identifies MAL as a pivotal biomarker for UCEC, providing new avenues for early detection, prognostic assessment, and individualized treatment. Although further validation is warranted, our findings establish a strong foundation for improving clinical outcomes and advancing personalized care in endometrial carcinoma.

Abbreviations

AUC	Area under the ROC curve
DEG	Differentially expressed genes
DMEM	Dulbecco's modified Eagle medium
DOI	Digital object identifier
DT	Decision tree
GEO	Gene Expression Omnibus
GLM	Generalized linear model
GO	Gene ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
KNN	K-Nearest neighbors
LASSO	Least absolute shrinkage and selection operator
MAL	Myelin and lymphocyte
NK	Natural killer
OS	Overall survival
PFI	Progression-free interval
PFS	Progression-free survival
RF	Random forest
ROC	Receiver operating characteristic
SD	Stable disease
SVM	Support vector machine
TCGA	The Cancer Genome Atlas
TFH	T follicular helper
WGCNA	Weighted gene coexpression network analysis

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-026-15573-7.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

We express gratitude to our laboratory members for their technical assistance and engaging discussions.

Authors' contributions

Hui Wang and Yenwen Gong: methodology, software, formal analysis. Xiu Han and Jiawei Zeng: wrote original draft.

Funding

This research received no external funding.

Data availability
The datasets analyzed in this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors have read and agreed to the published version of the manuscript.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Gynaecology, Guangyuan Central Hospital, Jingxiangzi, No. 16, Guangyuan Sichuan 628000, China
[2]Department of Clinical Laboratory Diagnostics, The Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan 646000, China
[3]Department of Pediatrics, Guangyuan Central Hospital, Jingxiangzi, No. 16, Guangyuan Sichuan 628000, China
[4]Department of Laboratory, Mianyang Central Hospital, Sichuan 621000, China

Received: 21 November 2025 / Accepted: 7 January 2026
Published online: 21 January 2026

References

1. Lai Z, Li C. TRIP13 is a potential prognostic marker and therapeutic target for endometrial cancer. Crit Rev Eukaryot Gene Expr. 2025;35(3):23–41.
2. He K, Li J, Huang X, et al. KNL1 is a prognostic and diagnostic biomarker related to immune infiltration in patients with uterine corpus endometrial carcinoma. Front Oncol. 2023;13:1090779.
3. Wang Y, Ren F, Song Z, Wang X, Ma X. Multiomics profile and prognostic gene signature of m6A regulators in uterine corpus endometrial carcinoma. J Cancer. 2020;11(21):6390–401.
4. Han Y, Huang Y, Luo D. An effective and validated prognostic modelfor uterine corpus endometrial cancer based on gene main effects and gene-gene interactions. Medicine Baltimore. 2025;104(23):e42583.
5. Djajadiningrat RS, Bergman AM, van Werkhoven E, Vegt E, Horenblas S. Neoadjuvant taxane-based combination chemotherapy in patients with advanced penile cancer. Clin Genitourin Cancer. 2015;13(1):44–9.
6. Unger KR, Romney DA, Koc M, et al. Preoperative chemoradiation for rectal cancer using capecitabine and celecoxib correlated with posttreatment assessment of thymidylate synthase and thymidine phosphorylase expression. Int J Radiat Oncol Biol Phys. 2011;80(5):1377–82.
7. Poulsen M, Aldren C, Tripcony L, Walker Q. Is surgery necessary in stage III and stage IV cancer of the head and neck that responds to induction chemotherapy? Arch Otolaryngol Head Neck Surg. 1996;122(5):467–71.
8. Yuan Y, Chen Z, Cai X, He S, Li D, Zhao W. Identification of hub genes correlated with poor prognosis for patients with uterine corpus endometrial carcinoma by integrated bioinformatics analysis and experimental validation. Front Oncol. 2021;11:766947.
9. Luo G, Bo C, Li J. Identification and validation of hub genes in uterine corpus endometrioid carcinoma: an observational study from TCGA and GEO. Medicine Baltimore. 2025;104(18):e42338.
10. Frank M. MAL, a proteolipid in glycosphingolipid enriched domains: functional implications in myelin and beyond. Prog Neurobiol. 2000;60(6):531–44.
11. Rubio-Ramos A, Labat-de-Hoz L, Correas I, Alonso MA. The MAL protein, an integral component of specialized membranes, in normal cells and cancer. Cells. 2021. https://doi.org/10.3390/cells10051065.
12. Frank M, Schaeren-Wiemers N, Schneider R, Schwab ME. Developmental expression pattern of the myelin proteolipid MAL indicates different functions of MAL for immature Schwann cells and in a late step of CNS myelinogenesis. J Neurochem. 1999;73(2):587–97.
13. Lara-Lemus R. On the role of myelin and lymphocyte protein (MAL) in cancer: a puzzle with two faces. J Cancer. 2019;10(10):2312–8.
14. Xin X, Zhou H, Huang S, et al. Identification of biomarkers and potential drug targets in DFU based on fundamental experiments and multi-omics joint analysis. Front Pharmacol. 2025;16:1561179.
15. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem. 2008;29:1859–65.
16. Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/Ewater models at 298 K. J Phys Chem A. 2001;105:9954–60.
17. Ye Y, Li H, Bian J, Wang L, Wang Y, Huang H. Exploring prognosis-associated biomarkers of estrogen-independent uterine corpus endometrial carcinoma by bioinformatics analysis. Int J Gen Med. 2021;14:9067–9081.
18. Wang L, Wang M, Wang Z, et al. UBE2T is a diagnostic and prognostic biombarker for endometrial cancer. Clin Transl Oncol. 2025;27(5):2067–83.
19. Wang X, Fang A, Peng Y, et al. *PHF6* promotes the progression of endometrial carcinoma by increasing cancer cells growth and decreasing T-cell infiltration. J Cell Mol Med. 2023;27(5):609–21.
20. Artioli G, Cassaro M, Pedrini L, et al. Rare presentation of endometrial carcinoma with singular bone metastasis. Eur J Cancer Care (Engl). 2010;19(5):694–8.
21. Amant F, Moerman P, Neven P, Timmerman D, Van Limbergen E, Vergote I. Endometrial cancer. Lancet. 2005;366(9484):491–505.
22. Franceschi S, La Vecchia C, Gallus G, et al. Delayed diagnosis of endometrial cancer in Italy. Cancer. 1983;51(6):1176–8.
23. Chen L, Chen Y, Feng YL, et al. Tumor circulome in the liquid biopsies for digestive tract cancer diagnosis and prognosis. World J Clin Cases. 2020;8(11):2066–80.
24. Tuncer SB, Erdogan OS, Erciyas SK, et al. MiRNA expression profile changes in the peripheral blood of monozygotic discordant twins for epithelial ovarian carcinoma: potential new biomarkers for early diagnosis and prognosis of ovarian carcinoma. J Ovarian Res. 2020;13(1):99.
25. Li D, Zhang J, Wu L, Yang X, Chen Z, Yuan J. Myelin and lymphocyte protein (MAL): a novel biomarker for uterine corpus endometrial carcinoma. Cancer Manag Res. 2021;13:7311–7323.
26. Kurashige J, Sawada G, Takahashi Y, et al. Suppression of MAL gene expression in gastric cancer correlates with metastasis and mortality. Fukuoka Igaku Zasshi. 2013;104(10):344–9.
27. Chen J, Li Q, Liu X, et al. Potential biomarkers and immune infiltration linking endometriosis with recurrent pregnancy loss based on bioinformatics and machine learning. Front Mol Biosci. 2025;12:1529507.

Publisher's Note