



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

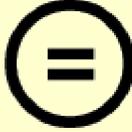
다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

**Unveiling the Prognostic Significance of Protein
Expression in Advanced High-grade Serous
Ovarian Cancer: A Comparative Study between
Long-Term Survivors and Early Mortal Patients**

Ji-Won Ryu

The Graduate School
Yonsei University
Department of Medicine

Unveiling the Prognostic Significance of Protein Expression in Advanced High-grade Serous Ovarian Cancer: A Comparative Study between Long-Term Survivors and Early Mortal Patients

A Dissertation Submitted
to the Department of Medicine
and the Graduate School of Yonsei University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Medical Science

Ji-Won Ryu

June 2024

This certifies that the Dissertation of Ji-Won Ryu is approved.

Thesis Supervisor Jae-Hoon Kim

Thesis Committee Member Hang-Seok Chang

Thesis Committee Member Joon Jeong

Thesis Committee Member Jae Yun Song

Thesis Committee Member Sung Jong Lee

**The Graduate School
Yonsei University
June 2024**

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Prof. Jaehoon Kim, a medical doctor and researcher, for his encouragement and direction despite my many shortcomings. I would also like to thank Prof. Hang Seok Jang, Prof. Joon Jeong, Prof. Jae Yun Sung, and Prof. Sung Jong Lee for their generous advice and support until the thesis was completed.

I would also like to thank Dr. Joon-Yong Jung for his great help and encouragement from far away in the United States when I was very confused at the beginning of my degree.

Dr. Jue Young Kim and Dr. Ha Yeon Shin, who worked together in the lab and were shining light for me during the difficult process of my degree, helped me a lot from the beginning to the end of my thesis, so thank you again.

My love, Mr. Il Ho Ryu, and my soul mate, Mrs. Cheong Sook Jang, who have given me the courage and the reason to get back on my feet when I am down. Mrs. Guiheon Ahn, who has taken care of the daughter-in-law, who is studying again at a late age, with more love than a daughter, and Mr. Han Young Shin, who has supported my passion and dreams.

I would like to thank my loving husband, Shin Dong Hyun, and my two daughters, Shin Jae In and Shin Hae In, for being a wonderful mom and woman.

TABLE OF CONTENTS

LIST OF FIGURES	iii
LIST OF TABLES	iv
ABSTRACT IN ENGLISH	v
1. INTRODUCTION	1
2. MATERIALS AND METHODS	2
2.1. Proteomic analysis	2
2.1.1. Selection of frozen tissue	2
2.1.2. Group curation	2
2.1.3. Proximity extension assay (PEA)	2
2.1.3.1. Protein extraction from fresh frozen tissue	2
2.1.3.2. Protein measurements	3
2.1.4 Data analysis and statistical	3
2.1.4.1. Machine learning based analysis	3
2.1.4.2. Statistical analysis	5
2.2. In vitro function assay	5
2.2.1. Cell culture	5
2.2.2. siTransfection	5
2.2.3. RNA extraction and quantitative real-time PCR	6
2.2.4. Cell proliferation assay	6
2.2.5. Boyden chamber invasion assay	6
2.2.6. Transwell migration assay	7
2.2.7. Western blot	7
2.3. Validation via Immunohistochemistry	7
2.3.1. Tissue microarray (TMA) construction	7
2.3.1.1. Clinical specimens	7
2.3.1.2. TMA	8
2.3.2. Immunohistochemistry (IHC)	8
2.3.2.1. IHC	8
2.3.2.2. Evaluation of immunohistochemical staining	8
3. RESULTS	9
3.1. Protein expression profiling via PEA	9

3.2. Key protein selection based on machine learning	12
3.2.1. Selection proteins with different features-proteins with low deviation	12
3.2.2. k-means clustering using the elbow method	12
3.2.3. Visualization model construction	12
3.2.4. NPTN and PPM1A are selected	12
3.2.5. Evaluation of the proposed scheme	12
3.3. Knock-down NPTN and PPM1A reduced cell proliferation, migration, and invasion ·	13
3.4. NPTN and PPM1A in IHC	17
3.4.1. NPTN, PPM1A expressed high in serous ovarian cancer in IHC	17
3.4.2. High expression on NPTN and PPM1A are associated with poor prognosis	18
4. DISCUSSION	21
5. CONCLUSION	23
REFERENCES	24
ABSTRACT IN KOREAN	29

LIST OF FIGURES

<Figure 1> System model for analyzing the relationship between patient prognosis and proteins	4
<Figure 2> PEA results	11
<Figure 3> Adaptive model construction based on machine learning	14
<Figure 4> Validation by in-vitro analysis	16
<Figure 5> Clinical validation using immunohistochemistry	19

LIST OF TABLES

<Table 1> Characteristics of patients.....	10
<Table 2> Performance analysis using heuristic techniques	13
<Table 3> Association of NPTN and PPM1A expression with patient clinical characteristics.....	18

ABSTRACT

Unveiling the Prognostic Significance of Protein Expression in Advanced High-grade Serous Ovarian Cancer: A Comparative Study between Long-Term Survivors and Early Mortal Patients

High-grade serous ovarian cancer (HGSOC), despite its high lethality, lacks reliable biomarkers for predicting poor prognosis, and limited progression has been made in personalized treatment. Targeted therapy based on genetic profile has not met expectations, as genomic alterations alone do not exclusively determine cancer cell phenotypes. Protein expression critically influences cellular processes. Recognizing proteomic alterations is even more crucial. This study proposes a novel technique, utilizing statistical deviation and machine-learning to select protein factors determining ovarian cancer prognosis.

In advanced HGSOC, divided into two groups with very good (n=24) and poor prognoses (n=23), proteins were extracted from fresh frozen tissue and subjected to proximity extension assay (PEA). We explored a novel approach called artificial intelligence (AI)-based machine learning to identify key proteins that could distinguish between groups with good and poor prognoses. By developing a model, we found that high levels of NPTN and PPM1A indicated a poor prognosis group, demonstrating remarkably high efficacy (Precision 0.857, Recall 0.818, F1-score 0.893). After IHC of NPTN and PPM1A in a tissue microarray (TMA), survival analysis showed that survival decreased when the expression was high. *In vitro* experiments with NPTN and PPM1A knockdown showed reduced cell proliferation, migration, and invasion. Our results suggest that it is feasible to select factors with significant differences between prognostic groups, particularly those that are amenable to clustering based on identified proteins. The research highlights the potential of proteomic markers to guide personalized therapeutic strategies to improve patient outcomes.

Keywords: high-grade serous ovarian cancer, machine-learning, prognosis, proteomics, proximity extension assay

1. INTRODUCTION

Cancer is the most prevalent genetic disease resulting from genetic mutations or chromosomal disorders^{1,2}. The genomic instability in cancer cells causes abnormalities in cellular processes, which leads to unrestrained cancer cell growth³. Scientists have therefore sought to discover biomarkers for cancer using genome-based approaches such as next-generation sequencing technologies⁴. Large-scale DNA sequencing efforts have led to advances in the development of targeted therapies based on the genomic profile of cancers⁵. Nevertheless, targeted therapy based on genomics has not been as successful as expected, largely due to the emergence of drug resistance⁶. Genomic variation is not the only determinant of cancer cell phenotype. The intricate process of alternative splicing enables genes to generate diverse proteins with different sequences and functions. In addition, post-translational modifications and interactions with other proteins contribute to the formation of mature proteins and complexes that play critical roles in various cellular activities. As a downstream consequence of gene regulation, alterations in protein expression and activity play a key role in shaping the phenotype of cancer cells². Therefore, in the realm of targeted therapy, the quantification of proteomic alterations is crucial. Epithelial ovarian cancer (EOC) represents the majority, accounting for 85-90% of all ovarian cancers. EOC is further classified into five subtypes: serous (the most common subtype with 70% of cases), endometrioid, clear cell, mucinous, Brenner, and undifferentiated tumors⁷⁻⁹. There is also a dualistic model of EOC classification based on clinical and genetic profiles. Type I tumors are low-grade serous, endometrioid, clear cell, mucinous, and Brenner tumors. They are commonly associated with somatic mutations in *BRAF*, *KRAS*, *PIK3CA*, and *PTEN*, and are generally indolent and poorly responsive to chemotherapy^{10,11}. In comparison, type II tumors, which include high-grade serous, high-grade endometrioid, carcinosarcoma, and undifferentiated carcinoma, are clinically aggressive with late presentation^{12,13}. Genomic instability and mutations in *TP53* (96%) and *BRCA1/2* (22%) are common features of HGSOC, which is the most common type of EOC, the leading cause of death in EOC, and the most extensively studied to date¹⁴. Patients with HGSOC have the worst outcome, with a 5-year survival rate of less than 40%^{15,16}. Proteomic markers that can predict the progression and prognosis of HGSOC may become novel therapeutic targets for HGSOC and play an important role in personalized treatment. Therefore, we aimed to identify key proteins that can predict prognosis by proteomic analysis of advanced HGSOC.

2. MATERIALS AND METHODS

2.1. Proteomic analysis

2.1.1. Selection of frozen tissue

We used fresh-frozen primary ovarian cancer specimens obtained intraoperatively and stored in the Korea Gynecologic Cancer Bank (KGCB) for study purposes. We identified patients who fulfilled the following inclusion criteria: (1) 18 years of age or older, (2) diagnosed with HGSOC between January 2007 and January 2022, and (3) advanced stage. The International Federation of Gynecology and Obstetrics (FIGO) stage III is more than, and (4) willing to donate biospecimens and provide written informed consent. We excluded patients with recurrent ovarian cancer, patients who received chemotherapy including neoadjuvant chemotherapy (NAC), patients with insufficient clinical data or lost to follow-up, and patients with severe comorbidities.

2.1.2. Group curation

We identified a good prognosis group and a poor prognosis group in the patient population. We defined the good prognosis group as follows. 1] Relapse after 2 years and no further progression for the next 5 years 2] No progression for 5 years after the first treatment. The poor prognosis group was defined as 1] Expired within 2 years of starting treatment, 2] Relapsed within 1 year of starting treatment and expired within 3 years, and 3] Refractory and resistant and expired within 3 years. The good prognosis group had a total of 24 organizations and the poor prognosis group had a total of 23 organizations. The normal tissue was obtained from the fallopian tubes of 6 patients undergoing surgery for benign conditions.

2.1.3. PEA (Proximity extension assay)

2.1.3.1. Protein extraction from fresh frozen tissue

The tissue, freshly collected and promptly snap-frozen on dry ice, was carefully excised with a scalpel while maintaining the frozen state. Subsequently, the tissue sample was transferred into a beads tube placed on ice. The homogenizer tube was filled with lysis T-PERTM buffer comprising 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, and 0.1% Sodium deoxycholate, supplemented with protease inhibitors. The sample was homogenized using a FastPrep-24TM classic homogenizer for 20 seconds and then placed on ice for 5 minutes. This homogenization and centrifugation process was repeated more than four times. Subsequently, the supernatant was collected, and centrifugation was performed at 14,000 rpm for 10 minutes at 4°C to

achieve centrifugal separation. Protein concentration was determined using bovine serum albumin (BSA), and samples were adjusted to a consistent concentration accordingly ($0.5\mu\text{g}/\mu\text{L}$). Subsequently, the samples underwent analysis using the Olink Explore Oncology II at the DNALink Lab in Seoul, Korea.

2.1.3.2. Protein measurements

Olink's PEA Explore technology employs antibody pairs with DNA oligonucleotide reporter molecules, binding to targets in proximity to generate double-stranded DNA amplicons. We used Olink's Oncology II panel, which consists of 386 antibodies that are associated with oncology. Following the initial immune reaction, amplicons were extended and amplified, incorporating individual sample index sequences in a subsequent step. The prepared libraries were pooled and sequenced using an Illumina NovaSeq 6000 instrument for analysis. Raw BCL files were converted into count files and further processed into normalized protein expression (NPX) values using a manufacturer-specified quality control and normalization method, wherein NPX values on a \log_2 scale reflect protein concentration doubling per unit increase. These NPX values underwent initial manufacturer-performed QC and were subsequently preprocessed using Olink Wizard for GENEX software at the Clinical Biomarker Facility for subsequent analyses¹⁷.

2.1.4. Data analysis and statistical

2.1.4.1. Machine learning based analysis

We aim to analyze the relationship between ovarian cancer prognosis in patients and associated protein factors that have been collected in the field. We prioritized proper data pre-processing to achieve clean data for analysis (Fig. 1 A). We divided the patients into distinct groups and applied a normalization and visualization process to check for the presence of outliers in the dataset (Fig.1 B (a)). Consequently, we identified these data as outliers and removed them to achieve a clean dataset, as illustrated in Fig. 1 B (b). Following the pre-processing scheme, descriptive statistics are used to describe the clinicopathological characteristics of the study population. In the clustering and trend analysis parts, we check whether the distribution of patients according to protein factor forms a certain clustering or tendency. The k-means clustering technique is used as a method to analyze the clustering of such data. The relationship between proteins and groups is also analyzed using regression techniques. Based on these methods, we select the most important factors that could identify the patient's prognosis.

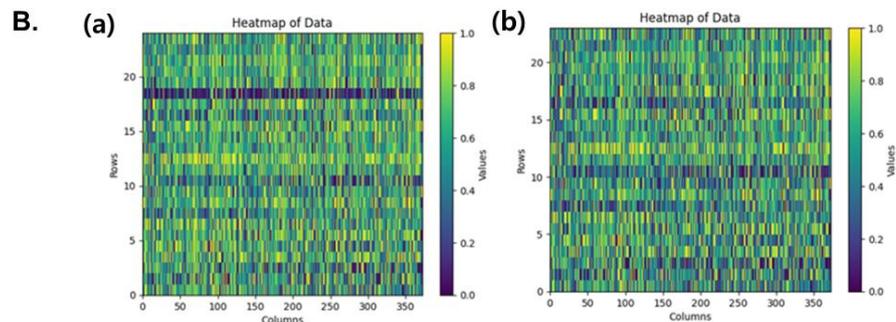
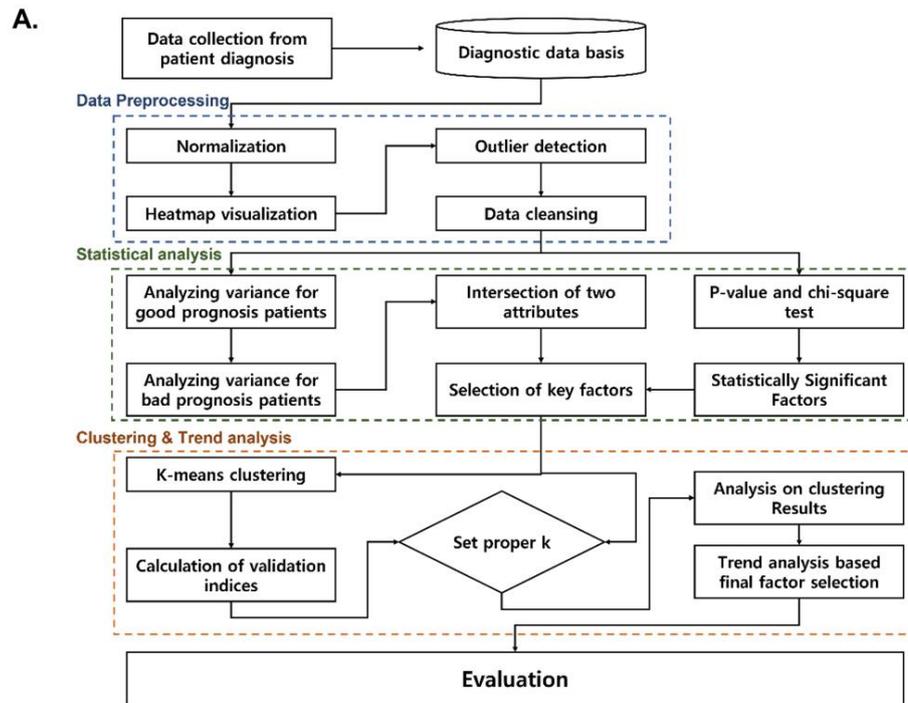


Figure 1. System model for analyzing the relationship between patient prognosis and proteins

A. Schematic for identifying candidate proteins. B. (a) Visualization of data without removing outliers. (b) Visualization of data with removing outliers. We used visualization to remove outliers, as their presence greatly affects the accuracy of data analysis.

2.1.4.2. Statistical analysis

Student's t-test, Mann-Whitney U test, Pearson's chi-square test, and Fisher's exact test were used for statistical analysis of patient clinical characteristics and proteomic results. The Mann-Whitney U test was used for IHC scoring data. Analysis of Kaplan-Meier survival curves was performed and statistical significance was calculated using the log-rank test. These analyses were performed using SPSS software (version 27.0, SPSS Inc., Chicago, IL, USA), R, and GraphPad Prism 10.0 software. All statistical tests were two-tailed and considered statistically significant if the *p-value* was less than 0.05.

2.2. In vitro functional assay

2.2.1. Cell Culture

Five immortalized human ovarian surface epithelial cells were used, which were described previously¹⁸. These cells were grown in DMEM containing 10% FBS with 1% penicillin/streptomycin and cultured at 37°C in 5% CO₂. A total of six serous-type ovarian cancer cell lines were used, all of which were obtained from the Korean cell line bank (Republic of Korea). SKOV3, PE01, and OVCAR3 cells were cultured at 37°C in 5% CO₂ containing RPMI 1640 (Corning; 10-041-CV, NY, USA), 10% fetal bovine serum (FBS; Gibco, #12483-020, Thermo Fisher Scientific Inc., NY, USA) and 1% penicillin/streptomycin (Gibco, #15140, Thermo Fisher Scientific Inc., NY, USA). CAOV3, OVCA429, and SNU119 cells were grown in Dulbecco's modified Eagles medium (DMEM) (Corning; 10-013-CV, NY, USA) containing 10% FBS with 50 U/mL penicillin and 50 µg/mL streptomycin. The cultured cells were sustained in a humidified, 5% CO₂ atmosphere at 37°C. Regular authentication of all cell lines was performed by monitoring their morphology.

2.2.2. siRNA Transfection

NPTN (#27020-1), PPM1A (#5494-1), and negative control (#SN-1003) knockdown were conducted using predesigned siRNA sequences purchased from Bioneer (Daejeon, Republic of Korea). Cells were seeded one day before siRNA transfection in 60 mm dish and were allowed to grow up to 60~70% confluency. On the day of siRNA transfection, cells were transfected with siRNA in Opti-MEM (Gibco, #31985-070, Thermo Fisher Scientific Inc., NY, USA), using Lipofectamine RNAiMax reagent (Invitrogen, #13778-075, Thermo Fisher Scientific Inc., NY, USA) according to the manufacturer's instructions. Transfection efficiencies were analyzed 48 h post-

transfection.

2.2.3. RNA extraction and quantitative real-time PCR

Total RNA was obtained using the RNeasy plus mini kit (Qiagen, Hilden, Germany), and cDNA was synthesized using the C1000 Thermal Cycler (BIO-RAD, Hercules, California, USA) according to the manufacturer's instructions. The beta-actin was used as an internal control. Quantitative PCR was performed and analyzed using the StepOnePlus real-time PCR system (Thermo Fisher Scientific, NY, USA). The primers used for real-time PCR were as follows: for NPTN, forward 5'-GCTCCTAAAGCAAACGCCACCA-3' and reverse 5'-TCTT GCGCCATATCCAGTCTGG-3'; for ADAMTS15, forward 5'-TCTTGCGCCATA TCCAGTCTGG-3' and reverse 5'-GGGAGCCAAG AACTGAGCAT-3'; for PPM1A, forward 5'-ATGCACCCAAAGTATCGCCA-3' and reverse 5'-GGCTGGGGATGTTCTC ACTC-3'; for RANGAP1, forward 5'-GGTGGTCTCTGCCTTCCTAA AG-3' and reverse 5'-CGAGGAGTTGAAAGCCTTCTGC-3'; for METAP2 5'-TGGTGTATTTCC CAAAGGACA-3' and reverse 5'-TGTCATCCCAGGCTTGA TCC-3'; for β -actin, forward 5'-CTCGCCTTTGCCGATCC-3' and reverse 5'-GGGGTACTTCA GGGTGAGG A-3'.

2.2.4. Cell proliferation assay

Cells stably transfected with either siRNA or negative control were seeded into 6-well plates at a density of 1×10^4 cells per well and cultured for 7 days. Subsequently, the cells were fixed using a solution consisting of 10% acetic acid and 10% methanol for 10 min, followed by staining with 0.5% crystal violet for 1 hour. After staining, photographs were taken, and the cells were extracted using a 2% SDS solution. The crystal violet extracted from the cells was then quantified by measuring the absorbance at 595 nm using the VERSA Max™ spectrophotometer. Additionally, the number of colonies containing more than 50 cells was enumerated using a digital microscope using a ZEN3.3 blue edition (Carl Zeiss, Thornwood, NY, USA). All experiments were performed in triplicate.

2.2.5. Boyden chamber invasion assay

To study cell invasion, 48-well micro chemotaxis chambers (Neuro Probe, Gaithersburg, MD, USA) were used. Culture medium containing 10% FBS was applied to the lower chambers, which were then coated with Matrigel (BD Biosciences, San Jose, CA, USA) coated membranes with 8 μ m pores. At 48 hours after transfection, siRNA-transfected cells (1×10^5 cells/56 μ l medium without FBS) were added to upper chambers. The membranes were fixed and stained with a Diff-quick solution (Sysmex, Kobe, Japan) after 24 hours. Non-invading cells were scraped from the top of the membrane and invading cells were counted under a digital microscope using a ZEN3.3 blue edition (Carl Zeiss,

Thornwood, NY, USA). Each experiment was repeated three times.

2.2.6. Transwell migration assay

SKOV3 cells transfected with 100 nM, 200 nM of NPTN and PPM1A, and negative control was seeded at 2×10^4 cells/200 μ l in FBS-free media in the upper chamber, and 500 μ l of media containing 10% FBS in the lower chamber. The chambers were incubated at 37°C for 48 h. At the end of incubation, cells on the upper side of the membrane were wiped off with a cotton swab. Cells under the membrane were fixed with 4% PFS and stained with 0.05% crystal violet. Images were taken on a ZEN3.3 microscope and analyzed using NIH Image J software.

2.2.7. Western blot

For cell lysate preparation, cells were washed with ice-cold phosphate-buffered saline (PBS) and lysed in ice-cold RIPA buffer (Thermo Fisher Scientific, NY, USA) with the addition of protease inhibitor (Sigma) and phosphatase inhibitor (Sigma). Cell lysate protein quantification was performed using BCA reagent (Thermo Fisher Scientific).

Briefly, 30 μ g of protein sample was mixed with an equal volume of $2 \times$ Laemmli buffer and boiled at 95°C for 10 minutes. Lysates were separated by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Sample volumes containing 10 μ g of protein were separated by 10% SDS-PAGE and transferred to PVDF membranes (Millipore, #IPVH00010, Sigma-Aldrich, St. Louis, Missouri, USA). After blocking non-specific binding with 5% non-fat milk for 1 hour at 25°C, the membranes were incubated overnight at 4°C with primary antibodies: anti-PPM1A (CST3549S, Cell Signaling, rabbit monoclonal Ab. 1:200) and anti-NPTN (TA504319, Origene, mouse monoclonal Ab. 1:100). After washing with TBST (Tris Buffered Saline with 0.1% Tween-20, #9997s; Cell Signaling Technology, Inc., Danvers, MA, USA), the membranes were incubated with HRP and developed with enhanced. Optical densities were analyzed using Amersham™ ImageQuant™ 800 (Cytiva, Marlborough, MA, USA). β -Actin was used as a loading control.

2.3. Validation via Immunohistochemistry (IHC)

2.3.1. Tissue Microarray (TMA) construction

2.3.1.1. Clinical specimens

We obtained the specimens from patients who underwent primary surgery at the Gangnam Severance Hospital between 2004 and 2021, and some of the samples were obtained from the KGCB

as part of the Bio and Medical. FIGO classification was applied for tumor staging, and clinical information, including surgical procedure, survival time, survival status, and age, was collected by reviewing the patient's medical records. Patients' response to therapy was determined by computed tomography using the Response Evaluation Criteria in Solid Tumors (RECIST; version 1.1). Tumor grade and cell type were assessed by reviewing pathology reports, and two gynecological pathologists histologically examined all tumor samples. All biological samples were collected after informed consent was provided by the participants.

2.3.1.2. TMA

Formalin-fixed paraffin-embedded (FFPE) tumor tissue blocks serve as the basis for generating TMA. Expert evaluation by an institutional pathologist of full-face H&E-stained sections identifies and defines representative tumor regions. Subsequently, these identified areas are used to construct TMA blocks. The TMA blocks were cut to 5- μ m thickness with a rotary microtome and comprised 1 mm tissue cores, ensuring a significant proportion of tumor cells.

2.3.2. Immunohistochemistry (IHC)

2.3.2.1. IHC

IHC was carried out using TMA. Sections were deparaffinized and rehydrated through graded ethanol. The sections were exposed to 3% H₂O₂ solution in methanol for 30 minutes to quench endogenous peroxidase activity. Heat-induced antigen retrieval was then conducted by incubating the sections in target retrieval buffer at pH 6.0 (Dako, Carpinteria, CA, USA) using a steam pressure cooker (Pascal; Dako) for 20 min, and the slides were stained with anti-PPMIA (CST3549S, cell signaling, rabbit monoclonal Ab 1:200), anti-NPTN (TA 504319, Origene, mouse monoclonal Ab. 1:100), EnVision+ Dual Link System HRP (Dako) and DAB+ (3, 3'-diaminobenzidine; Dako) were subsequently used to visualize the antigen-antibody reactions. Following dehydration and counterstaining with hematoxylin, the slides were embedded in Faramount Aqueous Mounting Medium (Dako). Appropriate positive and negative controls were added.

2.3.2.2. Evaluation of immunohistochemical staining

The stained tissue microarray sections were scanned with a high-resolution optical scanner (NanoZoomer 2.0 HT; Hamamatsu Photonic K.K., Hamamatsu City, Japan) at 20 \times objective magnification (0.5-mm resolution). The scanned sections were analyzed using VIS Image Analysis Software, version 4.5.1.324 (Visiopharm, Hørsholm, Denmark). For IHC scoring, the number of cells with positive cytoplasmic staining was counted and ranged from 0 to 100.

3. RESULTS

3.1. Protein expression profiling *via* PEA

Proteins were extracted from tissues of 24 patients with good prognoses and 23 patients with poor prognoses. Additionally, protein extraction was performed on fallopian tube tissue samples from 6 patients diagnosed with benign diseases. Table 1 presents the clinical characteristics of the samples. Analysis revealed a higher percentage of nulliparity and postoperative residual tissue exceeding 1 cm in patients with poor prognoses, while other characteristics showed no significant differences. When examining the PEA results, of the eight candidates with significantly differential expression between the groups, the top five, ADAMTS15, NPTN, PPM1A, METAP2, and RANGAP1, were more highly expressed in the poor prognosis group, while CD38 was more highly expressed in the good prognosis group (Fig. 2 A). We examined protein expression in other subgroups such as recurrence, stage, CA125 expression, progression-free survival (PFS), BRCA mutation, and overall survival (OS), but did not find any significant differences. (Fig. 2 B). Using the above five proteins, we examined their expression in three IHOSE cell lines (IHOSE79A, IHOSE1421, IHOSE85A) and five serous ovarian cancer cell lines (SKOV3, OVCAR3, OVCA429, PE01, SNU119) by real-time PCR. NPTN and METAP2 were significantly highly expressed in tumor cell lines ($p=0.035$, $p=0.035$). PPM1A and ADAMTS15 were more highly expressed in tumor cell lines, although not statistically significant ($p=0.142$, $p=0.250$). RANGAP1 expression was slightly higher in IHOSE cell lines, but not significant ($p=0.571$) (Fig. 2 C).

Table 1. Characteristics of patients

	Good-prognosis (24)	Poor-prognosis(23)	<i>p</i> -Value
Age			
65≥	19(79.2%)	19(28.6%)	0.529
65<	5(20.8%)	4(17.4%)	
Parity			
nulliparity	0 (0 %)	5(21.7%)	0.02
multiparity	24(100%)	18(78.3%)	
Stage			
stageIII	21(87.5%)	17(73.9%)	0.286
stage IV	3(12.5%)	6(26.1)	
Residual tumor			
1cm <	17(70.8%)	8 (34.8%)	0.02
1cm ≥	7(29.2%)	15(65.2%)	
BRCA mutation			
BRACA(-/-)	6(60%)	11(73.3%)	0.27
BRCA1+	3(30%)	1(6.7%)	
BRCA2+	1(10%)	3(20%)	
NA	14	8	
CA125 (U/ml)	1975.97	2466.63	0.688

p-values were performed by unpaired t-test or Mann-Whitney
 NA, not applicable.

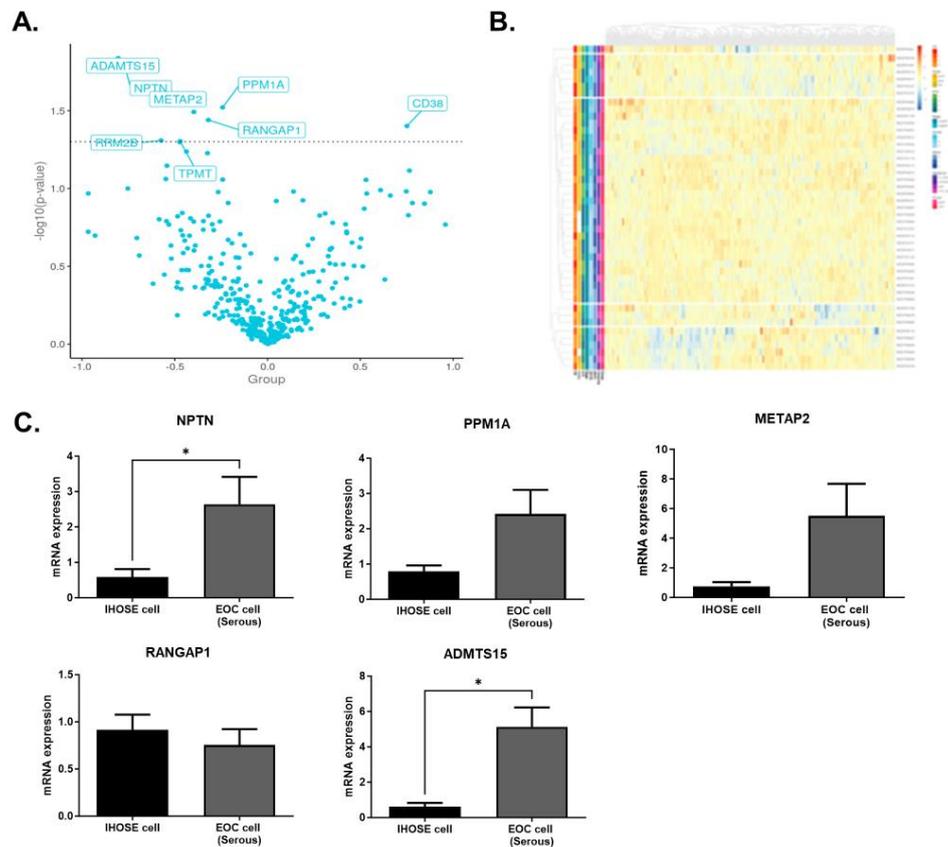


Figure 2. PEA results.

A. Volcano plot of protein expression. Mean differences in NPX (poor-good) are shown on the x-axis and p -values ($-\log_{10}$, two-sided Welch's t -test) on the y-axis. Light dot-line represents cut-offs for p -value and fold change. B. Heatmap of proteins. Protein expression was characterized by various subgroups such as overall survival, Progression-free survival, stage III or IV, ca125 level, BRCA mutation status, and recurrence status, but no meaningful classification was achieved. C. Determined mRNA expression by real-time PCR in three IHOSE cell lines and five epithelial ovarian cancer cell lines. Data are expressed as the mean \pm standard error (S.E.). Mann-Whitney U test was performed. $*p < 0.05$

3.2. Key protein selection based on machine learning

3.2.1. Select proteins with different features - proteins with low deviations

Even though there are many different features in the subgroups, they are relatively small. Therefore, we decided to select features based on their standard deviation. The expression distribution of proteins with high variation shows that even within the same group of patients, the degree of expression is very different (Fig. 3 A (a)). On the other hand, the expression of the proteins with low deviation shows that the distribution of the proteins is similar in the patients (Fig. 3 A (b)). We, therefore, identified about 50 proteins with low deviations in each group and selected proteins that intersected, i.e., proteins with low deviations in both good and bad prognosis groups, as candidate proteins.

3.2.2. k-means clustering using the elbow method.

We utilize k-means clustering to analyze protein data, acknowledging that a narrow data distribution does not ensure effective clustering. To address this, we employ the elbow method to determine the optimal number of clusters (Fig. 3 B). Demonstrates this method, revealing a rapid decrease in the cost function around $k=3$, indicating its suitability for our dataset.

3.2.3. Visualization model construction

We designed a visualization model to discriminate between good and poor groups based on protein expression. Since it is not possible to create a visualization model with all proteins, we chose the above five proteins that had statistical significance and the proteins that clustered well in both groups with low deviation after data preprocessing as candidates. Each set of models was visualized using the K-means clustering method (via the elbow point method), and a heuristic approach was used to find meaningful results among their models (Fig. 3 D).

3.2.4. NPTN and PPM1A are selected

We determine that *NPTN* and *PPM1A* are major factors in determining the patient's prognosis. In Fig. 3 C, we visualize the diagnostic status of each patient based on *NPTN* and *PPM1A* factors. In the heuristic approach, it is possible to determine that the condition of patients is good when the *NPTN* factor is less than 0 or the *PPM1A* factor is less than -0.2.

3.2.5. Evaluation of the proposed scheme

As shown in Table 2, selecting the proper factors allows us to classify two groups using a heuristic technique. Here, Precision, Recall, and F1-score are key metrics for evaluating the performance of

classification methods. Precision indicates the proportion of patients with a poor prognosis among all the patients predicted to have a poor prognosis. Recall refers to the proportion of patients predicted as having a poor prognosis among all the patients with an actual poor prognosis. The F1-score represents the harmonic average of Precision and Recall, which refers to a statistical index used to evaluate the performance of the classification model. When each indicator is close to 1, it means that model classification accuracy and stability are high. This means that it is possible to derive meaningful classification results using only selected protein factors and thresholds on a given data set.

Table 2. Performance analysis using heuristic techniques

Index	Value
Precision	0.857
Recall	0.818
F1-score	0.837

3.3. Knockdown NPTN and PPM1A reduced cell proliferation, migration, and invasion.

NPTN and PPM1A, which were identified as candidate proteins, were expressed in 5 IHOSE and 6 serous ovarian cancers by real-time PCR. We found that both proteins were increased in tumor cells and that their expression was increased in SKOV3 and OVCAR3. Cell experiments were conducted with these two types of cells. SKOV3 was transfected with siNPTN, and siPPM1A at 100nM and 200nM, and negative control was also processed at 200nM. To ensure that NPTN and PPM1A were properly knocked down, we evaluated their efficiency by western blot and real-time PCR.

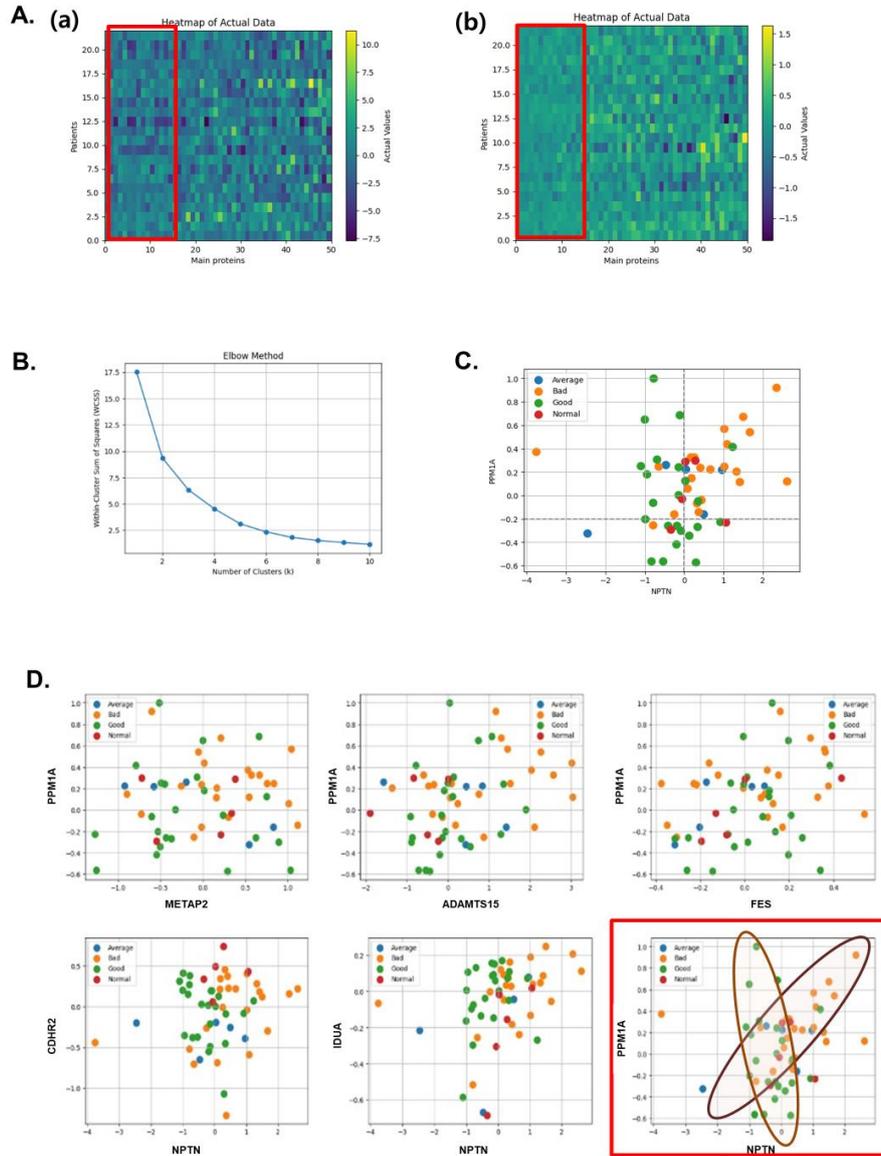


Figure 3. Adaptive model construction based on machine learning

A. Validation of selected factors for the case of poor patients (a) Factors with high deviation values (b) Factors with low deviation values. B. Clustering results with different k values. Clustering results when k = 3. C. Visualizations model made with NPTN and PPM1A. D. Various visualization models.

In SKOV3 cells, the expression of siNPTN 100nM, siNPTN 200nM siPPM1A 100nM, and siPPM1A 200nM were all significantly decreased. ($p=0.004$, $p=0.004$, $p=0.003$, $p=0.003$). In OVCAR3 cells, siNPTN 200nM and siPPM1A 200nM significantly decreased expression ($p=0.006$, $p=0.0001$). Whereas siNPTN 100nM and siPPM1A 100nM decreased expression but were not statistically significant ($p=0.421$, $p=0.23$) (Fig. 4 A, B).

To investigate the association with cell proliferation, which is associated with poor prognosis in cancer, OVCAR3 cells were transfected with siNPTN and siPPM1A and compared to negative control. siNPTN and siPPM1A both significantly decreased cell proliferation at 200nM (siNPTN 200nM $p=0.028$, siPPM1A 200nM $p=0.02$). While at siNPTN 100nM the expression was reduced but not significant ($p=0.213$). At siPPM1A 100nM the expression was significantly reduced ($P=0.027$) (Fig. 4 C).

Transwell migration experiments were performed with SKOV3 cells. SKOV3 was transfected with siNPTN, siPPM1A, and negative control. siNPTN significantly reduced migration at both 100nM and 200nM (siNPTN 100nM $p=0.0003$, siNPTN200nM $p=0.0007$). siPPM1A significantly reduced migration at 200nM (siPPM1A 200nM $p=0.015$). Meanwhile, at 100nM of PPM1A, a slight decrease was visible but not significant (siPPM1A 100nM $p=0.015$) (Fig. 4 C).

Invasion assay was performed in SKOV3 cells. After transfection with siNPTN, siPPM1A, and negative control, SKOV3 was used as before. siNPTN 200nM and siPPM1A 200nM both significantly reduced the invasion (siNPTN200nM $p=0.005$, siPPM1A 200nM $p=0.008$) (Fig. 4 D).

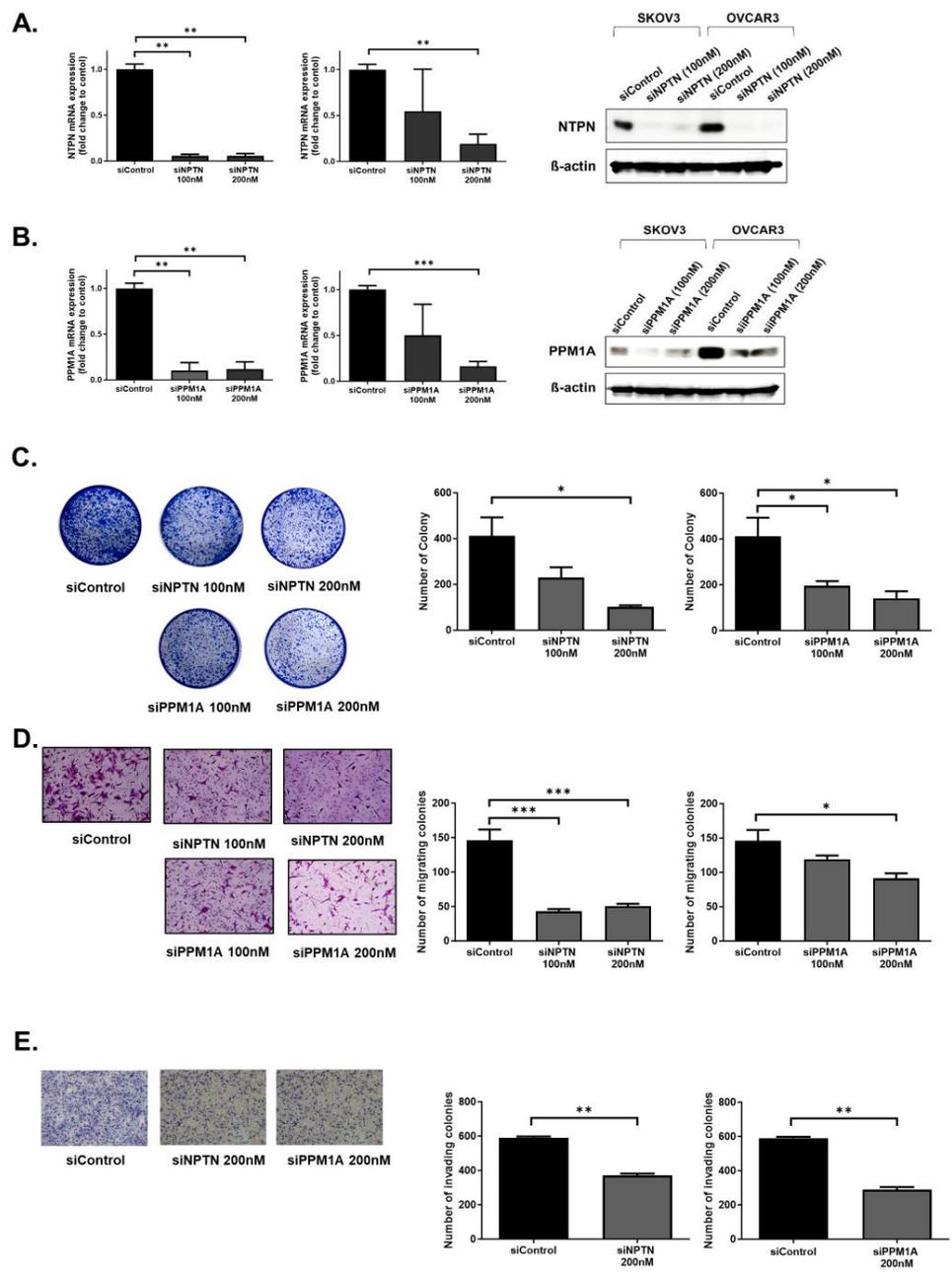


Figure 4. Validation by in-vitro analysis

A. Expression levels after transfection of the *NPTN* gene into the SKOV3 and OVCAR3 cell lines. mRNA expression was determined by real-time PCR. In SKOV3 cells, both 100nM and 200nM significantly decreased the expression. However, in OVCAR3, although both decreased expressions, the decrease in expression at 100nM was not significant. Data are expressed as the mean \pm standard error (S.E.). Statistical significance was determined using the Mann-Whitney U test. * $p=0.05$, ** $p=0.01$. Simultaneously, protein expressions were detected using immunoblotting with β -actinin as an internal loading control. B. *PPM1A* gene was knocked down in SKOV3 and OVCAR3, and the expression was significantly reduced at 100nM and 200nM in SKOV3 cells. On the other hand, in OVCAR3, it was significantly decreased at 200nM, but not at 100nM, but a decrease was confirmed. Data are expressed as the mean \pm standard error (S.E.). Statistical significance was determined using the Mann-Whitney U test. * $p=0.05$, ** $p=0.01$, *** $p=0.001$. C. Cell proliferation assay of siNPTN, siPPM1A, siNegative control in OVCAR3. Data are expressed as the mean \pm standard error (S.E.). Statistical significance was determined using the Mann-Whitney U test. * $p=0.05$. D. Transwell migration assay of siNPTN, siPPM1A, siNegative control in SKOV3. E. Boyden chamber invasion assay with siNPTN, siPPM1A, siNegative control in SKOV3. Data are expressed as the mean \pm standard error (S.E.). Statistical significance was determined using the Mann-Whitney U test. * $p=0.05$, ** $p=0.001$.

3.4. NPTN and PPM1A in IHC

3.4.1. NPTN, PPM1A expressed high in serous ovarian cancer.

IHC was performed both on TMA samples from 91 patients with serous ovarian tumors and on TMA samples from 78 normal ovarian tissues. Tissues from patients used for proteomics were excluded. The association of NPTN and PPM1A expression with the characteristics of the group of patients with serous ovarian tumors was not significant (Table 3). NPTN was expressed in the cytoplasm, nucleus, and membrane, while PPM1A was expressed in the cytoplasm and nucleus (Fig. 5 A). NPTN was significantly expressed in tumor tissues with a value of 8.823 in normal tissues and 54.383 in serous ovarian cancer patients ($p<0.0001$), and PPM1A with a value of 3.54 in normal tissues and 77.824 in serous ovarian cancer patients ($p<0.0001$) (Fig. 5 B).

		NPTN				PPM1A			
		No.	IHC score		<i>p</i> -value	No.	IHC score		<i>p</i> -value
Variables	Mean		SD	Mean			SD		
All		91				91			
Histology	serous	91				91			
Age	65>	67	56.118	32.347	0.637	68	78.254	23.2	0.762
	65≤	23	52.393	32.983		23	67.552	23.289	
CA125	35U/ml<	11	48.148	37.619	0.416	11	78.955	26.021	0.863
	35U/ml≤	78	56.666	31.577		79	77.652	22.998	
Stage	I/II	17	46.15	38.961	0.515	17	74.509	29.601	0.204
	III/IV	73	57.266	30.559		74	78.585	21.516	
Grade	well/moderate	42	50.423	29.132	0.195	42	72.967	25.211	0.063
	poor	48	59.317	34.721		49	81.987	20.486	
Chemo sensitivity	Sensitive	70	54.013	31.666	0.605	70	77.717	22.99	0.945
	Resistant	17	58.5	33.381		17	77.278	25.82	

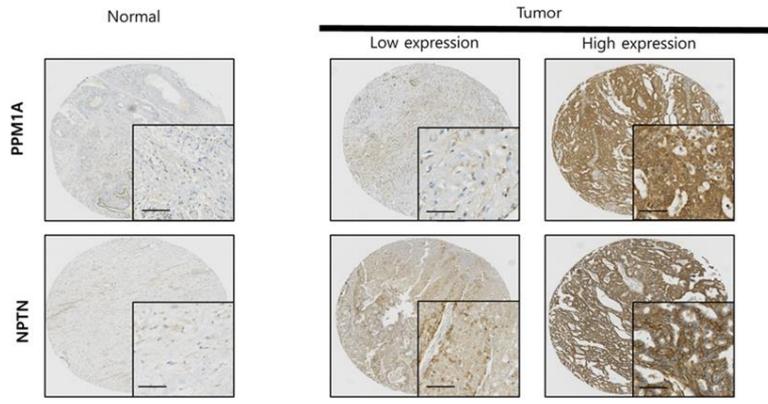
Table 3. Association of NPTN and PPM1A expression with patient clinical characteristics.

P-values were performed by t-test or Mann-Whitney test. SD, standard deviation

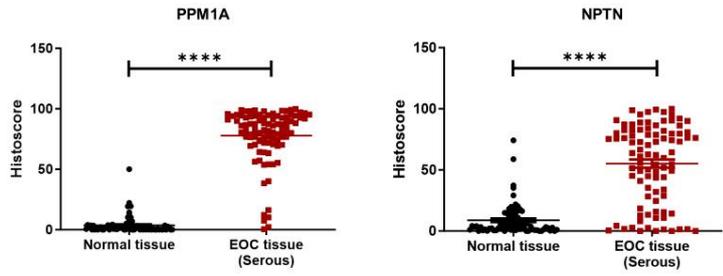
3.4.2. High expression of NPTN, and PPM1A are associated with poor prognosis

We used a cut-off via the receiver operating characteristics (ROC) curve to divide the expression into two groups, high and low, and the survival between the two groups was determined by the Kaplan-Meier curve (Cut-off value of NPTN; 82.255, Cut-off value of PPM1A; 63.579). Survival analysis showed that NPTN had a higher mortality rate when its expression was significantly higher ($p=0.03$), and PPM1A also had a higher mortality rate when its expression was higher, although it was not statistically significant ($p=0.17$). When both NPTN and PPM1A were evaluated simultaneously, the survival curves of both low and high-expression groups showed that mortality was significantly higher in the high-expression group ($p=0.025$) (Fig. 5 C). The group with low NPTN and high PPM1A had worse survival than those with both, although this was not significant (Fig. 5 C).

A.



B.



C.

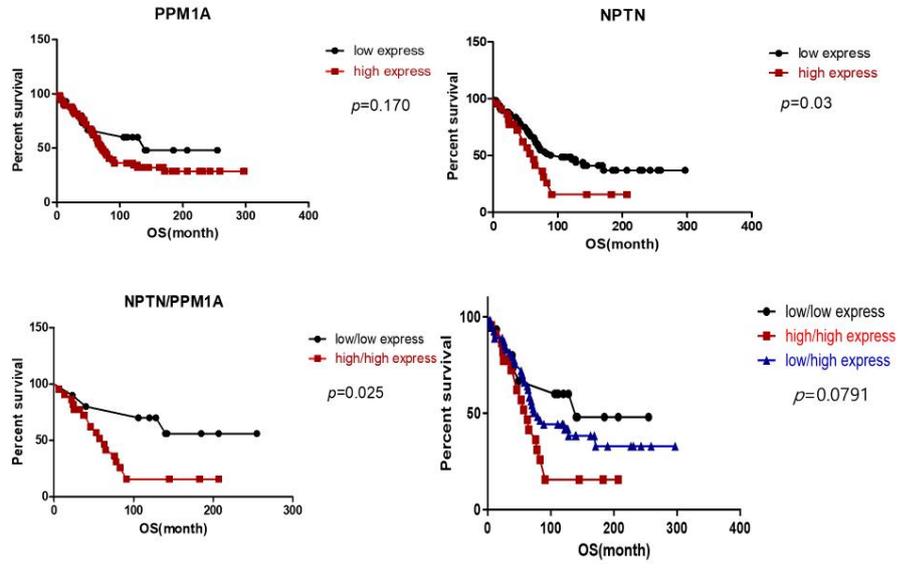


Figure 5. Clinical validation using immunohistochemistry

A. Immunohistochemistry staining of NPTN and PPM1A. (Magnification $\times 200$; scale bar, $10\mu\text{m}$)
B. The expression of both NPTN and PPM1A was significantly increased in serous ovarian cancer tissues. Unpaired t-test was performed. -, ****. C. Kaplan–Meier curve according to NPTN and PPM1A expression in 91 patients. Determined the cut-off value using the Younden index on the ROC curve (Cut-off value of NPTN 82.255, Cut-off value of PPM1A; 63.579). For NPTN there are 69 low and 22 high expression groups and for PPM1A there are 15 low and 76 high expression groups. Together, 15 are low in both, 54 are low in NPTN and high in PPM1A, and 22 are high in both. There were no cases with high PPM1A and low NPTN. Groups with low or high expression were compared. Statistical analysis was performed using the log-rank test.

4. DISCUSSION

Our study presents a novel approach in the realm of ovarian cancer prognosis by integrating PEA with machine learning analysis. This combination allowed us to delve deeper into the proteomic landscape of advanced HGSOC, to identify key proteins associated with prognosis. Our results underscore the significance of protein expression profiles in determining patient outcomes, shedding light on the limitations of solely relying on genomic alterations for prognostic predictions.

Several biomarkers have been identified by proteomic analysis in ovarian cancer, currently recognized biomarkers such as CA125, HE4, and OVERA². Numerous biomarkers associated with ovarian cancer have been identified through various biological approaches like proteomics, lipidomics, and genomics¹⁹. According to the National Cancer Institute's Early Detection Research Network (EDRN) website, there are currently 616 active clinical trials focusing on 232 biomarkers associated with ovarian cancer (<https://edrn.nci.nih.gov/data-and-resources/biomarkers>). Recently, efforts have been made to improve the sensitivity of ovarian cancer screening by interpreting the results in combination with CT or radiologic image²⁰⁻²². However, the majority of these are focused on screening tests that can be used to identify patients with ovarian cancer in the general population at an early stage, as early detection of ovarian cancer is the most important step in improving patient outcomes^{23,24}. In HGSOC, investigations into various biomarkers are predominantly focused on differential diagnosis²⁵. However, there is a scarcity of biomarkers related to prognosis, particularly in advanced stage HGSOC. Kim et al. reported predictors of prognosis in HGSOC by label-free quantitative protein analysis, but there is a limitation that they only analyzed proteins from frozen tissues of 10 patients with PFS 18 and 10 patients with PFS less than or equal to 18^{26,27}. We analyzed proteins by PEA of 386 cancer-related protein biomarkers provided by Olink Bioscience (Uppsala, Sweden), rather than by mass spectrometry, which is often used in novel proteomic studies²⁸⁻³⁰. Combining the sensitivity of PCR with the specificity of antibody-based detection methods, PEA technology is suitable for multiplex biomarker detection and high-throughput quantification³¹. PEA also has very high sensitivity, allowing the detection of low-abundance proteins^{17,32,33}. This approach provided us with a nuanced understanding of the intricate molecular mechanisms underlying ovarian cancer progression, surpassing the conventional genomic profiling methods. Using machine learning algorithms, we were further able to distinguish subtle but crucial differences in protein expression profiles between groups of patients with very good and poor prognoses.

Notably, our analysis identified NPTN and PPM1A as potential prognostic markers for advanced

HGSOC. High expression levels of these proteins were strongly associated with poor prognosis, as corroborated by both clinical validation and in vitro experiments. Our findings suggest that these proteins could serve as reliable indicators of disease progression and could potentially guide personalized therapeutic interventions.

NPTN, a glycoprotein of the immunoglobulin superfamily³⁴, has limited research in tumors, particularly in ovarian cancer. Studies in breast cancer suggest that NPTN- β subtype overexpression promotes tumor growth and angiogenesis, possibly through vascular endothelial growth factor induction under hypoxic conditions³⁵. Its role in promoting metastasis has been observed in breast and lung cancers^{36,37}, indicating its potential as a growth regulator in malignancies, likely through tumor microenvironmental signaling, though precise mechanisms remain unclear.

PPM1A, a magnesium-dependent phosphatase, functions as a negative regulator in several signaling pathways. One notable signaling pathway it regulates is TGF- β ³⁸⁻⁴¹. It has been implicated as a tumor suppressor in breast, pancreatic, and renal cancer⁴²⁻⁴⁵. However, the function of TGF- β is inconsistent with the tumor microenvironment⁴⁶, and the role of PPM1A may vary depending on the tumor type. There are currently no reports on PPM1A in ovarian cancer or ovarian cancer cell lines. Though our study presents promising results, several limitations warrant consideration. First, it is usually convenient to divide patients into groups for prognostic purposes and interpret the results using clear criteria of OS or PFS. However, we looked at three factors simultaneously: recurrence, progression, and overall survival. This is because there are many cases of chemoresistance, refractoriness, and recurrence in HGSOC^{16,47}, so we looked at these three factors simultaneously rather than just PFS and OS, which is not enough to distinguish between those with good prognosis and those with poor prognosis. Secondly, there is the limitation of the small number of tissues. However, the number of tissues in our study is not small among studies that have included long-term survivors of more than 10 years in experiments with chemotherapy-naive frozen tissues, and this is something that should be overcome in future collaborative studies with other institutions. Third, in the validation of our NPTN and PPM1A, we would like to see replication with PEA in other HGSOC patient populations. This would require a long time and high cost, which we have addressed by performing IHC on TMA from patients with serous ovarian cancer. IHC is less expensive and easier to use in the clinical setting than whole genome sequencing or RNA sequencing

48,49

.5. CONCLUSION

In conclusion, our study demonstrates the potential of integrating PEA with machine learning analysis as a robust strategy for identifying prognostic markers in ovarian cancer. The identification of NPTN and PPM1A as potential prognostic indicators highlights the importance of proteomic profiling in guiding individualized therapeutic strategies. Future translational efforts are warranted to harness the clinical utility of these findings and improve patient outcomes in advanced high-grade serous ovarian cancer.

REFERENCES

1. Corbo C, Cevenini A, Salvatore F. Biomarker discovery by proteomics-based approaches for early detection and personalized medicine in colorectal cancer. *Proteomics Clin Appl* 2017;11.
2. Ryu J, Thomas SN. Quantitative Mass Spectrometry-Based Proteomics for Biomarker Development in Ovarian Cancer. *Molecules* 2021;26.
3. Augoff K, Hryniewicz-Jankowska A, Tabola R, Stach K. MMP9: A Tough Target for Targeted Therapy for Cancer. *Cancers (Basel)* 2022;14.
4. Callahan N, Tullman J, Kelman Z, Marino J. Strategies for Development of a Next-Generation Protein Sequencing Platform. *Trends Biochem Sci* 2020;45:76-89.
5. Chin L, Andersen JN, Futreal PA. Cancer genomics: from discovery science to personalized medicine. *Nat Med* 2011;17:297-303.
6. Meador CB, Hata AN. Acquired resistance to targeted therapies in NSCLC: Updates and evolving insights. *Pharmacol Ther* 2020;210:107522.
7. Lisio MA, Fu L, Goyeneche A, Gao ZH, Telleria C. High-Grade Serous Ovarian Cancer: Basic Sciences, Clinical and Therapeutic Standpoints. *Int J Mol Sci* 2019;20.
8. Bukłaho PA, Kiśluk J, Wasilewska N, Nikliński J. Molecular features as promising biomarkers in ovarian cancer. *Adv Clin Exp Med* 2023;32:1029-40.
9. Berek JS, Renz M, Kehoe S, Kumar L, Friedlander M. Cancer of the ovary, fallopian tube, and peritoneum: 2021 update. *Int J Gynaecol Obstet* 2021;155 Suppl 1:61-85.
10. Babaier A, Mal H, Alselwi W, Ghatage P. Low-Grade Serous Carcinoma of the Ovary: The Current Status. *Diagnostics (Basel)* 2022;12.
11. Hollis RL. Molecular characteristics and clinical behaviour of epithelial ovarian cancers. *Cancer Lett* 2023;555:216057.
12. Armstrong DK, Alvarez RD, Bakkum-Gamez JN, Barroilhet L, Behbakht K, Berchuck A, et al. Ovarian Cancer, Version 2.2020, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* 2021;19:191-226.
13. Grisham RN, Slomovitz BM, Andrews N, Banerjee S, Brown J, Carey MS, et al. Low-grade serous ovarian cancer: expert consensus report on the state of the science. *Int J Gynecol Cancer* 2023;33:1331-44.
14. Irodi A, Rye T, Herbert K, Churchman M, Bartos C, Mackean M, et al. Patterns of clinicopathological features and outcome in epithelial ovarian cancer patients: 35 years of

- prospectively collected data. *Bjog* 2020;127:1409-20.
15. Le Page C, Rahimi K, Köbel M, Tonin PN, Meunier L, Portelance L, et al. Characteristics and outcome of the COEUR Canadian validation cohort for ovarian cancer biomarkers. *BMC Cancer* 2018;18:347.
 16. Dao F, Schlappe BA, Tseng J, Lester J, Nick AM, Lutgendorf SK, et al. Characteristics of 10-year survivors of high-grade serous ovarian carcinoma. *Gynecol Oncol* 2016;141:260-3.
 17. Wik L, Nordberg N, Broberg J, Bjorkesten J, Assarsson E, Henriksson S, et al. Proximity Extension Assay in Combination with Next-Generation Sequencing for High-throughput Proteome-wide Analysis. *Mol Cell Proteomics* 2021;20:100168.
 18. Shin HY, Yang W, Lee EJ, Han GH, Cho H, Chay DB, et al. Establishment of five immortalized human ovarian surface epithelial cell lines via SV40 T antigen or HPV E6/E7 expression. *PLoS One* 2018;13:e0205297.
 19. Atallah GA, Abd Aziz NH, Teik CK, Shafiee MN, Kampan NC. New Predictive Biomarkers for Ovarian Cancer. *Diagnostics (Basel)* 2021;11.
 20. Gu R, Tan S, Xu Y, Pan D, Wang C, Zhao M, et al. CT radiomics prediction of CXCL9 expression and survival in ovarian cancer. *J Ovarian Res* 2023;16:180.
 21. Hong Y, Liu Z, Lin D, Peng J, Yuan Q, Zeng Y, et al. Development of a radiomic-clinical nomogram for prediction of survival in patients with serous ovarian cancer. *Clin Radiol* 2022;77:352-9.
 22. Hu J, Wang Z, Zuo R, Zheng C, Lu B, Cheng X, et al. Development of survival predictors for high-grade serous ovarian cancer based on stable radiomic features from computed tomography images. *iScience* 2022;25:104628.
 23. Zheng H, Zhang L, Zhao Y, Yang D, Song F, Wen Y, et al. Plasma miRNAs as diagnostic and prognostic biomarkers for ovarian cancer. *PLoS One* 2013;8:e77853.
 24. Huh S, Kang C, Park JE, Nam D, Kim SI, Seol A, et al. Novel Diagnostic Biomarkers for High-Grade Serous Ovarian Cancer Uncovered by Data-Independent Acquisition Mass Spectrometry. *J Proteome Res* 2022;21:2146-59.
 25. Dieters-Castator DZ, Rambau PF, Kelemen LE, Siegers GM, Lajoie GA, Postovit LM, et al. Proteomics-Derived Biomarker Panel Improves Diagnostic Precision to Classify Endometrioid and High-grade Serous Ovarian Carcinoma. *Clin Cancer Res* 2019;25:4309-19.
 26. Kim SI, Jung M, Dan K, Lee S, Lee C, Kim HS, et al. Proteomic Discovery of Biomarkers

- to Predict Prognosis of High-Grade Serous Ovarian Carcinoma. *Cancers (Basel)* 2020;12.
27. Kim SI, Hwangbo S, Dan K, Kim HS, Chung HH, Kim JW, et al. Proteomic Discovery of Plasma Protein Biomarkers and Development of Models Predicting Prognosis of High-Grade Serous Ovarian Carcinoma. *Mol Cell Proteomics* 2023;22:100502.
 28. Filbin MR, Mehta A, Schneider AM, Kays KR, Guess JR, Gentili M, et al. Longitudinal proteomic analysis of severe COVID-19 reveals survival-associated signatures, tissue-specific cell death, and cell-cell interactions. *Cell Rep Med* 2021;2:100287.
 29. Enroth S, Ivansson E, Lindberg JH, Lycke M, Bergman J, Reneland A, et al. Data-driven analysis of a validated risk score for ovarian cancer identifies clinically distinct patterns during follow-up and treatment. *Commun Med (Lond)* 2022;2:124.
 30. Gyllensten U, Hedlund-Lindberg J, Svensson J, Manninen J, Ost T, Ramsell J, et al. Next Generation Plasma Proteomics Identifies High-Precision Biomarker Candidates for Ovarian Cancer. *Cancers (Basel)* 2022;14.
 31. Boylan KLM, Geschwind K, Koopmeiners JS, Geller MA, Starr TK, Skubitz APN. A multiplex platform for the identification of ovarian cancer biomarkers. *Clin Proteomics* 2017;14:34.
 32. Franzen B, Kamali-Moghaddam M, Alexeyenko A, Hatschek T, Becker S, Wik L, et al. A fine-needle aspiration-based protein signature discriminates benign from malignant breast lesions. *Mol Oncol* 2018;12:1415-28.
 33. Lundberg M, Eriksson A, Tran B, Assarsson E, Fredriksson S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res* 2011;39:e102.
 34. Empson RM, Buckby LE, Kraus M, Bates KJ, Crompton MR, Gundelfinger ED, et al. The cell adhesion molecule neuroligin-1 inhibits hippocampal long-term potentiation via a mitogen-activated protein kinase p38-dependent reduction in surface expression of GluR1-containing glutamate receptors. *J Neurochem* 2006;99:850-60.
 35. Rodriguez-Pinto D, Sparkowski J, Keough MP, Phoenix KN, Vumbaca F, Han DK, et al. Identification of novel tumor antigens with patient-derived immune-selected antibodies. *Cancer Immunol Immunother* 2009;58:221-34.
 36. Bajkowska K, Sumardika IW, Tomonobu N, Chen Y, Yamamoto KI, Kinoshita R, et al. Neuroligin-mediated upregulation of solute carrier family 22 member 18 antisense

- (SLC22A18AS) plays a crucial role in the epithelial-mesenchymal transition, leading to lung cancer cells' enhanced motility. *Biochem Biophys Rep* 2020;22:100768.
37. Powell AA, Talasaz AH, Zhang H, Coram MA, Reddy A, Deng G, et al. Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines. *PLoS One* 2012;7:e33788.
 38. Sun W, Yu Y, Dotti G, Shen T, Tan X, Savoldo B, et al. PPM1A and PPM1B act as IKKbeta phosphatases to terminate TNFalpha-induced IKKbeta-NF-kappaB activation. *Cell Signal* 2009;21:95-102.
 39. Li M, Xu X, Su Y, Shao X, Zhou Y, Yan J. A comprehensive overview of PPM1A: From structure to disease. *Exp Biol Med (Maywood)* 2022;247:453-61.
 40. Lu X, An H, Jin R, Zou M, Guo Y, Su PF, et al. PPM1A is a RelA phosphatase with tumor suppressor-like activity. *Oncogene* 2014;33:2918-27.
 41. Ge Q, Shi Z, Zou KA, Ying J, Chen J, Yuan W, et al. Protein phosphatase PPM1A inhibition attenuates osteoarthritis via regulating TGF- β /Smad2 signaling in chondrocytes. *JCI Insight* 2023;8.
 42. Mazumdar A, Tahaney WM, Reddy Bollu L, Poage G, Hill J, Zhang Y, et al. The phosphatase PPM1A inhibits triple negative breast cancer growth by blocking cell cycle progression. *NPJ Breast Cancer* 2019;5:22.
 43. Fan J, Yang MX, Ouyang Q, Fu D, Xu Z, Liu X, et al. Phosphatase PPM1A is a novel prognostic marker in pancreatic ductal adenocarcinoma. *Hum Pathol* 2016;55:151-8.
 44. Hong Y, Gong L, Yu B, Dong Y. PPM1A suppresses the proliferation and invasiveness of RCC cells via Smad2/3 signaling inhibition. *J Recept Signal Transduct Res* 2021;41:245-54.
 45. Gifford CC, Tang J, Costello A, Khakoo NS, Nguyen TQ, Goldschmeding R, et al. Negative regulators of TGF- β 1 signaling in renal fibrosis; pathological mechanisms and novel therapeutic opportunities. *Clin Sci (Lond)* 2021;135:275-303.
 46. Syed V. TGF-beta Signaling in Cancer. *J Cell Biochem* 2016;117:1279-87.
 47. Gockley A, Melamed A, Bregar AJ, Clemmer JT, Birrer M, Schorge JO, et al. Outcomes of Women With High-Grade and Low-Grade Advanced-Stage Serous Epithelial Ovarian Cancer. *Obstet Gynecol* 2017;129:439-47.
 48. Duraiyan J, Govindarajan R, Kaliyappan K, Palanisamy M. Applications of immunohistochemistry. *J Pharm Bioallied Sci* 2012;4:S307-9.

49. Masuda S, Nakanishi Y. Application of Immunohistochemistry in Clinical Practices as a Standardized Assay for Breast Cancer. *Acta Histochem Cytochem* 2023;56:1-8.

ABSTRACT (IN KOREAN)

진행된 고등급 장액성 난소암에서 단백질 분석을 통한 예후 인자의 발굴: 장기 생존자와 조기 사망 환자 간의 비교 연구

고등급 장액성 난소암은 높은 치사율에도 불구하고 예후를 예측할 수 있는 신뢰할 만한 바이오마커가 부족해 개인 맞춤형 치료의 진전이 제한적이었다. 계놈을 기반으로 한 많은 연구가 이어지고 있지만, 유전적 변형만으로는 암세포의 표현형을 결정하지 못하고 직접적인 역할을 하며 암의 표현형을 결정하는 것은 결국 단백질이기에 그 무엇보다도 단백질 발현을 확인하는 것이 더 중요하다고 할 수 있다. 이 연구는 머신러닝을 활용하여 난소암 예후를 결정하는 단백질 인자를 선별하는 새로운 기법을 제안했다. 우리는 고등급 장액성 난소암 환자군들 중 예후가 매우 좋은 군과 예후가 좋지 않은 군을 설정한 뒤, 진행성 고등급 장액성 난소암 환자의 항암 전 신선 동결 조직을 예후가 좋은군 24개, 예후가 안좋은 군 23개씩 확보하였다. 이들 조직에서 단백질을 추출한 뒤 microarray 기법을 활용하여 단백질을 정량화 하는 Olink사의 Proximity extension assay를 이용하여 단백질 분석을 하였다. 그리고

그 결과를 바탕으로 인공지능 기반 머신 러닝을 통하여 분석해 두개의 군을 예측할 수 있는 모델을 만들었다. 그 결과 NPTN과 PPM1A가 발현량이 높을 때 예후가 나쁜 군과 좋은 군을 분류해서 볼 수 있었고 그 모델의 성능은 Precision 0.857, Recall 0.818, F-1 score 0.893로 훌륭했다.

우리는 NPTN과 PPM1A를 장액성 난소암 환자의 조직 마이크로어레이에서 면역 염색을 시행하였고, 발현량이 높을수록 생존 지수가 낮아지는 결과를 확인하였다. 또한, NPTN과 PPM1A의 발현을 감소시킨 세포를 가지고 세포 증식 실험, 세포 이동 실험, 보이든 챔버 침투 실험을 시행하였다. 그 결과, 두개의 단백질이 발현이 적을 때 증식, 이동, 침투가 감소됨을 확인하였다. 위의 결과들을 통해 우리는 그룹 간에 유의미한 차이가 있는 인자, 특히 단백질을 기반으로 클러스터링에 적합한 인자를 선별하는 것이 가능하다는 것을 시사하였다. 이 연구는 환자 치료 결과를 개선하기 위한 개인 맞춤형 치료의 발전에 있어 단백질 마커의 잠재적인 가능성을 다시 한번 확인하게 해주었다.

핵심되는 말

고등급 장액성 난소암, 단백질분석, 머신러닝, 예후인자, 예후예측