Classification and molecular biological characterization of serous ovarian cancer according to lncRNA expression

Seonhyang Jeong

Department of Medicine

The Graduate School, Yonsei University

Classification and molecular biological characterization of serous ovarian cancer according to lncRNA expression

Directed by Professor Young Suk Jo

The Doctoral Dissertation
submitted to the Department of Medicine,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Medical Science

Seonhyang Jeong

December 2023

This certifies that the Doctoral Dissertation of Seonhyang Jeong is approved.

Thesis Supervisor : Young Suk Jo

Thesis Committee Member#1 : Jan Dee Lee

Thesis Committee Member#2 : Sang Geun Jung

Thesis Committee Member#3 : Jung Hyun Yoon

Thesis Committee Member#4 : Soo Han Bae

The Graduate School
Yonsei University

December 2023

<TABLE OF CONTENTS>

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Classification and molecular biological characterization of serous ovarian cancer according to lncRNA expression

Seonhyang Jeong

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Young Suk Jo)

Ovarian cancer (OvCa) is the deadliest gynecological malignant tumor, resulting in more than 130,000 cancer-related deaths annually worldwide. High-grade serous carcinoma (HGS_OvCa), the most common histological subtype of OvCa, constitutes 70% of diagnosed Epithelial Ovarian Cancer cases and is first diagnosed at an advanced stage when the tumor has spread to the abdomen or outside the abdominal cavity because it has no specific symptoms. Although surgery combined with chemotherapy is common, 75% of treated patients exhibit short survival rates due to drug resistance and relapse. New targets are being discovered to overcome these therapeutic limitations, and one of them is lncRNAs. LncRNAs are being investigated as new diagnostic and therapeutic targets in various types of human cancers, and recent studies suggest that lncRNAs play an important role in regulating tumor progression, metastasis, estrogen response, and drug resistance in OvCa. The purpose of this study is to confirm the characteristics of lncRNA expression through functional analysis based on the results of cluster classification using HGS_OvCa RNA-seq data. It also aims to identify factors that reflect these characteristics. CNMF clustering was performed using the data of 367 HGS_OvCa patients in the TCGA database, and clusters reflecting prognosis were selected through survival analysis. In addition, each cluster was classified into "Immune group", "EMT group", "Estrogen response group", "EMT-Androgen response group", and "Differentiation group" through functional analysis. To identify the factors affecting these characteristics, DNA mutations, somatic copy

number alterations, and miRNA and DNA methylation expression patterns were analyzed. In addition, transcription factors regulating lncRNAs and mRNAs were classified according to cluster. Among the J4 clusters, MSC, AEBP1, and CREB3L1 were selected as master transcription factors (MTFs) based on Eigen centrality. Additionally, I identified seven lncRNAs (LINC01614, LINC00702, AL109924.2, LINC02544, AL356417.2, AC112721.2, LINC01929) that exert a stronger influence than the selected master transcription factors. Through *in vitro* studies, I validated that these lncRNAs regulate both the master transcription factors and EMT-related genes. This study suggests that the seven lncRNAs regulating MTFs contribute to identifying the transcription factor regulation mechanism of lncRNAs in HGS_OvCa and can be indicators for personalized medicine.

Key words : ovarian cancer, high grade serous ovarian cancer, long noncoding RNA, biomarker, multi-omics data

Classification and molecular biological characterization of serous ovarian cancer according to lncRNA expression

Seonhyang Jeong

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Young Suk Jo)

I. INTRODUCTION

Ovarian cancer (OvCa) is the deadliest gynecological malignant tumor, causing more than 130,000 cancer-related deaths annually worldwide[1]. The overall 5-year survival rate was only about 40% after surgical and systemic chemotherapy treatment[2]. The survival rate of patients with OvCa is low because most of the patients are diagnosed at the advanced stage of metastatic tumors[3]. Depending on the type of cell in which the tumor develops, OvCa can be classified as stromal, embryonic or epithelial (EOC), the latter being the most common and accounting for 90% of cases[4]. Within EOC, five histological subtypes (high-grade serous, low-grade serous, mucinous, clear-cell and endometrioid) exist[5,6]. They are distinguished on the basis of histological structure, mutations in specific proto-oncogenes or tumor suppressor genes, chemosensitivity, metastatic behavior, and prognosis[7,8]. High-grade serous carcinoma (HGS_OvCa), the most common histological subtype, constitutes 70% of diagnosed EOC cases and is first diagnosed at an advanced stage when the tumor has spread to the abdomen or outside the abdominal cavity because it has no specific symptoms[9-11]. Although surgery combined with chemotherapy is commonly used to treat this malignancy, 75% of treated patients show short survival rates due to drug resistance and relapse[11]. Therefore, many biomarkers have been suggested, as early tumor detection can improve overall 5-year survival rates.

Long non-coding RNAs (lncRNAs) are transcripts that were identified in genomic studies in the late 1990s and 2000s. They are defined as longer than 200 nucleotides and are presumed not to encode proteins[12]. The hallmark of lncRNAs is that they regulate gene expression through chromatin remodeling, transcription, alternative splicing, and generating microRNAs (miRNAs) or short biologically active peptides[13-15]. Recently, research on lncRNA as a biomarker of ovarian cancer has been extensively reported. For example, overexpression of the lncRNA ABHD11-AS1 promotes tumor progression in EOC through targeted regulation of RhoC in ovarian cancer cells A2780 and OVCAR3[16]. Casc2 and FLJ33360, respectively, distinguish the serous and high-grade serous subtypes from the others[17,18]. DANCR[19] induce angiogenesis by activating the expression of VEGF and SEMA4D. lncGHET1[20] and LINK-A[21] are known to regulate HIF1 and HIF1a to regulate VEGF expression, respectively. Thus, these two lncRNAs are involved in angiogenesis. lnc-miR503HG may promote methylation of miR-31-5p and act as a sponge to inhibit OvCa cell invasion and migration[22].

Class discovery using gene expression profiling has identified clinically relevant subtypes in solid malignancies, such as breast and lung cancer[23,24]. Ovarian cancer molecular subtypes were first identified using four HGSC subtypes with microarray data from 285 Australian patients[25]. Later, The Cancer Genome Atlas (TCGA) also reported four subtypes classified as “Mesenchymal”, “Differentiated”, “Proliferative” and “Immunoreactive” based on RNA sequencing data. Most subtypes have been classified based on microarray data or mRNA expression data[26]. In 2016, a study was published that classified subtypes using DNA methylation, protein, microRNA, and gene expression data for ovarian cancer provided by TCGA. Subtypes were classified according to gene expression and the results of confirming the correlation with the subtypes of the remaining data were presented[27]. lncRNA have not been as deeply studied as their counterparts, miRNAs, and many questions remain about their mechanisms of action and effects in the context of cancer, including ovarian cancer.

In recent studies, transcription factors or transcription factor groups have been identified

as key drivers of biological or disease processes' development[28,29]. Even though there is a group of transcription factors that regulate oncogenes and disease processes, there appears to be a hierarchy in the regulatory activities of these transcription factors Master transcription factors often appear to control most regulatory activities of other transcription factors and related genes and are known to play a role in promoting cell survival, metastasis or tumorigenesis[30-32]. As such, these factors represent crucial therapeutic targets for cancer treatment. Although the critical TFs driving HGSC development have not been characterized, studies have recently been published that identified and validated MTFs, including PAX8, SOX17, and MECOM[33-35]. In particular, PAX8 and SOX17 cooperate to positively regulate cell cycle progression and angiogenesis in HGSCs[36,37]. However, Master Transcription Factors (MTFs) that regulate lncRNAs and are regulated by lncRNAs have not yet been identified.

Therefore, in this study, clusters according to lncRNA expression were separated using high-grade serous ovarian cancer (HGS_OvCa) RNA-seq data, and transcription factors representing each characteristic were identified. Among the selected transcription factors, MTFs that regulate transcription factors were identified, and lncRNAs that regulate MTFs were additionally identified and proposed. This study provides new insight into the impact of lncRNAs that regulate MTFs in cancer malignancies by regulating the expression of downstream target genes through various signaling mechanisms.

II. MATERIALS AND METHODS

1. Public Databases

Data pertaining to 367 ovarian cancer samples were obtained from The Cancer Genome Atlas (TCGA) Genomic Data Commons (GDC) portal [38]. The downloaded dataset included total RNA-seq, miRNA, methylation, and clinical data. TCGA-OV GISTIC2 gene-level copy number and somatic mutation data were downloaded from The Broad Institute TCGA GDAC Firehose [39] with no further processing.

Within the Cancer Cell Line Encyclopedia (CCLE) project[40], I identified 47 cell lines representing major subtypes of ovarian cancer and analyzed them using RNA sequencing data. Among them, cell lines were isolated according to the widely accepted ovarian cancer classification paradigm based on clinico-pathological and molecular evidence that Type I and Type II tumors develop through different pathways[41-43]. As a result, 28 Type I cell lines and 19 Type II cell lines could be identified. Since all Type II ovarian cancer cell lines correspond to HGSC, RNA-seq data of the 19 cell lines classified were analyzed.

2. RNA-seq data processing and Subtype discovery

Among the total RNA-seq data 60,433 identifiers were downloaded using TCGA GDC; lncRNAs and coding RNAs were classified using the GTF file (v30) classified lncRNAs provided by GENCODE[44]. As a result, 15,171 lncRNAs were classified. Among the remaining 45,262 genes, 20,531 coding genes were classified using the HUGO probe map. Among the total 15,171 classified lncRNAs, the top 1,500 lncRNAs were selected through the Median Absolute Deviation (MAD) method[45].

Clustering was performed using the consensus nonnegative matrix factorization (CNMF) method for the selected top 1,500 lncRNAs. The analysis was conducted using the ConsensusClusterPlus R package (parameters: max$K$=6, reps = 100)[46]. This method computes multiple k-factor decompositions of expression matrices and evaluates their stability. In the consensus matrix, consensus values ranging from 0 (never clustered together) to 1 (always clustered together) were marked by blue to red. The sample

correlation matrix ranges from -1.0 to 1.0 and is colored white to blue. Clustering results were verified using cophenetic coefficient[47], average silhouette width[48], and total within sum of square[49]. When cophenetic coefficient and average silhouette width are closer to 1, they indicated a suitable cluster, and total within sum of square is a method of determining the point where the WSS (within sum of square) rapidly decreases based on the sum of squares of distances within the cluster as the appropriate number of clusters. The appropriate number of clusters was determined based on the results of the three test methods.

After that, the mRNA expression pattern was confirmed based on the group according to lncRNA expression. From the 20,531 coding genes classified above, the top 1,500 were selected using the MAD method, similar to lncRNA, and then differentially expressed genes were identified according to the group.

3. Identification of a signature predictive of survival in ovarian cancer

Overall survival and disease-free survival, were calculated as the number of years between the year of diagnosis and the year of all-cause death, the date of last follow-up, or 5-year censored survival data. Survival analysis was assessed by comparing overall survival and disease-free survival according to clusters classified using Kaplan-Meier curves and applying Log-rank[50]. The survival curve was plotted using the R package 'survminer'[51].

4. Functional enrichment analysis

Functional analysis was conducted using different tools for lncRNAs and mRNAs. First, the function of lncRNAs was checked using the funcpred database[52]. Funcpred proposed a method to indirectly confirm the function of lncRNAs based on the function of the matching coding gene by matching the lncRNA sequence with the coding gene sequence. The lncRNAs matching the coding genes were reclassified into gene sets to which the coding genes belonged. Analysis was performed using the hallmark gene annotation source.

Hallmark geneset results satisfying p-value < 0.05 were selected and the number of lncRNAs contained in each geneset was counted.

Second, mRNA functional analysis was performed using gprfiler[53]. Gprofiler is a public web server used for characterizing and manipulating gene lists resulting from high-throughput mining of genomic data. mRNAs expressed in each cluster were selected, and functional analysis was conducted using hallmark genesets, similar to the approach used for lncRNAs.

5. Identification of significantly mutated genes

Analysis was performed using the R package 'MAFtools' to investigate mutational differences between groups[54]. This tool facilitated variant analysis on Mutation Annotation Format (MAF) data, allowing mutation patterns to be explored and compared across different groups in a study. Mutations were confirmed by dividing the entire HGSC sample and each cluster.

6. Analysis of Copy Number Alteration

To determine copy number alterations in TCGA ovarian cancer (OVCA), Genomic Identification of Significant Targets in Cancer (GISTIC) data from the Broad Firehose infrastructure were used[55]. GISTIC statistically calculated the copy number alterations occurring in many patient specimens. The threshold used for DNA copy number amplification and deletion was 0.1, the confidence level was 0.99, and the q-value cut-off was 0.25.

7. miRNA data preprocessing and analysis

Among a total of 1,881 miRNA identifiers, miRNAs that maintained gene expression in 10% of the samples with a count value of 10 or more were selected. As a result, a total of 486 miRNAs were classified. Differentially expressed genes were identified through

DESeq using the selected miRNAs[56]. Group information was analyzed using the previously analyzed lncRNA group.

8. DNA methylation data for model construction

DNA methylation profile was measured experimentally using the Illumina Infinium HumanMethylation27 platform. DNA methylation values, described as beta values, are recorded for each array probe in each sample via BeadStudio software. The DNA methylation beta value is a continuous variable between 0 and 1 and represents the ratio of the intensity of methylation to a bound gene. Thus, higher beta values represent higher level of DNA methylation, i.e. hypermethylation, and lower beta values represent lower level of DNA methylation, i.e. hypomethylation[57].

9. Identification of Transcription Factors

Master regulator inference analysis was performed based on the ARACNe and viper algorithms[58,59]. Based on the official tutorial on GitHub (https://github.com/califano-lab/ARACNe-AP), the analysis was conducted based on the ARACNe-AP package with default parameters excluding the Transcription Factor list. For transcription factor lists, GO transcription factors, active gene lists, TRRUST[60] and GTRD[61] lists were collected. Among the collected lists, those known as transcritpion factors were reclassified to finally identify 2,192 transcription factors[62]. In each cluster, the lncRNAs and mRNAs with increased expression were selected, and positive correlation (Pearson's Correlation $R \geq 0.4$) lncRNAs and mRNAs were selected again. Since lncRNA as a transcription factor target gene does not exist in the list, an indirect verification method through mRNA expression was used. A regulon object was created from an ARACNe network file and the corresponding expression dataset using the aracne2regulon function from the viper package with default parameters. To compare gene expression changes between groups, we employed a t-test and generated a null model through sample permutations and subsequent t-tests. Master regulator inference analysis was performed with the msviper function in the viper package

using t-statistics and corresponding p-values and a null model. Analysis results were judged to be significant when p-value was less than 0.05.

10. Construction of a network of master transcription factors

Based on the results of the selected master transcription factors, the eigen centrality of each TF was identified using the tidygraph R package[63]. After confirming centrality, lncRNAs with a positive correlation (Pearson's Correlation $R \geq 0.4$) with the target gene of the selected Master Transcription Factor were selected. Through this method, master TF - mRNA - lncRNA could be linked, and the result was visualized using Cytoscape (version 3.9.1)[64]. In addition, the topological properties of the regulatory network were visualized and analyzed, and the directionality of the target gene was confirmed using the edge bundle function, a Cytoscape plug-in.

11. Transcription factors target related gene enrichment analysis

Based on the result of confirming the master transcription factor, the target gene of each TF was selected. A single-sample Gene Set Enrichment Analysis (ssGSEA) was performed using the CCLE data set to confirm expression at the cellular level of selected target gene sets[65]. The ssGSEA is a rank-based algorithm that calculates a score illustrating the level of absolute enrichment of a particular gene set in each sample. ssGSEA was analyzed through GenePattern (https://www.genepattern.org/), and the analysis results were visualized using the heatmap R package[66].

12. Cell culture

High-grade serous carcinoma (HGSC), Caov3 (cat#30075), OVCAR3 (cat#30161), and SNU8 (cat#00008) cell lines were purchased from the Korean Cell Line Bank (Seoul, Korea) and 59M (cat#89081802) cell line was purchased from ECACC (European Collection of Authenticated Cell Cultures, UK). All cell lines were authenticated using STR

profiling. The OVCAR3, SNU8 and 59M cell lines were cultured in RPMI (cat# 10040CV, Corning, NY, USA), supplemented with 10% fetal bovine serum (cat# 35015CV, Corning, NY, USA), 1% penicillin, and streptomycin. The CAOV3 cell line was cultured in DMEM (cat# 10013CV, Corning, NY, USA) supplemented with 10% fetal bovine serum (cat# 35015CV, Corning, NY, USA) and 1% penicillin and streptomycin (cat# 15140122, Thermo Fisher Scientific, Waltham, MA, USA).

13. Short interfering RNA (siRNA) transfection

Three different small interfering RNAs (siRNAs) targeting MSC (cat# 9242-1, 9242-2, 9242-3), AEBP1 (cat# 165-1, 165-2, 165-3), and CREB3L1 (cat# 90993-1, 90993-2, 90993-3) were purchased from Bioneer (Daejeon, Korea). Non-targeting siRNA was used as a negative control. The RNAi oligonucleotide or RNAi negative control was transfected into the cells using a Lipofectamine RNAiMAX Transfection Reagent (Thermo Fisher Scientific,Waltham, MA, USA) according to the manufacturer's instructions.

14.RNA Isolation and Real-Time PCR

Total RNA from cells was extracted using TRIzol® (Invitrogen), and cDNA was synthesized from total RNA with SuperScript® III First Strand Synthesis Kit (Invitrogen). Real-time PCR was performed using Power SYBR® Green Master Mix (Applied Biosystems, Foster City, CA, USA) on the Step One™ Real-Time PCR System (Applied Biosystems, FosterCity). The comparative cycle threshold (CT) method was used to evaluate relative quantification. Primers used in qRT-PCR are listed in Table 1. Real-time PCR experiments were repeated three times, and each experiment was performed in triplicate. GAPDH was used as an internal control.

15.Statistical Analysis

SPSS Statistics Version 26 (IBM, Armonk, NY, USA) or GraphPad Prism (GraphPad

Software, San Diego, CA, USA) were used for all statistical analyses. Statistical comparisons of mean values were performed using Student's t-test (two-tailed) and group comparisons were performed using ANOVA. Pearson's correlation coefficient was used to examine the relationship between lncRNAs with the genes of interest. Data are presented as the means ± SD, and all p-values are two-sided. $p < 0.05$ was considered significant.

Table 1. Primer sequences used for the qRT-PCR

Gene Name		Primer Sequence
GAPDH	Forward	GCCGTCTAGAAAAACCTGCC
	Reverse	ACCACCTGGTGCTCAGTGTA
MSC	Forward	AGGACCGCTATGAGAACGGCTA
	Reverse	GTGGTTCCACATAGTCTGTTGGC
AEBP1	Forward	CTACGCACAGAATGAGGTGGTG
	Reverse	CACTCCTCGTTCACCACCTTCA
CREB3L1	Forward	GCCTTGTGCTTTGTTCTGGTGC
	Reverse	CCGTCATCGTAGAATAGGAGGC
COL3A1	Forward	GAAGTCAAGGAGAAAGTGGTCG
	Reverse	ACCTCGTTCTCCATTCTTACCA
LUM	Forward	TGGTCTCCCTGTCTCTCTTCTA
	Reverse	CTATCAGCCAGTTCGTTGTGAG
DCN	Forward	GCTCTCCTACATCCGCATTGCT
	Reverse	GTCCTTTCAGGCTAGCTGCATC
FNDC1	Forward	TGCATCTTGGGATGCGCTACCA
	Reverse	GGCAGAAGTAGTGTCTCCAGGA
THBS2	Forward	CAGTCTGAGCAAGTGTGACACC
	Reverse	TTGCAGAGACGGATGCGTGTGA
COL1A1	Forward	GATTCCCTGGACCTAAAGGTGC
	Reverse	AGCCTCTCCATCTTTGCCAGCA
MMP2	Forward	AGCGAGTGGATGCCGCCTTTAA
	Reverse	CATTCCAGGCATCTGCGATGAG
FN1	Forward	ACAACACCGAGGTGACTGAGAC
	Reverse	GGACACAACGATGCTTCCTGAG
FAP	Forward	GGAAGTGCCTGTTCCAGCAATG
	Reverse	TGTCTGCCAGTCTTCCCTGAAG
LINC01614	Forward	CAGTTGTTTTGGGGCGATCTAG
	Reverse	AGAGAAAGAGCCTATTCCCCAG
LINC00702	Forward	ACTCAATGGGAAATGGCTTG
	Reverse	GTACCACAAGGTTGGCAGGT
LINC02544	Forward	GTTCTCATTCGTGGCTGGAT
	Reverse	TCTGCAAGCACAAAGACAGC
AL356417.2	Forward	GCTACAGTGACCTAAGCTCTAG
	Reverse	GTTTTGGGTTGTCATGAGGGAG
AC112721.2	Forward	GATTTGCACTAGACGCTCTCTC
	Reverse	GGTGCAATAGGAGAGCTTCATG
LINC01929	Forward	GTGTGGTCCTGTTTCAGTCAAA
	Reverse	GAAAAGATGCCCATACCAGACC

Abbreviations: GAPDH, glyceraldehyde-3-phosphate dehydrogenase; MSC, Musculin;

AEBP1, AE binding protein 1; CREB3L1, CAMP responsive element binding protein 3 like 1; COL3A1, Collagen Type III Alpha 1 Chain; LUM, Lumican; DCN, Decorin; FNDC1, Fibronectin Type III Domain Containing 1; THBS2, Thrombospondin-2; COL1A1, Collagen, type I, alpha 1; MMP2, matrix metalloproteinase-2; FN1, Fibronectin 1; FAP, fibroblast activation protein alpha; LINC01614, Long Intergenic Non-Protein Coding RNA 1614; LINC00702, Long Intergenic Non-Protein Coding RNA 702; LINC02544, Long Intergenic Non-Protein Coding RNA 2544; LINC01929, Long Intergenic Non-Protein Coding RNA 1929

III. RESULTS

1. HGS_OvCa is classified into five clusters according to the expression patterns of lncRNAs

First, among the total RNA-seq 60,433 identifiers, lncRNAs and coding RNAs were classified using the GTF file (v30) classification lncRNAs provided by GENCODE. As a result, 15,171 lncRNAs were identified, and among them, the top 1,500 lncRNAs were selected for clustering (Figure 1A). Clustering was performed using the CNMF method, and three verification methods were considered to select an appropriate *k*-factor (Figure 1B). Consensus matrices and sample correlation matrices are shown for $k = 2$ to $k = 6$. Clustering with $k = 5$ gave the most consistent result in both sets (Figure 1B). The cophenetic coefficient shows a consistently high value between $k = 2$ and $k = 6$. Moreover, average silhouette widths suggest optimal results for $k = 2$, $k = 4$, $k = 5$, and $k = 6$. Finally, in the results of total within sum of square, $k = 3$ was suggested as an appropriate result (Figure 1C and 1E). Among the methods for selecting an appropriate cluster, it was confirmed that it was defined based on the most stable k-factor decomposition of the sample by the sample correlation matrix and visual inspection.

Since the purpose of checking through clustering is to check whether each cluster reflects the prognosis and classify clusters related to prognosis, additional k-factor survival analysis was performed (Figure 2A and 2B). As a result of survival analysis, it was confirmed that $k = 5$ showed a significant level compared to other *k*-factors (OS Log Rank p-value = 0.034, DFS Log Rank p-value = 0.023). Based on this, $k = 5$ clustering revealed five distinct and robust clusters with limited overlap between clusters.

Therefore, it was confirmed that HGSC was classified into five clusters according to lncRNA expression (J1 n = 92, J2 n = 66, J3 n = 86, J4 n = 57, J5 n = 66, samples for each cluster) (Figure 2C). The number of lncRNAs corresponding to the J2 cluster was more than twice as high as that of other clusters (J1 cluster n = 284, J2 cluster n = 536, J3 cluster

n= 297, J4 cluster n = 212, J5 cluster n = 171, number of lncRNAs in each cluster). As a result of checking the mRNA expression based on the cluster classified according to the confirmed lncRNA expression, it was confirmed that the expression pattern was similar to that of the lncRNA expression pattern (Figure 2D). Like the number of lncRNAs, there were twice as many mRNAs in the J2 cluster as compared to in other clusters (J1 cluster n = 331, J2 cluster n = 577, J3 cluster n = 208, J4 cluster n = 201, J5 cluster n = 183, number of mRNAs in each cluster).

HGS_OvCa was classified into five clusters according to lncRNA expression, and each sample reflects prognosis according to the classified cluster. The J2 cluster is a cluster that contains a lot of lncRNAs, and mRNAs as confirmed above, but it is confirmed that it does not affect prognosis. On the other hand, the cluster with the worst prognosis was confirmed to correspond to the J4 cluster.

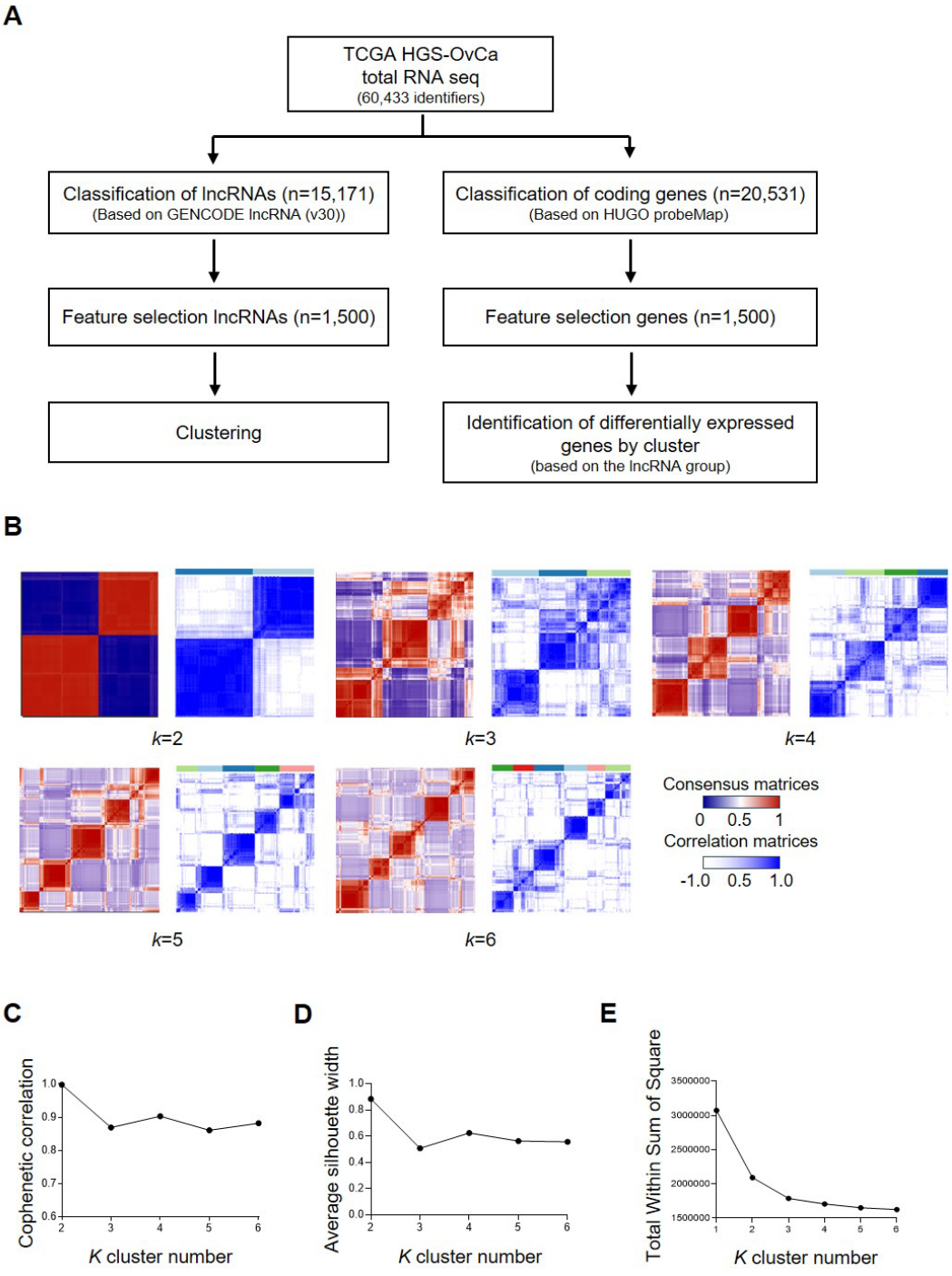


Figure 1. CNMF clustering of 1,500 variably expressed genes and 367 HGS_OvCa samples. (A) Clustering analysis strategy according to lncRNA expression pattern using HGS-OvCa total RNA-seq data. **(B)** Consensus matrices (left panel) and correlation matrices (right panel) are shown for clustering with k=2 to k=6. Cluster fitness evaluation items according to k-factor: **(C)** Cophenetic correlation, **(D)** Average silhouette width, **(E)** Total Within Sum of Square.

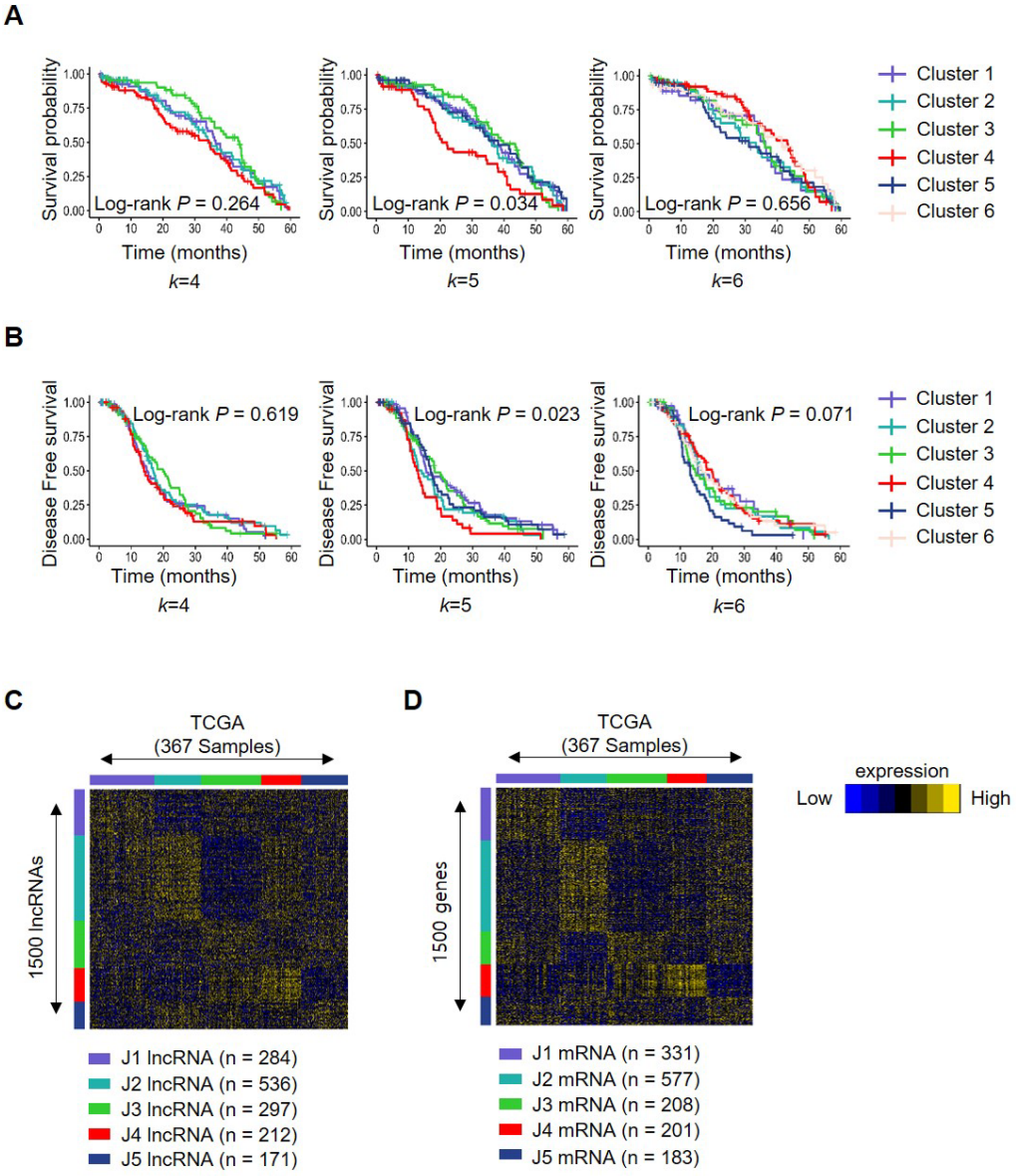


Figure 2. Selection of *k*-factors suitable for lncRNA expression patterns according to survival analysis results. (A) Survival probability and **(B)** Disease Free survival curve according to *k*-factor. **(C)** lncRNA, **(D)** mRNA expression patterns of HGS-OvCa 367 samples according to selected *k*-factors.

2. The five clusters of HGS_OvCa are classified according to function

Functional analysis was performed to confirm the function of each of the five clusters identified using the lncRNA expression pattern and prognosis. As a result of selecting lncRNAs with increased expression by cluster and classifying them using the funcpred analysis tool, it was confirmed that all lncRNAs corresponding to the five clusters were related to SPERMATOGENESIS in common (Figure 3A and 3E). Hallmark's spermatogenesis-related genes are included in development categories such as angiogenesis and EMT, and in previous studies, Notch related lncRNAs in ovarian cancer were shown to have a function in spermatogenesis. As a result of confirming the specific function of each cluster except for spermatogenesis-related gene sets identified in all clusters, the J1 cluster was related to Immune, J2 cluster was related to EMT, and J3 cluster was related to Estrogen response (Figure 3A and 3C). In the J4 cluster, APICAL JUNCTION related to EMT and cellular components, ADIPOGENESIS and MYOGENESIS related to development, and HYPOXIA and ANDROGEN RESPONSE related to signaling were identified (Figure 3D). It was confirmed that the difference between J2 and J4 clusters was related to the same EMT gene, but the J4 cluster was related to genes reflecting multiple cancer cell aggressiveness besides EMT. Bile acid metabolism and peroxisomes in the J5 cluster are known to affect cancer cell growth and differentiation (Figure 3E)[67].

The direct result of lncRNA function analysis has not been known so far, and funcpred is also matched with a coding gene to guess its function in an indirect way. Based on the previous analysis that the expression patterns of lncRNAs and mRNAs were similar for each cluster, the increased mRNA expression for each cluster was reconfirmed in the same way as in the lncRNA function analysis. Increased mRNA expression for each cluster was selected and analyzed using gprofiler, and p-value < 0.05 and q-value < 0.25 were judged as a significant result. Most of them were confirmed to be consistent with the results of analyzing lncRNA function previously (Figure 4A and 4E). In the J1 cluster, gene sets of

INTERFERON GAMMA RESPONSE, INTERFERON ALPHA RESPONSE, ALLOGRAFT REJECTION, and INFLAMMATORY RESPONSE related to the immune response were identified, just as the lncRNA function was confirmed (Figure 4A). In addition to the EMT gene set, the J2 cluster identified a gene set with increased oncogene KRAS signaling genes (Figure 4B)[68,69]. The J3 cluster contains a set of estrogen-responsive genes and a set of immune-related genes, and the J4 cluster identified the same set of genes in lncRNA functional analysis (Figure 4C and 4D). In the J5 cluster, a gene set different from the gene set identified in the results of lncRNA functional analysis was identified (Figure 4E).

By combining the two results, the functions identified in common with lncRNA and mRNA, lncRNA specific functions, and lncRNA common gene sets were classified. As a result, J1 cluster could be classified as the Immune group (Figure 5A), J2 cluster as the EMT group (Figure 5B), J3 cluster as the Estrogen response group (Figure 5C), and J4 cluster as the EMT-Androgen response group (Figure 5D). The common gene set of lncRNA or mRNA in the J5 cluster could be identified (Figure 5E). As a result of comparison with the cluster sample of the previous study, it was confirmed that the J5 cluster was about 51% consistent with the sample of the differentiation cluster (Figure 6A), and the differentiation marker expression was also increased (Figure 6B and 6C). Therefore, the J5 cluster was classified as a Differentiation group.

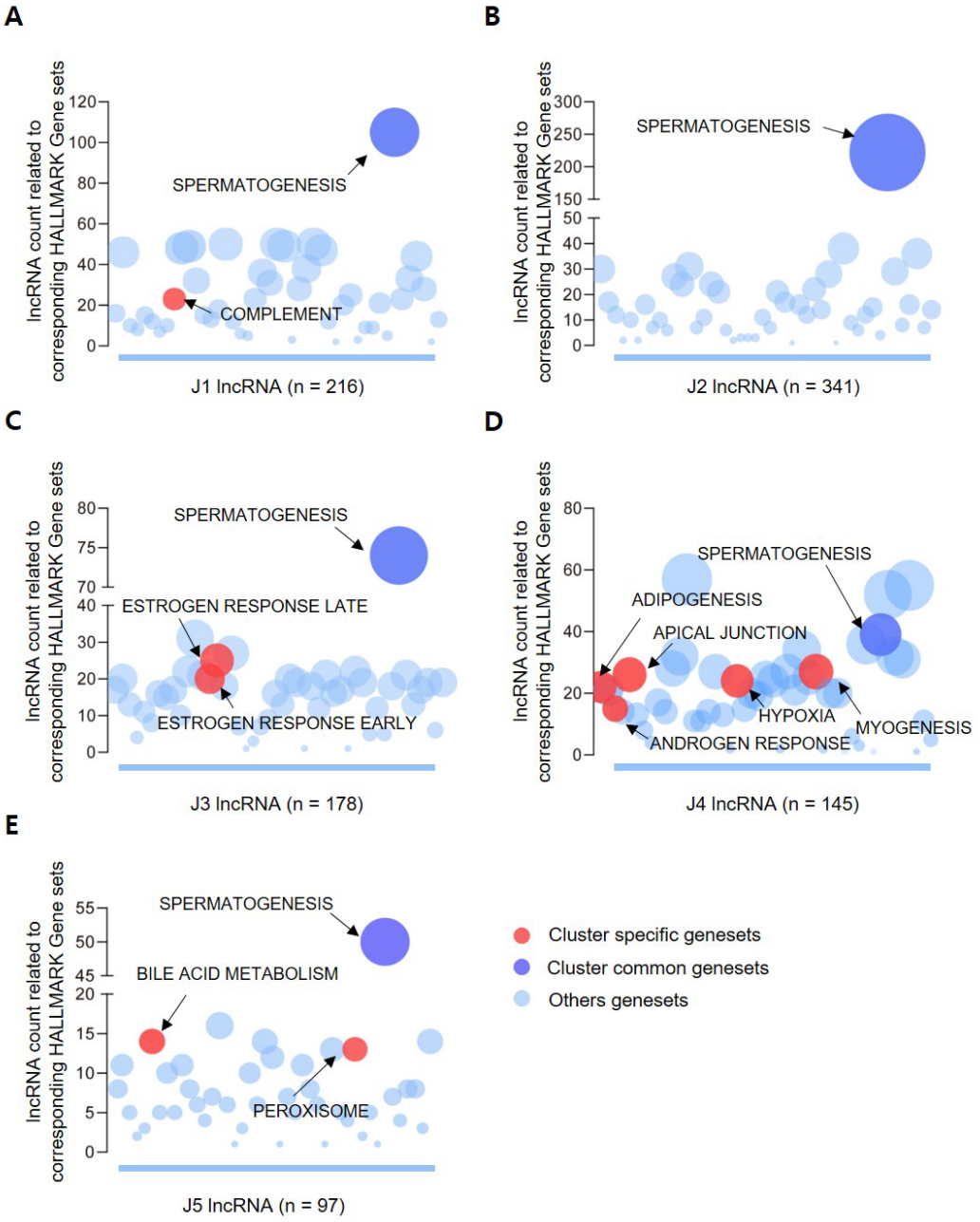


Figure 3. Functional analysis of lncRNAs in each cluster. (A and E) Functional analysis of lncRNAs in each lncRNA cluster using the Funcpred hallmark geneset. The size of the circle represents the number of lncRNAs associated with the corresponding HALLMARK gene set. Red circles represent genesets identified only in the corresponding cluster, and purple circles represent genesets identified in all clusters, and the light blue circle represents the rest of the geneset.

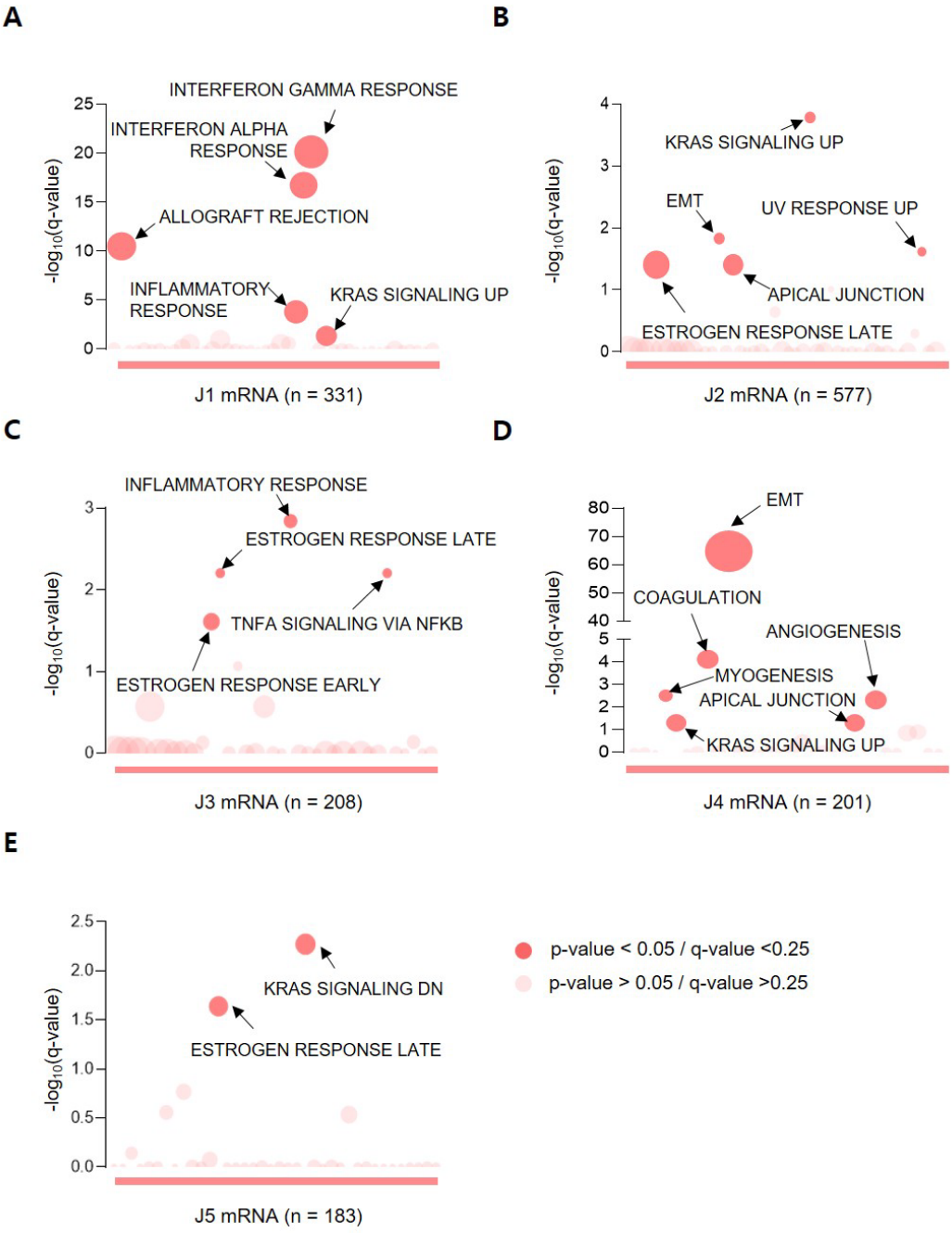


Figure 4. Functional analysis of mRNA in each cluster. (A and E) Functional analysis of mRNA included in each cluster using hallmark geneset with gprofiler. The size of the circle represents the q-value, and the salmon-colored circle represents the set of genes with p-value < 0.05 and q-value < 0.25. Light pink indicates gene sets with p-value > 0.05 and q-value > 0.25.

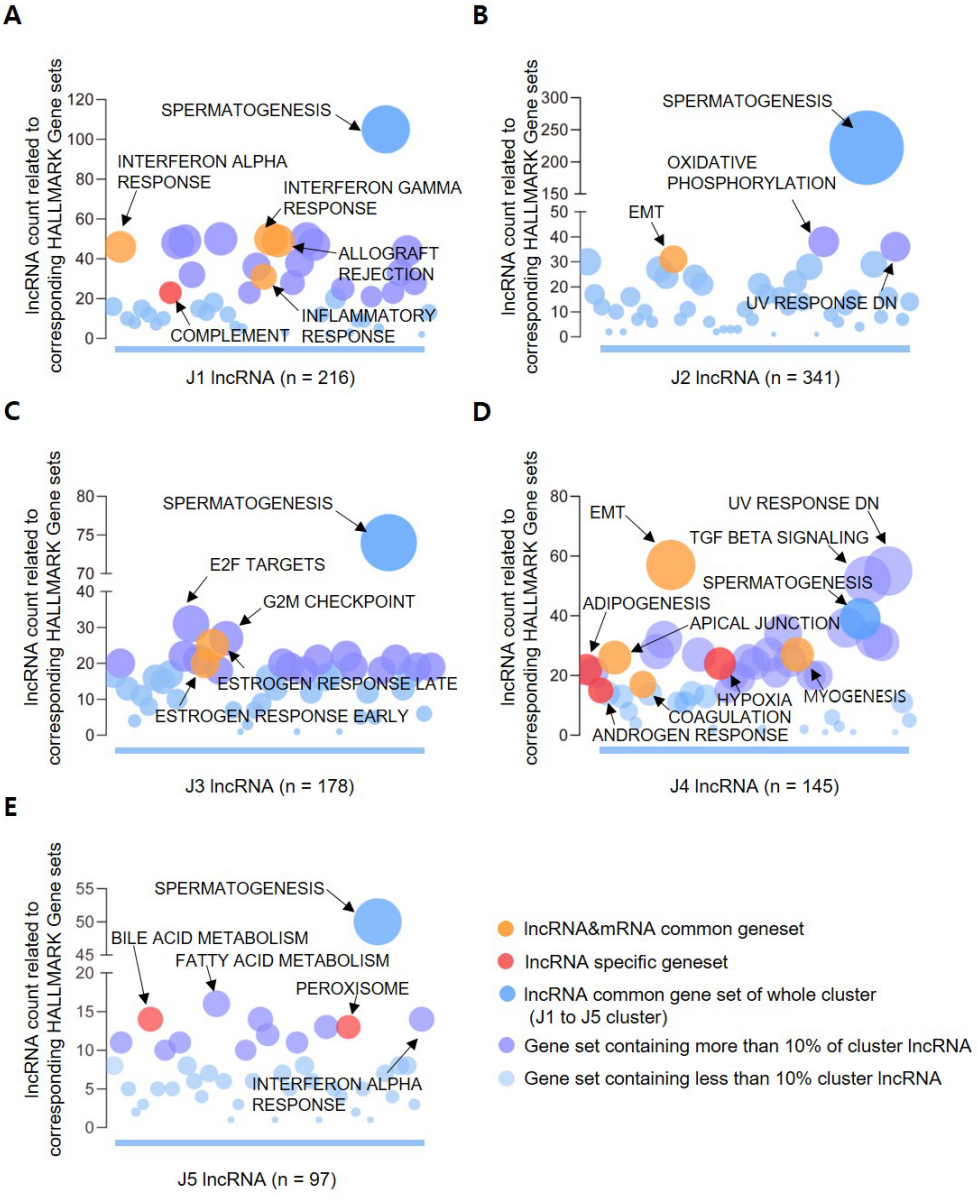


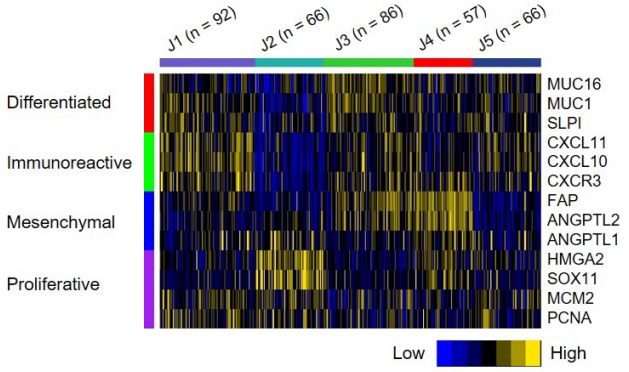
Figure 5. Analysis of lncRNA and mRNA functions according to each cluster. (A and E) Identification of specific gene sets and common gene sets by merging the functional analysis results of lncRNA and mRNA corresponding to each cluster. Orange circles represent sets of genes commonly identified in lncRNA and mRNA function analyses, and red circles represent specific gene sets identified in lncRNA functional analyses. Blue circles indicate lncRNA common gene sets corresponding to the J1 and J5 clusters, and purple circles indicate gene sets containing 10% of the lncRNAs in each cluster. Light blue circles indicate gene sets containing less than 10% of the lncRNAs corresponding to each cluster.

A

TCGA cluster

	Differentiated	Immunoreactive	Mesenchymal	Proliferative
J1 cluster	24	39	10	19
J2 cluster	6	2	4	54
J3 cluster	34	23	25	4
J4 cluster	0	4	46	7
J5 cluster	34	12	1	19

B



C

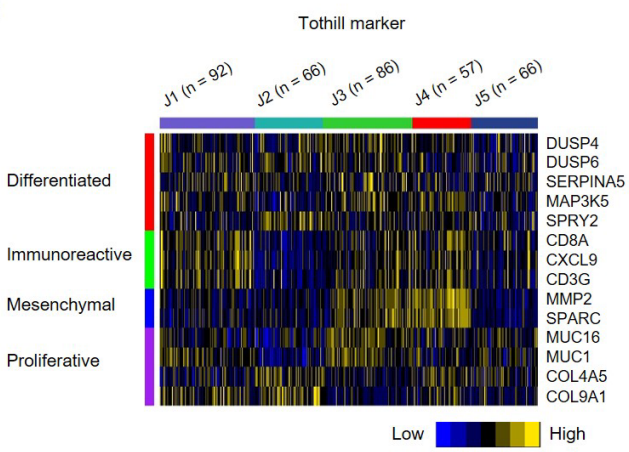


Figure 6. J5 cluster involved in differentiation of HGS_OvCa. (A) Sample matching between clusters classified according to lncRNA expression and clusters presented by the TCGA group. Confirmation of TCGA marker **(B)** and Tothill marker **(C)** expression in clusters classified according to lncRNA expression.

3. There are no significant differences in the mutant genes for each HGS_OvCa cluster

Regarding the characteristics of the classified clusters, mutation profiles were first identified for each cluster. There was no difference between the results of confirming mutations in the entire HGSC sample without cluster classification and the results of confirming mutations by classifying each cluster (Figure 7A and 7F). Missense mutations were the most frequently observed mutations in all clusters, and TP53 mutations accounted for more than 80% of each cluster sample.

A dualistic model of carcinogenesis was proposed for EOC based on histopathological and molecular characteristics. This model broadly categorizes epithelial ovarian tumors into two groups: Type I tumors in EOC are characterized by relative chromosomal stability and specific mutations in genes such as KRAS, BRAF, ERBB2, CTNNB1, PTEN, PIK3CA, ARID1A, and PPPR1A. These tumors rarely exhibit mutations in TP53, which encodes the tumor suppressor p53. Type II tumors are much more frequent and aggressive than Type I, tend to grow faster, are present at an advanced stage, and have a very high frequency of TP53 mutations. Titin (TTN), a mutation gene that occupies the second highest ratio in all clusters along with TP53, is known to be most frequently mutated in all cancers.

BRCA1 mutations have been identified in the estrogen response cluster, and BRCA1,2 mutations have been found to be present in approximately 5-15% of ovarian cancers (Figure 3D). Therefore, BRCA1 and BRCA2 mutations are relatively rare and account for only a subset of all ovarian cancers. However, it is considered a sign for hope, because it means that personalized treatment is possible if there is a mutation in the BRCA1 and BRCA2 genes. Recently, PARP inhibitors have been shown to be effective in reducing tumor burden in patients with BRCA1 and/or BRCA2 mutations[70,71]. Therefore, it was confirmed that there was no difference in mutation profiles in the remaining clusters except for the estrogen response cluster, and that there was no characteristic distinguishing each cluster.

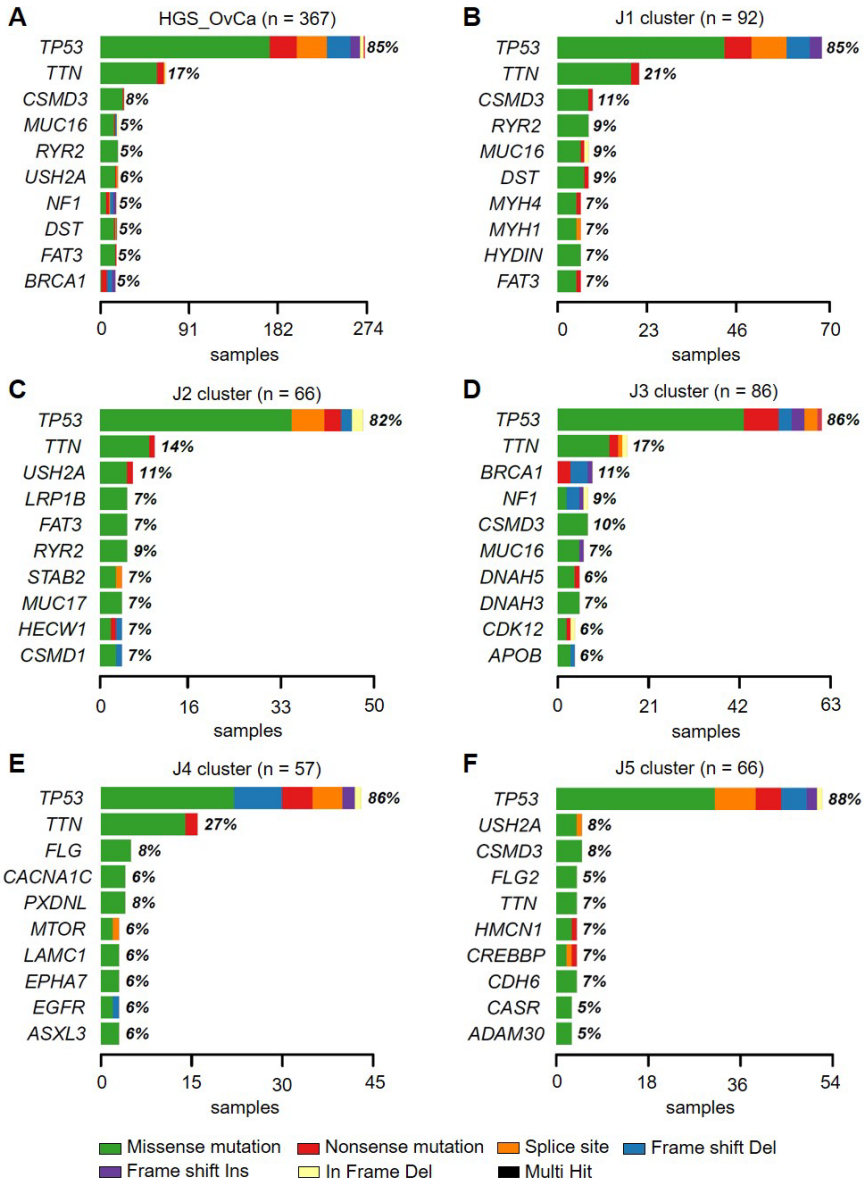
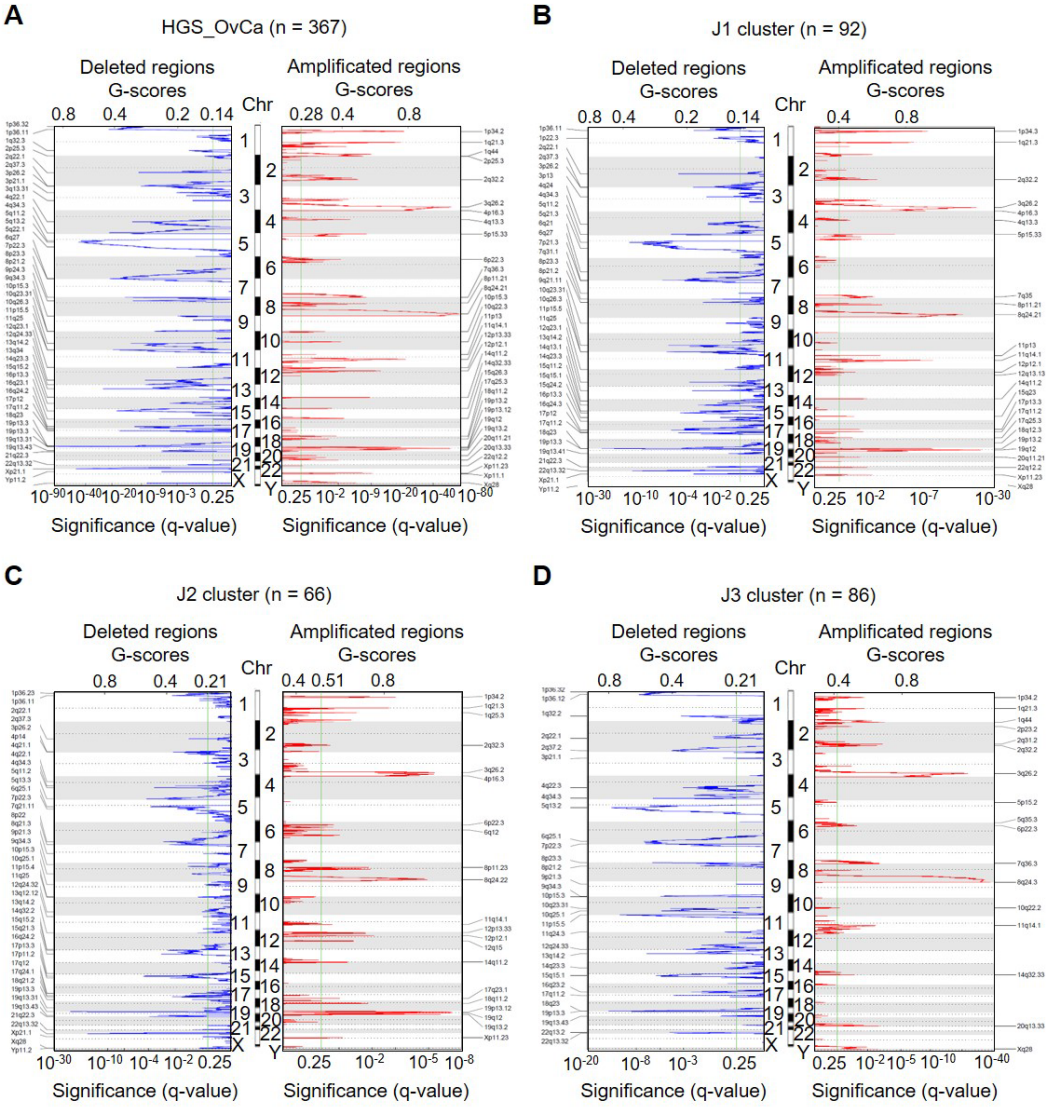


Figure 7. Identification of significantly mutated genes according to each lncRNA cluster. Top 10 mutant genes corresponding to the **(A)** total HGS_OvCa sample and **(B and F)** each cluster sample. Graphs were ordered by the total number of variants in each gene, and the percentages represent the proportion of tumor samples with genetic variants relative to the total sample. Colored squares represent mutant genes.

4. Functional differences between the HGS_OvCa clusters are not due to changes in Somatic Copy Number Alteration

Somatic Copy Number Alteration (SCNA) plays an important role in activating oncogenes and inactivating tumor suppressors, and understanding of the biological and phenotypic effects of SCNAs has resulted in significant advances in cancer diagnosis and treatment. Recently, TCGA identified changes in SCNA in the context of a large genome-wide integrated study of approximately 500 high-grade serous ovarian cancers (HGSOC). Among them, the oncogenic lncRNA FAL1 with amplified gene copy number shows increased expression levels in many types of cancer. FAL1 expression is associated with ovarian cancer outcome and interacts with the PRC1 component BMI1 to repress numerous genes including CDKN1A.

Genomic Indentification of Significant Targets In Cancer (GISTIC2.0) was used to identify focal CNA regions and potential drive genes. As a result of checking the total HGSC copy number, the most frequent gain regions included chromosomal regions 8q24.21, 3q26.3, 19q12, 11q14.1, and 1p34.2, and the most frequent loss regions included 19p13,3, 22q13.32, 13q14.2, 6q27, and 5q11.2 (Figure 7A). Similarly, as a result of checking the copy number of each cluster, the regions where the most gain and the most loss occurred were not significantly different from the results of checking the copy number of the entire HGSC (Figure 7B and 7F). When comparing the top 10 somatic copy number alteration amplification and deletion chromosomal regions in the entire HGS_OvCa sample and each cluster sample, similar results were confirmed without any difference (Table 2 and Table 3). Through the analysis results, it was confirmed that the difference between the classified clusters was not due to SCNA.



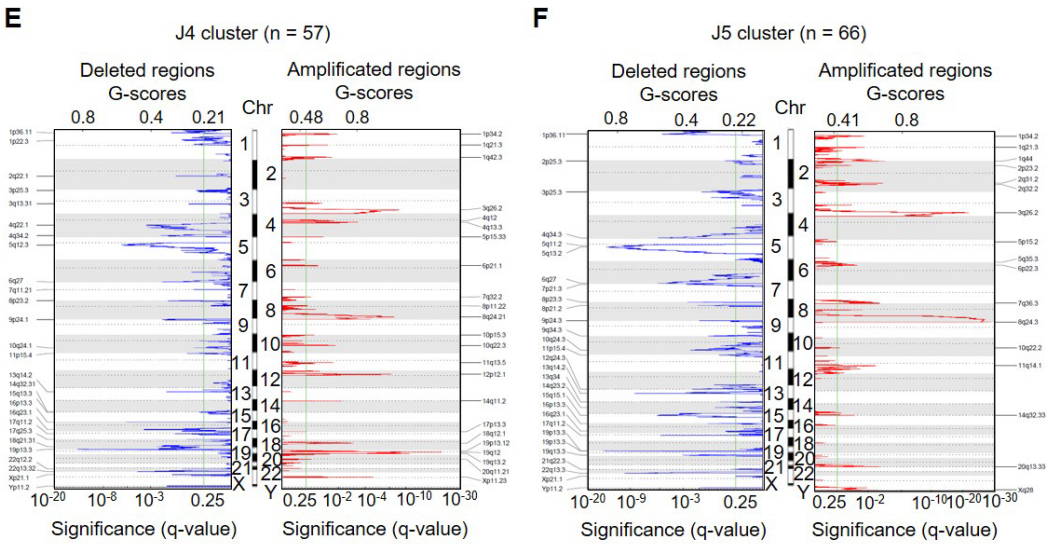


Figure 8. Identification of somatic copy number alternation in each lncRNA cluster. Genomic Identification of Significant Targets in Cancer (GISTIC) amplification (right) and deletion (left) plots for **(A)** total HGS_OvCa samples and **(B and F)** each cluster sample. Gains and losses are shown in red and blue, respectively, and are ordered according to genomic location. The significance threshold (q-value < 0.25) is indicated at the bottom of the plot.

Table 2. Top 10 chromosomal regions with somatic copy number alteration amplifications in each cluster

HGS_OvCa (n = 367)	J1 cluster (n = 92)	J2 cluster (n = 66)	J3 cluster (n = 86)	J4 cluster (n = 57)	J5 cluster (n = 66)
8q24.21	19q12	19p13.12	8q24.3	19q12	8q24.21
3q26.2	3q26.2	3q26.2	3q26.2	3q26.2	19q12
19q12	8q24.21	19q12	7q36.3	8q24.21	11q14.1
11q14.1	19p13.2	8q24.22	20q13.33	12p12.1	3q26.2
1p34.2	11q14.1	19q13.2	Xq28	Xp11.23	1p34.2
1q21.3	1p34.3	1p34.2	11q14.1	19q13.2	4q13.3
12p12.1	1q21.3	4p16.3	1p34.2	19p13.12	7q36.1
5p15.33	5p15.33	1q21.3	1q44	18q12.1	1q44
7q36.3	11p13	8p11.23	6p22.3	14q11.2	15q26.3
19p13.2	2q32.2	18q11.2	14q32.33	1p34.2	14q11.2

Table 3. Top 10 chromosomal regions with somatic copy number alteration deletions in each cluster

HGS_OvCa (n = 367)	J1 cluster (n = 92)	J2 cluster (n = 66)	J3 cluster (n = 86)	J4 cluster (n = 57)	J5 cluster (n = 66)
19p13.3	19p13.3	19p13.3	19p13.3	19p13.3	22q13.32
22q13.32	22q13.32	22q13.32	5q13.2	5q12.3	5q13.2
13q14.2	5q11.2	15q15.2	11p15.5	16p13.3	13q14.2
6q27	11p15.5	4q34.3	18q23	22q13.32	19p13.3
5q11.2	16q24.3	4q22.1	6q25.1	16q23.1	15q15.1
11p15.5	10q23.31	6q25.1	13q14.2	4q22.1	6q27
15q15.2	6q27	5q11.2	4q34.3	18q21.31	4q34.3
4q34.3	17q11.2	19q13.43	15q15.1	6q27	17q11.2
17q11.2	18q23	7p22.3	8p23.3	9p24.1	1p36.11
5q13.2	Yp11.2	13q14.2	9q34.3	13q14.2	19p13.3

5. miRNA expression and DNA methylation do not explain the differences observed between HGS_OvCa clusters

Previous analyses failed to identify the cause of functional differences between HGSC clusters. As an additional analysis to demonstrate functional differences between the clusters, I examined the relationship between miRNA expression patterns and DNA methylation. LncRNAs are involved in pattern regulation of expressed proteins through specific mechanisms, including a variety of biological interactions, such as lncRNA-ncRNA, lncRNA-mRNA, and lncRNA-protein interactions. Therefore, the construction of an inferred biological interaction network mediated by lncRNAs should be desirable to uncover the potential mechanisms and biological functions of lncRNAs. In other words, miRNAs have a significant influence on the molecular mechanisms of lncRNAs. Numerous studies have demonstrated that miRNAs and lncRNAs are involved in pathological processes, including a variety of diseases, and the regulatory role of miRNA-lncRNA interactions in some human complex diseases has been systematically investigated.

Therefore, it was additionally confirmed whether miRNA expression was different depending on the lncRNA group. When the expression patterns of the 486 miRNAs selected based on the classified lncRNA clusters were visually inspected, it was confirmed that there was no difference between the clusters (Figure 9A). In order to quantify and confirm that there is no difference, the median value of miRNA expression corresponding to each cluster was confirmed (J1 = 5.84±3.83, J2 = 5.93±3.81, J3 = 5.80±3.88, J4 = 5.87±3.86, J5 = 5.83±3.86), and as a result, it was reconfirmed that there was no difference between each cluster (Figure 9B).

DNA methylation and carcinogenesis are interrelated. The best known mechanism through which DNA methylation affects carcinogenesis is the silencing of tumor suppressors through hypermethylation, which is apparently localized to promoters and other regulatory regions due to increased DNA methyltransferase levels. DNA methylation has been shown to play an important role in OvCa, and several tumor suppressor genes

have been shown to be hypermethylated. Methylation of the BRCA1 promoter has received a lot of attention because it is known that BRCA1 mutations are involved in the inherited OvCa.

Based on the results of previous studies, I used TCGA data to analyze DNA methylation between clusters. The methylation data are quantified and provided as a β-values (methylation ratio, proportion of methylated/unmethylated), with a total number of probes of 24,956. As a result of checking the distribution of methylation β-values for each group, the overall pattern was similar (Figure 9C). The results were reconfirmed by quantifying whether there was no difference for each cluster by combining the β-values of each cluster (Figure 9D). As a result of the analysis, it was confirmed that the β-values of each cluster were different, and that the β-values of the J3 cluster were increased compared to other clusters (J1 = 0.23±0.30, J2 = 0.23±0.31, J3 = 0.25±0.32, J4 = 0.23±0.30, J5 = 0.24±0.31). In general, a β-value < 0.2 is considered hypomethylation, a β-value > 0.8 is considered hypermethylation, and a β-value of 0.5 is considered partially methylated. The average of the beta values in our data set is approximately 0.238. Therefore, if the β-value is expressed as a heat map, it can be confirmed that most of them are hypomethylated. Based on the above results, it was determined that DNA methylation did not affect HGSC cluster functional characteristics.

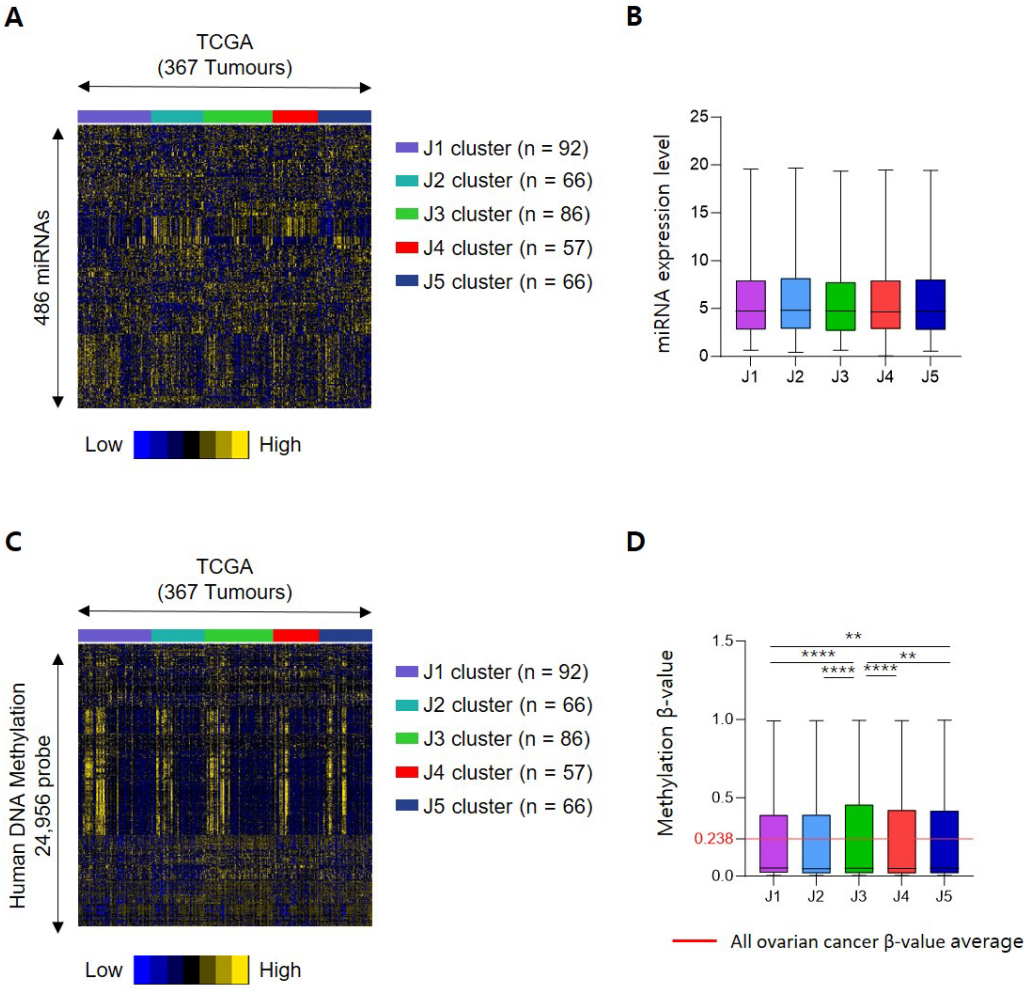


Figure 9. Confirmation of miRNA and DNA methylation according to lncRNA clusters. **(A)** Expression patterns and **(B)** expression levels of 486 miRNAs according to lncRNA clusters. **(C)** β-value patterns and **(D)** β-value levels of human DNA methylation 24,956 probes according to lncRNA clusters. Data is shown as mean±SD and p-values were calculated by ANOVA. ** $p < 0.01$, **** $p < 0.0001$. The results of Not Significance were not separately indicated.

6. HGS_OvCa clusters are regulated by different transcription factors specific to each cluster

Several factors potentially affecting cluster functional differences were examined, but none of them were found to be relevant. Therefore, additional investigations were conducted to identify potential factors that could indicate functional differences among the clusters. Recent studies have revealed that a transcription factor or group of transcription factors play important roles in biological processes and disease progression. Therefore, I investigated the characteristic transcription factors of each cluster.

Before starting the analysis to identify transcription factors, a data screening process was performed. I attempted to identify transcription factors within clusters classified according to lncRNA expression, but so far, lncRNAs have not been included in the list of transcription factor target genes. Therefore, I selected mRNAs that had a positive correlation with lncRNAs and confirmed that most of them covered half of the cluster (Figure 10A and 10B). The number of lncRNAs and mRNAs corresponding to each cluster was more than twice as high in the J2 cluster as in the other clusters, but the cluster containing many mRNAs with a significant positive correlation with lncRNAs was identified as the J4 cluster. Based on the above results, I speculated that the cluster with strong interactions between lncRNAs and mRNAs would be the J4 cluster.

As a result of transcription factor inference analysis using these data, it was possible to identify unique transcription factors for each cluster (Figure 11A and 11F, Table 4 and Table 8). The J1 cluster identified three transcription factors based on a p-value < 0.05. CDKN2A is known to affect EMT, immune reactivity and immune cell infiltration; affect poor prognosis of hepatocellular carcinoma; and is associated with metastasis of colorectal cancer. Similarly, ETV7 and IFI27 are also related to immune modulation. The analysis results are consistent with the function of the J1 cluster identified above (Figure 11A and Table 4). A total of 21 transcription factors were identified in the J2 cluster, and among the corresponding transcription factors, DACH1 is known to promote distant metastasis in

ovarian cancer. Similarly, identified transcription factors are known to promote EMT or accelerate the onset of ovarian cancer (Figure 11B and Table 5). The J3 cluster was classified into a group related to estrogen response based on the previous functional analysis results. According to the result, the transcription factors RUNX1, TRAP2C, and TRPS1 identified in the J3 cluster are related to estrogen response, and RUNX1 is increased in ovarian cancer and is known to regulate ER-mediated genes. TFAP2C is also known to regulate ER-mediated genes, including RUNX1, and TRPS1 is known to regulate epithelial marker expression in ER-positive cancer (Figure 11C and Table 6). RUNX2, a transcription factors corresponding to the J4 cluster with the worst prognosis, is associated with tumor invasion and metastasis in ovarian cancer and is known to promote metastasis by regulating EMT in breast cancer. ZFHX4 is also known to show poor prognosis in ovarian cancer, and it also includes SNAI2 and TWIST1,2 related to EMT (Figure 11D and Table 7). The transcription factor EHF identified in the J5 cluster is involved in epithelial differentiation and carcinogenesis (Figure 11F and Table 8).

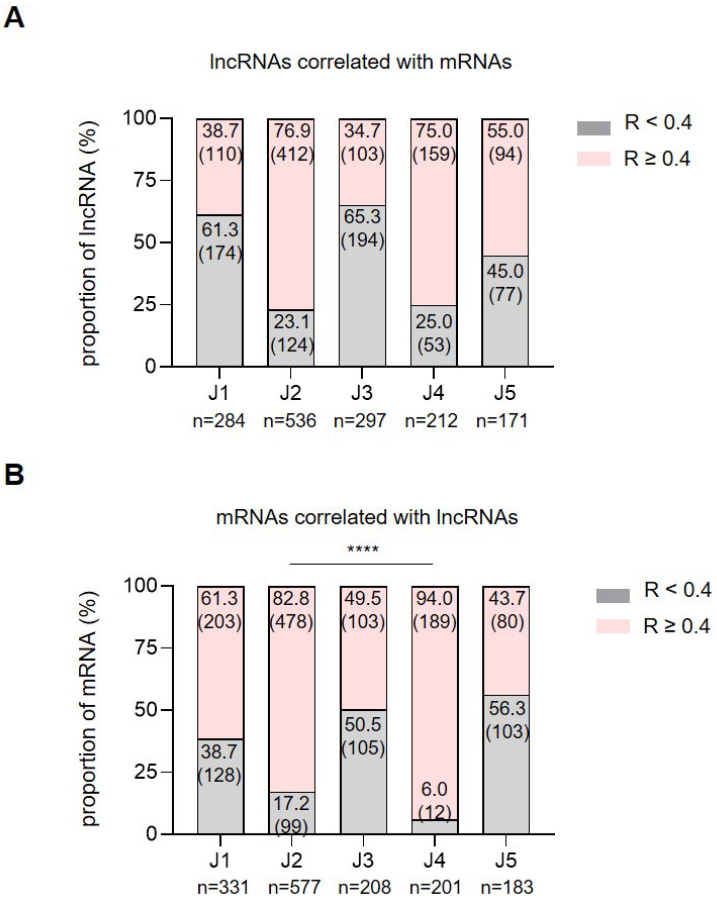


Figure 10. Identification of positively correlated lncRNAs and mRNAs between lncRNA clusters. (A) Proportion of mRNA positively correlated with lncRNA for each cluster. **(B)** Proportion of lncRNAs positively correlated with mRNAs for each cluster. Data are presented as respective ratios (number of samples). Positive correlations based on Pearson's correlation R $\geq$ 0.4 are shown in pink, others are shown in gray. p-values calculated by linear-by-linear association. **** $p < 0.001$. The results of Not Significance were not separately indicated.

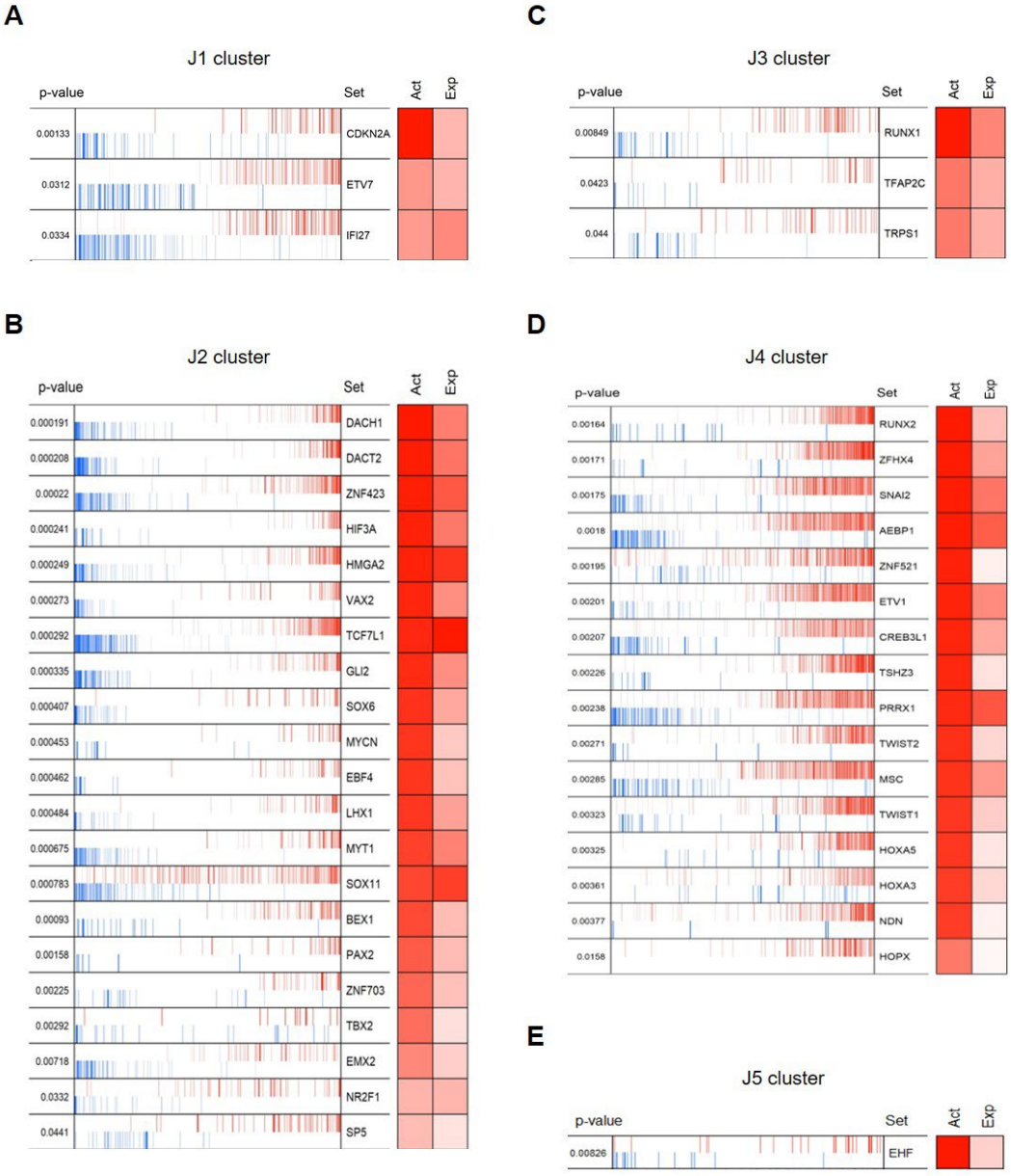


Figure 11. Transcription factors regulating each lncRNA cluster in HGS_OvCa. (A and E) VIPER plot showing the projection of the negative (repressed, shown in blue color) and positive (activated, shown in red color) targets for each TF, as inferred by ARACNe and correlation analysis when reverse engineering the regulatory network (vertical lines resembling a barcode). The "Act" and "Exp" columns show the normalized enrichment score (NES) of the expected activity of the entire TF network and the expression level of the TF itself. p-values were determined using the enrichment method in the VIPER algorithm. All data were collected according to the p-value < 0.05 criterion.

Table 4. Identification of J1 cluster transcription factors using Viper

J1 cluster (n = 7)		
Transcription factor	NES	p-value
CDKN2A	3.209	0.001
ETV7	2.155	0.031
IFI27	2.127	0.033
MESP1	1.632	0.103
HOXD1	1.581	0.114
OASL	1.353	0.176
HEY2	0.656	0.512

Abbreviations: NES, Normalized Enrichment Score

Table 5. Identification of J2 cluster transcription factors using Viper

J2 cluster (n = 37)		
Transcription factor	NES	p-value
DACH1	3.730	<0.0001
DACT2	3.709	<0.0001
ZNF423	3.695	<0.0001
HIF3A	3.672	<0.0001
HMGA2	3.663	<0.0001
VAX2	3.639	<0.0001
TCF7L1	3.622	<0.0001
GLI2	3.587	<0.0001
SOX6	3.535	<0.0001
MYCN	3.507	<0.0001
EBF4	3.502	<0.0001
LHX1	3.490	<0.0001
MYT1	3.399	0.001
SOX11	3.359	0.001
BEX1	3.311	0.001
PAX2	3.159	0.002
ZNF703	3.055	0.002
TBX2	2.977	0.003
EMX2	2.689	0.007
NR2F1	2.130	0.033
SP5	2.013	0.044
MLXIPL	1.626	0.104
PEG3	1.570	0.116
RXRG	1.504	0.133
NPAS3	1.447	0.148
GATA4	1.327	0.184
NR5A1	1.245	0.213

FOXL2	1.227	0.220
MDFI	1.116	0.264
ID4	1.062	0.288
SATB1	1.025	0.305
GREB1	0.975	0.330
PLAGL1	0.934	0.350
FOXJ1	0.865	0.387
PGR	0.842	0.400
MYCL	0.399	0.690
HOXB8	0.154	0.878

Abbreviations: NES, Normalized Enrichment Score

Table 6. Identification of J3 cluster transcription factors using Viper

J3 cluster (n = 7)		
Transcription factor	NES	p-value
RUNX1	2.632	0.008
TFAP2C	2.031	0.042
TRPS1	2.014	0.044
CIITA	1.192	0.233
RUNX3	0.752	0.452

Abbreviations: NES, Normalized Enrichment Score

Table 7. Identification of J4 cluster transcription factors using Viper

J4 cluster (n = 17)		
Transcription factor	NES	p-value
RUNX2	3.730	<0.0001
ZFHX4	3.709	<0.0001
SNAI2	3.695	<0.0001
AEBP1	3.672	<0.0001
ZNF521	3.663	<0.0001
ETV1	3.639	<0.0001
CREB3L1	3.622	<0.0001
TSHZ3	3.587	<0.0001
PRRX1	3.535	<0.0001
TWIST2	3.507	<0.0001
MSC	3.502	<0.0001
TWIST1	3.490	<0.0001
HOXA5	3.399	0.001
HOXA3	3.359	0.001

NDN	3.311	0.001
HOPX	3.159	0.002
ZFPM2	3.055	0.002

Abbreviations: NES, Normalized Enrichment Score

Table 8. Identification of J5 cluster transcription factors using Viper

J5 cluster (n = 2)		
Transcription factor	NES	p-value
EHF	2.641	0.008
ZBED2	1.480	0.139

Abbreviations: NES, Normalized Enrichment Score

7. The J4 cluster has stronger transcription factor activity than the other clusters

It is difficult to determine that a transcription factor is activated based on its own expression. According to recent studies, it is considered that transcription factors are activated according to the target gene expression of transcription factors rather than transcription factors themselves. Therefore, based on the results of analyzing the master transcription factor, the correlation between the target gene of the transcription factor and the selected mRNA and lncRNA was analyzed.

After examining the proportion of mRNAs with a positive correlation to transcription factor target genes in each cluster, I found that approximately 27.9% of mRNAs were included in the J1 cluster, 11.5% in the J2 cluster, 17.8% in the J3 cluster, 69.2% in the J4 cluster, and 7.5% in the J5 cluster. Among them, in J2 and J4, it was confirmed that there was a difference between the groups with p-value < 0.0001 (Figure 12A and 12F). Using the same method, I confirmed that in the J1 cluster, 46.0% of lncRNAs had a correlation with transcription factor target genes, in the J2 cluster, 30.2% had a correlation, in the J3 cluster, 18.8% had a correlation, in the J4 cluster, 88.7% had a correlation, and in the J5 cluster, 7.4% had a correlation with transcription factor target genes. Similarly, it was confirmed that there was a significant difference (p-value < 0.0001) between groups in the proportion of lncRNAs correlated with transcription factors in the J2 and J4 clusters (Figure 13A and 13F). Based on the results of correlation analysis with transcription factor target genes, it was judged that the J4 cluster had transcription factors activated, even though the lncRNA and mRNA levels corresponding to the J2 cluster were more than twice as high as those of other clusters.

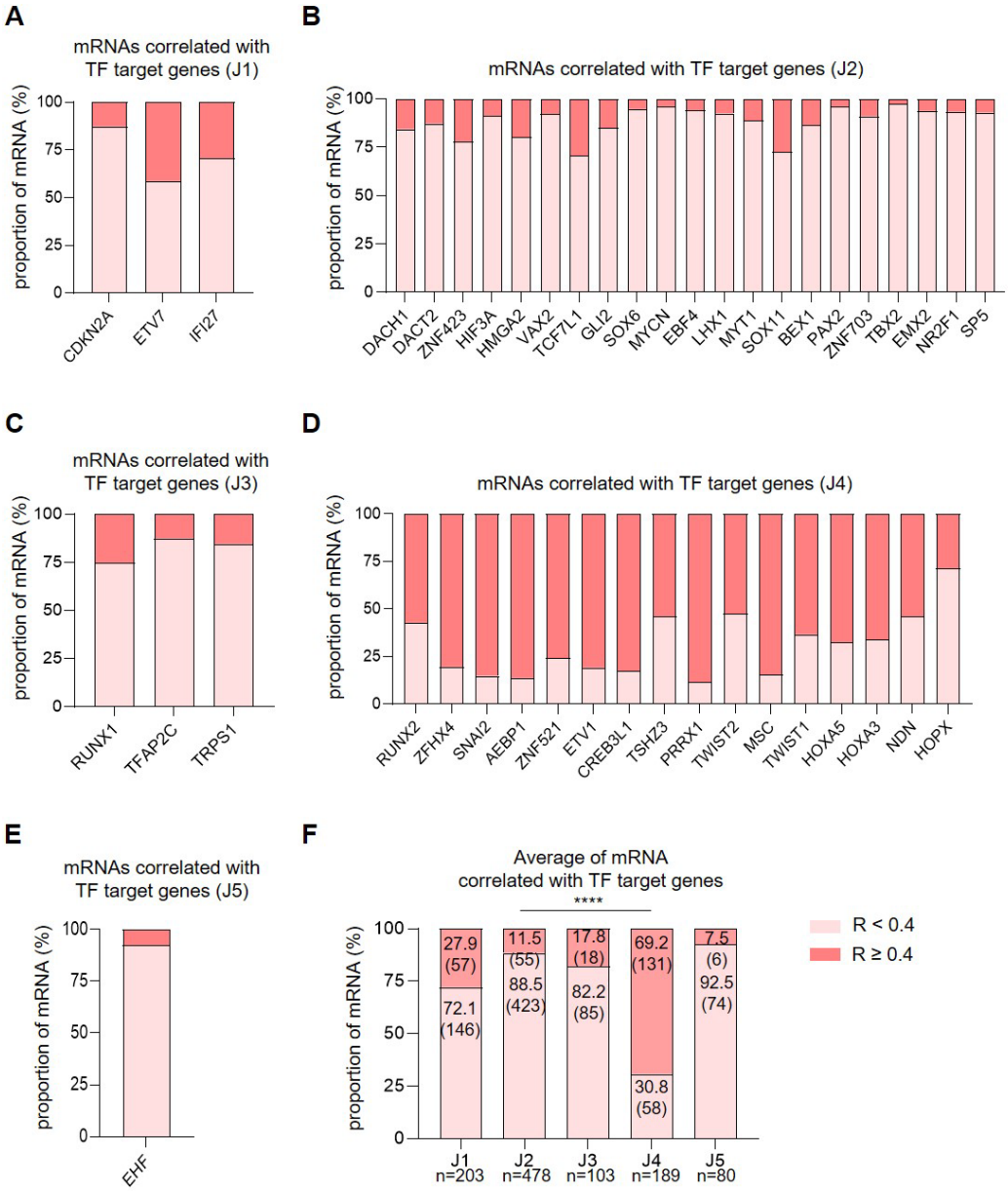


Figure 12. Correlation between target genes of transcription factors and mRNA. (A and E) Proportion of Transcription factor target genes positively correlated with mRNA for each cluster. **(F)** The ratio of mRNA that are positively correlated with the transcription factor target genes of each cluster is shown as an average value. Positive correlations based on Pearson's correlation $R \geq 0.4$ are shown in salmon and others are shown in pink. Data are presented as respective ratios (number of samples). p-values calculated by linear-by-linear association. **** $p < 0.001$. The results of Not Significance were not separately indicated.

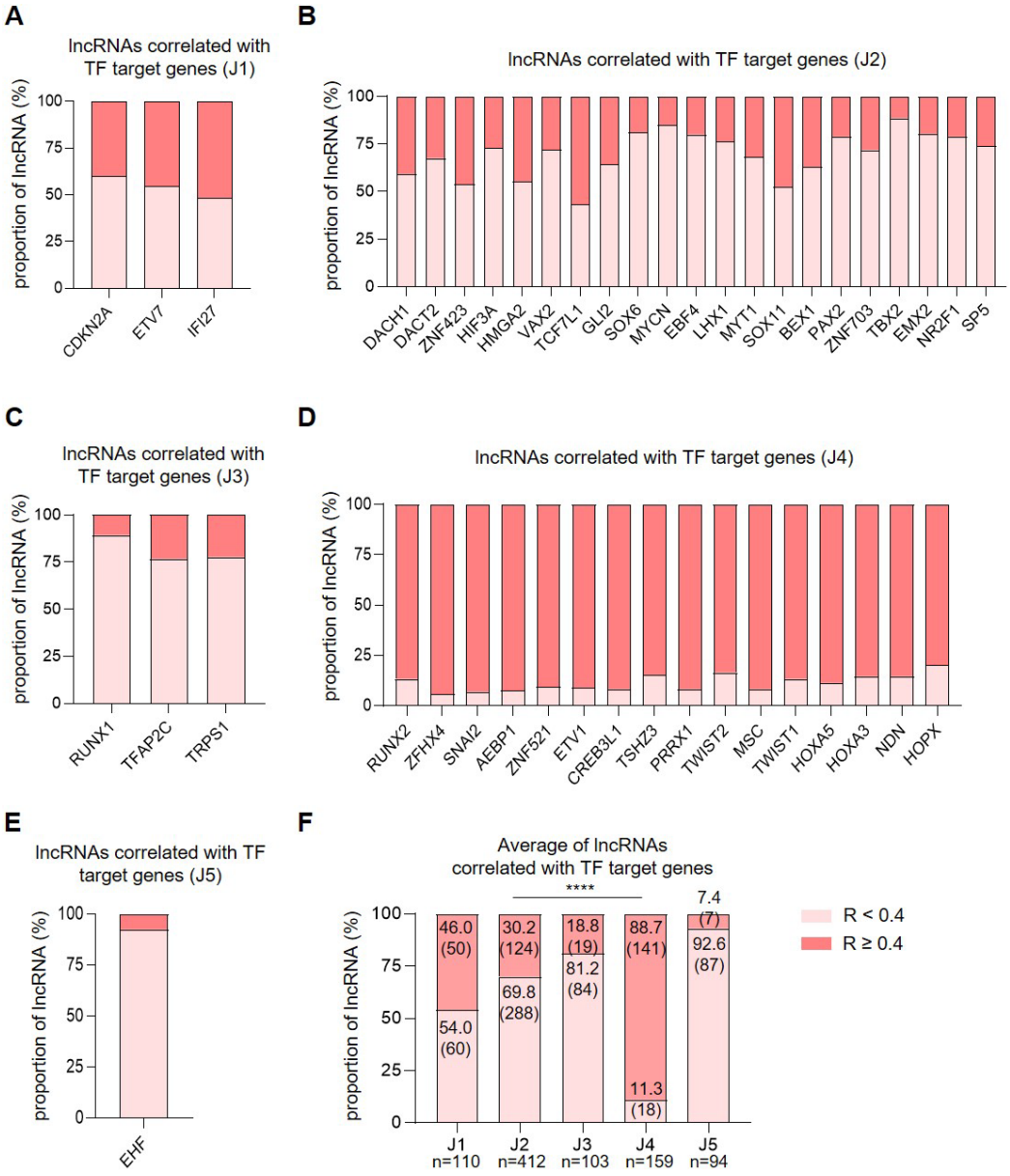


Figure 13. Correlation between target genes of transcription factors and lncRNAs. (A and E) Proportion of Transcription factor target genes positively correlated with lncRNA for each cluster. **(F)** The ratio of lncRNAs that are positively correlated with the transcription factor target genes of each cluster is shown as an average value. Positive correlations based on Pearson's correlation $R \geq 0.4$ are shown in salmon and others are shown in pink. Data are presented as respective ratios (number of samples). p-values calculated by linear-by-linear association. **** $p < 0.001$. The results of Not Significance were not separately indicated.

8. The J4 cluster has a stronger transitional influence than other clusters

To identify the cluster with the strongest influence in HGS_OvCa, the influence of lncRNAs and mRNAs positively correlated with the transcription factor selected from each cluster were analyzed using eigen centrality, and a co-expression network was constructed based on this (Figure 14A and 14E). Eigen centrality is an algorithm used to measure the transitional influence of nodes. Relationships originating from nodes with higher scores contribute more to the node's score than connections from nodes with lower scores. A high eigen centrality means that the node itself is connected to many nodes with high centrality and has strong influence.

As a result of the eigen centrality analysis of the entire cluster, it was confirmed that the node closest to 1 was the lncRNA LINC01614 of the J4 cluster. Subsequently, COL3A1 (eigen centrality = 0.997), the mRNA of the J4 cluster, was confirmed, and MSC (eigen centrality = 0.918) of the J4 cluster was identified as the top node in the MTF (Table 9). The J4 cluster contained most of the top nodes, followed by the corresponding nodes in the J2 cluster. Thus, the J4 cluster contains the most influential MTFs, lncRNAs, and mRNAs among all HGS_OvCa, explaining why the J4 cluster is more transcriptionally active than the J2 cluster, which contains many lncRNAs and mRNAs.

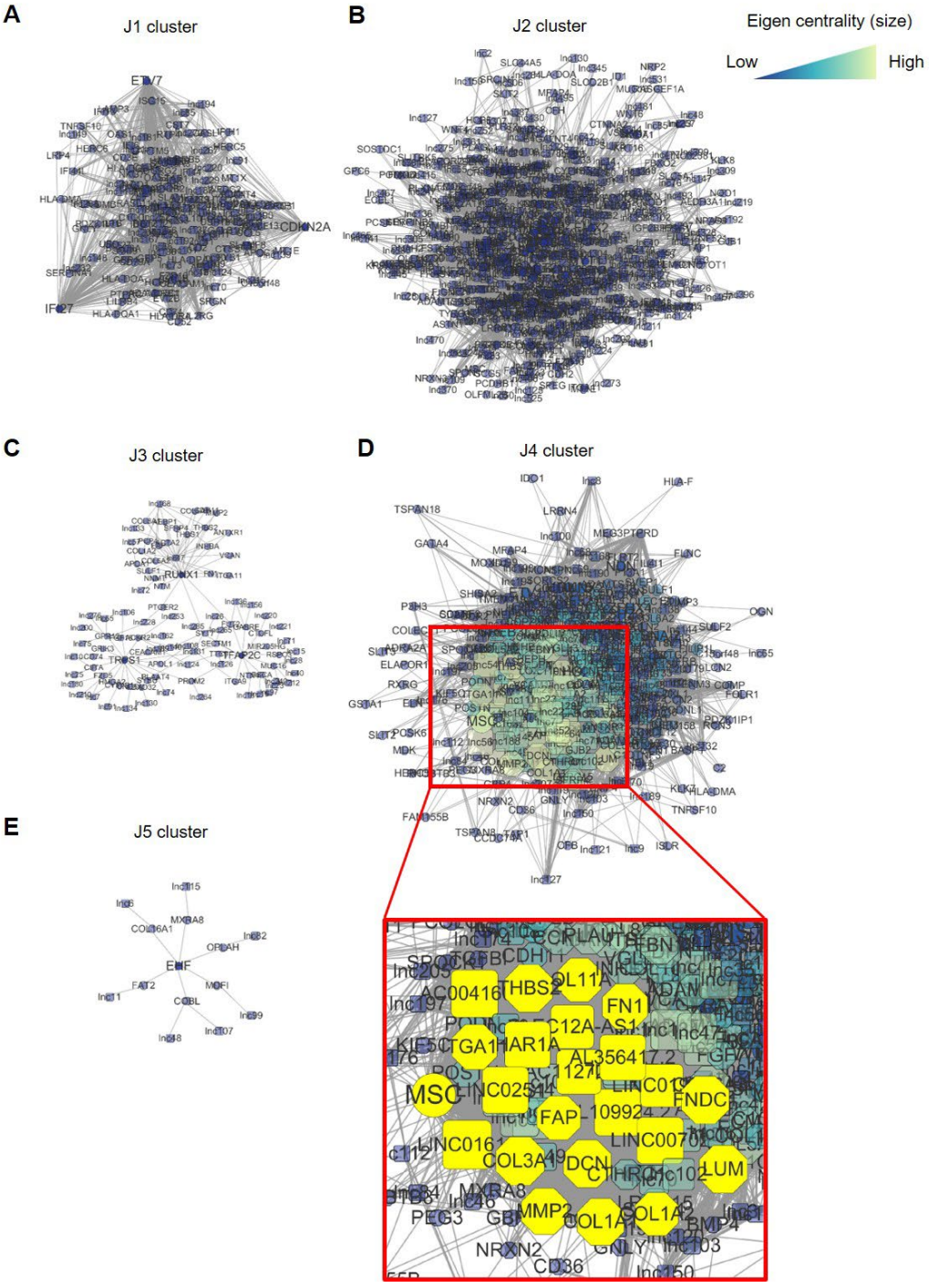


Figure 14. Construction of the master TF-mRNA-lncRNA co-expression network. (A and E) TF-mRNA-lncRNA co-expression networks for each cluster, based on eigenvector centrality, with higher scores representing larger nodes and closer to yellow, and lower scores representing smaller nodes and closer to blue. The red box shows an enlarged view of the area where the high eigenvector centrality nodes of the J4 cluster are concentrated. Yellow nodes are nodes with an eigenvector centrality of 0.9 or higher.

Table 9. List of lncRNAs, mRNAs and transcription factors with eigen centrality ≥ 0.9 in the HGS_OvCa cluster

Rank[1]	Name	Eigen centrality	Type	Cluster
1	LINC01614	1.000	lncRNA	J4
2	COL3A1	0.998	mRNA	J4
3	LUM	0.984	mRNA	J4
4	DCN	0.982	mRNA	J4
5	FNDC1	0.980	mRNA	J4
6	THBS2	0.977	mRNA	J4
7	AC004160.1	0.971	lncRNA	J4
8	COL1A1	0.970	mRNA	J4
9	LINC00702	0.962	lncRNA	J4
10	MMP2	0.961	mRNA	J4
11	AL109924.2	0.961	lncRNA	J4
12	LINC02544	0.959	lncRNA	J4
13	AL356417.2	0.951	lncRNA	J4
14	HAR1A	0.950	lncRNA	J4
15	FN1	0.941	mRNA	J4
16	ITGA11	0.938	mRNA	J4
17	AC112721.2	0.928	lncRNA	J4
18	FAP	0.919	mRNA	J4
19	MSC	0.918	TF	J4
20	COL11A1	0.916	mRNA	J4
21	COL1A2	0.908	mRNA	J4
22	CLEC12A-AS1	0.906	lncRNA	J4
23	LINC01929	0.900	lncRNA	J4

[1] Rank is the rank based on eigen centrality.

9. Master regulators MSC, AEBP1, CREB3L1 that regulate transcription factors of the J4 cluster

As a result of the centrality analysis, it was confirmed that among the transcription factors, theose with the greatest influence and those with the lowest influence were classified. Based on these results, it was predicted that among transcription factors, there may be upper transcription factors that regulate transcription factors. According to the definition provide, the “master regulator” transcription factor is at the top of a regulatory hierarchy and must not be under the regulatory influence of any other gene or transcription factor. Based on previous studies, I identified top transcription factor regulators that regulate transcription factors.

Among the transcription factors, MSC, AEBP1, and CREB3L1 were selected based on the median eigen centrality of 0.5. Additionally, target genes commonly regulated by the corresponding MTFs were identified. As a result, I identified nine common genes, namely COL3A1, LUM, DCN, FNDC1, THBS2, COL1A1, MMP2, FN1, and FAP, regulated by three MTFs (Figure 15A and 15C). The selected common genes had high eigenvector centrality and were genes associated with EMT.

Additionally, the expression of commonly known target genes of the corresponding MTF was confirmed for each cluster. Gene sets commonly identified in each study that identified the target genes of the three master transcription factors were extracellular matrix, cell adhesion, and collagen remodeling, which are related to EMT[72-78]. By evaluating the scores of each gene set within clusters using ssGSEA, I confirmed an increase in the J4 cluster, which includes the master transcription factor (Figure 16A).

Based on previous results, I used TCGA HGS_OvCa data to reconfirm the expression of selected master transcription factors for each cluster and the expression patterns of experimentally validated target genes (Figure 16B and 16C). As expected, I confirmed that the expressions of the three master transcription factors and their target genes are

upregulated in the J4 cluster compared to those in all other clusters. Furthermore, the nine common genes identified through network analysis were validated to be part of the previously mentioned gene set using gprofiler (Figure 16D).

The transcription factors selected in this way were predicted as master transcription factors (MTFs), and the relationship between each MTF and the transcription factors affecting it was confirmed (Figure 17A). As a result, it was confirmed that the top three MTFs commonly target six transcription factors (SNAI2, RUNX2, PRRX1, ZFHX4, ETV1, TWIST1). Therefore, the top three transcription factors were considered to be MTFs that regulate other transcription factors.

In the results of the previous analysis of eigenvector centrality (Table 9), lncRNAs comprised the top rankings with the highest centrality. Although mRNAs were identified as the targets for the selected MTF, lncRNAs were not included in the transcription factor target gene list and could not be identified. Therefore, lncRNAs with a positive correlation (Pearson's Correlation R $\geq$ 0.4) with each MTF were classified, and among the classified lncRNAs, lncRNAs commonly included in the three MTFs were reclassified (Figure 17B). A total of fifteen lncRNAs were classified as common lncRNAs of the three MTFs, and the eigen centrality of the lncRNAs was confirmed. All of the corresponding lncRNAs showed eigen centrality $\geq$ 0.5, and among them, LINC01614 was confirmed to have the highest centrality in all clusters (Table 10). I predicted an overall relationship regulating EMT-related genes in the J4 cluster, including common lncRNAs predicted to regulate master transcription factors (Figure 17C).

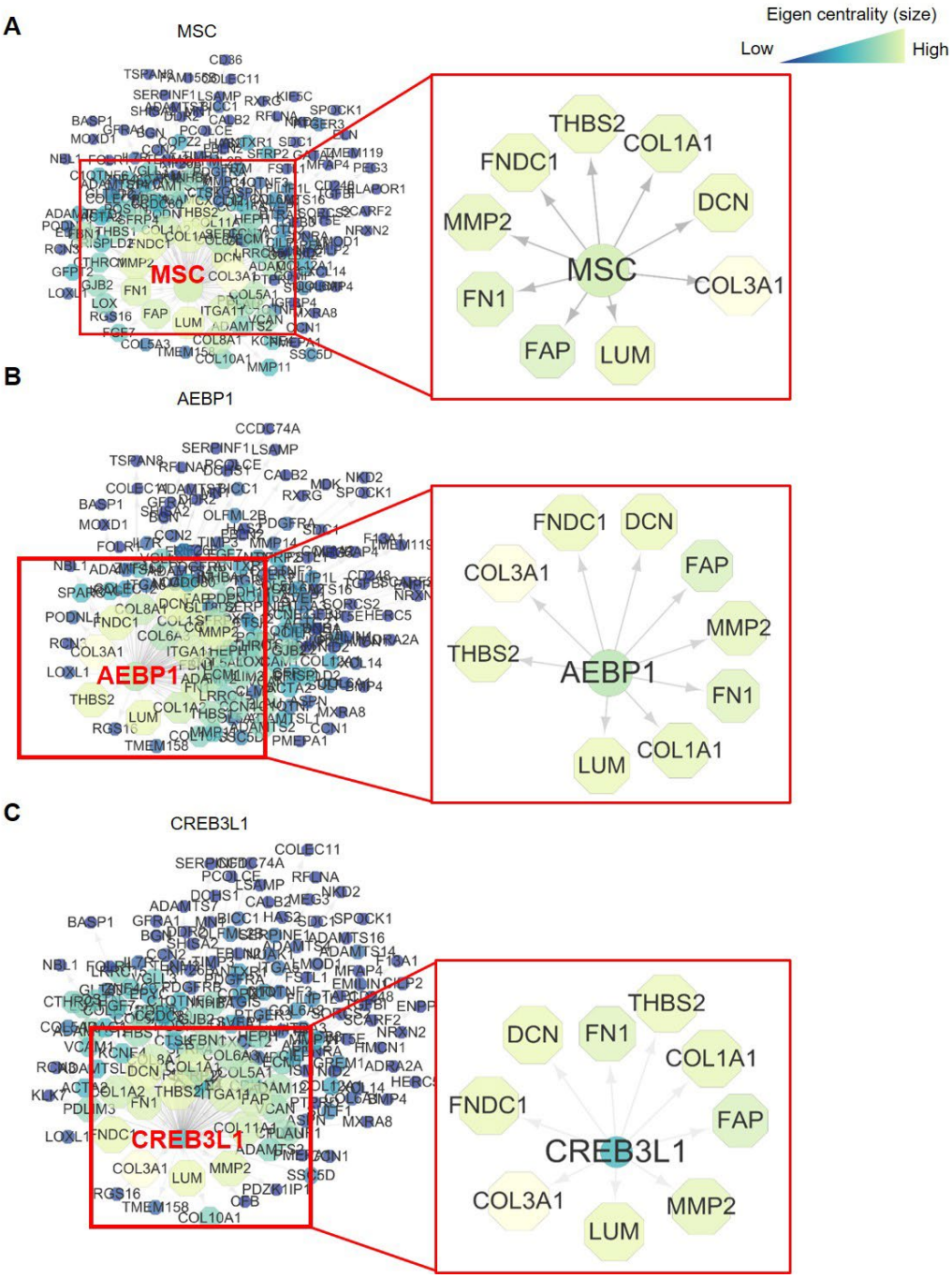


Figure 15. Identification of common target genes of three selected master transcription factors. (A and C) Target genes of transcription factors MSC, AEBP1, and CREB3L1. Red boxes represent common target genes of the three transcription factors. Based on eigenvector centrality, high scores represent larger nodes and are closer to yellow, while lower scores represent smaller nodes and are closer to blue.

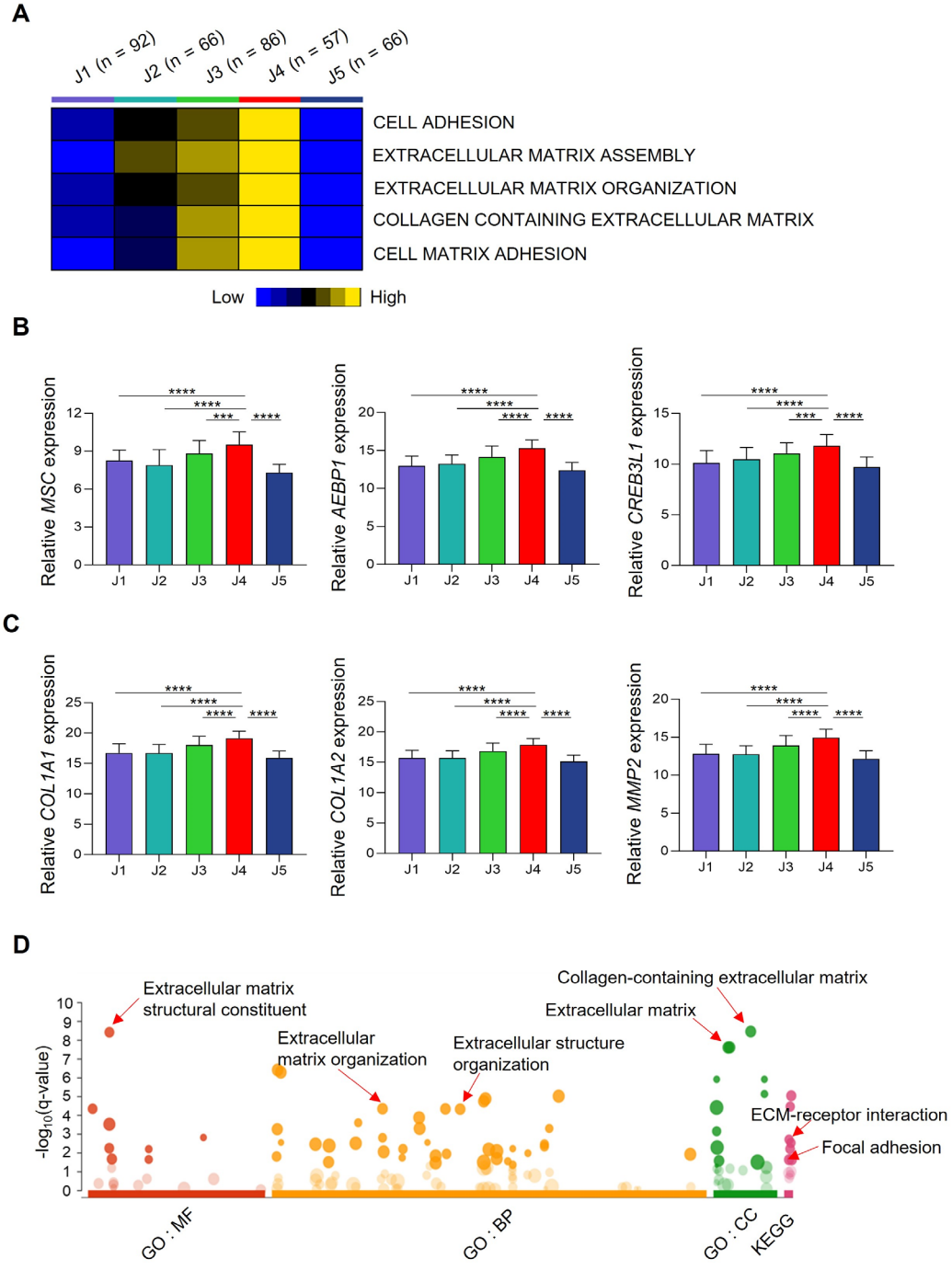


Figure 16. Confirmation of target gene expression of predicted master transcription factors. (A) Results of confirming the expression of gene sets related to master transcription factors for each cluster through ssGSEA. **(B and C)** Verification of expression of master transcription factors and common target genes for each cluster using TCGA HGS_OvCa data. **(D)** A gene set containing nine common genes identified through network analysis was identified using gprofiler. Red arrows indicate gene sets identified in previous studies. Data are presented as mean ± SD and p-values were calculated by ANOVA. *** $p < 0.001$, **** $p < 0.0001$. The results of Not Significance were not separately indicated.

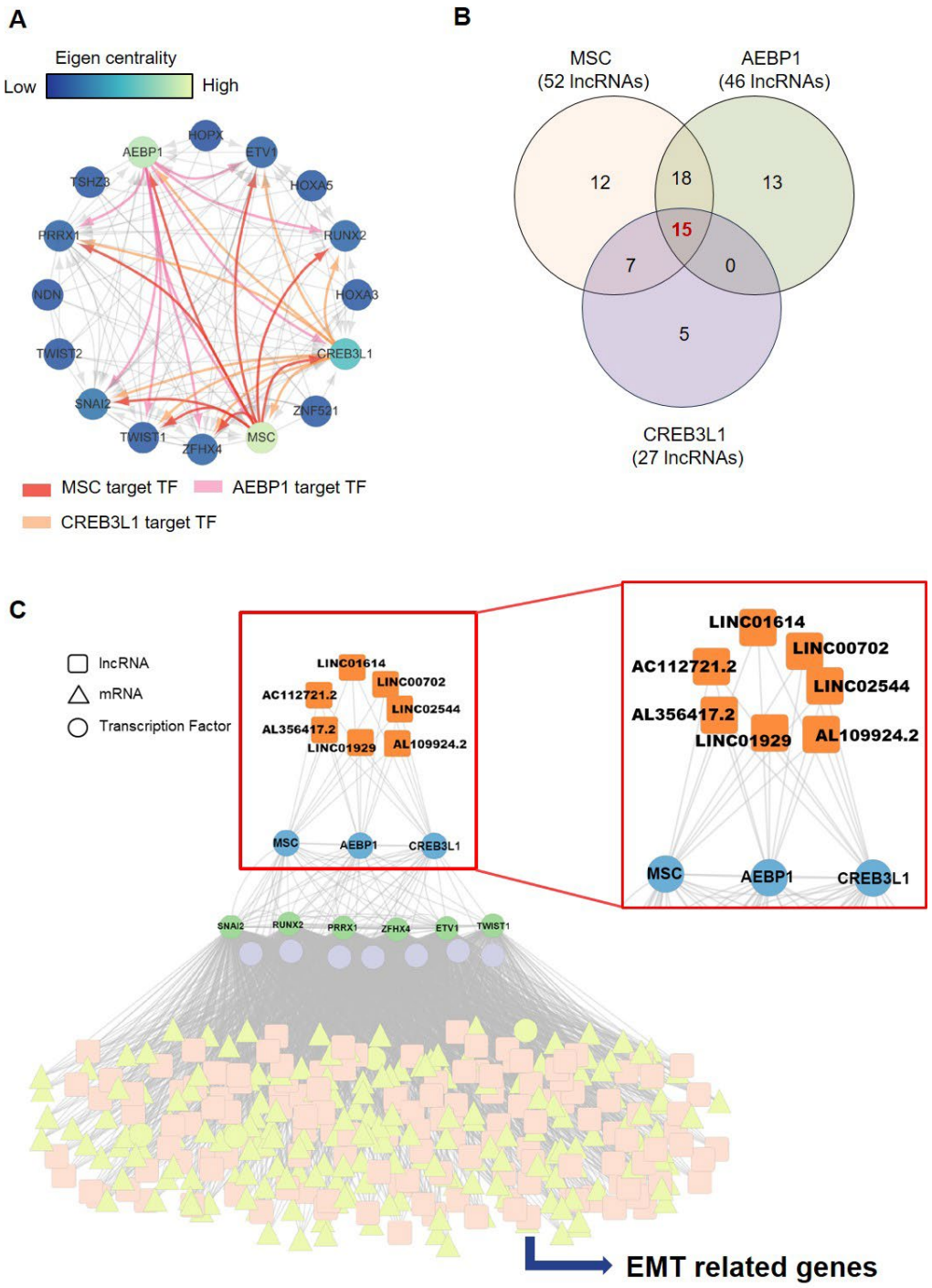


Figure 17. Discovery of lncRNAs regulating master transcription factors. (A) Target transcription factors of top MSC, AEBP1, CREB3L1 among transcription factors of J4 cluster. Salmon color indicates the target Transcription factors of MSC, pink indicates the Transcription factors of AEBP1, and orange indicates the Transcription factors of CREB3L1. Nodes are colored yellow for high scores and blue for low scores based on eigenvector centrality. **(B)** Venn diagram showing the logical relationship of positively correlated (based on Pearson's correlation $R \geq 0.4$) lncRNAs with master transcription factors (MTFs). Red numbers indicate the number of common lncRNAs. **(C)** Illustration showing the overall relationships regulating EMT-related genes in the J4 cluster, including common lncRNAs predicted to regulate master transcription factors. The orange round squares are seven lncRNAs with eigenvector centrality ≥ 0.9 among fifteen common lncRNAs. The three master transcription factors are represented by light blue circles.

Table 10. List of common lncRNAs of the three master transcription factors

lncRNA name	Eigenvector centrality
LINC01614	1.000
LINC00702	0.962
AL109924.2	0.961
LINC02544	0.959
AL356417.2	0.951
AC112721.2	0.928
LINC01929	0.900
HECW2-AS1	0.889
MSC-AS1	0.871
LINC00519	0.858
ACTA2-AS1	0.826
AC104083.1	0.811
AP000892.3	0.801
AC106739.1	0.799
LINC01615	0.760

10. The top regulator of the J4 cluster involved in the expression of EMT genes are lncRNAs

For experimental validation based on previous results, CCLE data were used to confirm the expression of mRNAs and lncRNAs correlated with transcription factor target genes in each cluster (Figure 18A and 18B). It was confirmed that the expression of the J4 cluster transcription factor target gene increased in the 59M cell line among the HGS_OvCa cell lines compared to other cell lines (Figure 18C and 18D). In the J5 cluster, the scores of the target genes of the transcription factor EHF also increased in the 59M cell line, but the scores of lncRNAs increased significantly in the J4 cluster. The results of ssGSEA confirmed the expression of MTF in five cell lines, including the HGS_OvCa cell line possessed by the laboratory. It was confirmed that the expression of the master transcription factors selected above, MSC, AEBP1, CREB3L1 all increased significantly in the 59M cell line (Figure 18E and 18G).

As mentioned above, according to the definition that MTF is not affected by other genes or transcription factors, each MTF was silenced in the 59M cell line. As a result of silencing, it was confirmed whether the expression of the target gene and the selected lncRNA changed as the expression of MTF decreased. It was confirmed that the expression of each MTF was significantly reduced by siRNA, and it was confirmed that all nine common target genes showed significantly decreased expressions (Figure 19A and 19F). Among the common lncRNAs, the seven most influential lncRNAs with intrinsic centrality greater than 0.9 were selected to determine their expression. As a result, I confirmed that the expression of common lncRNAs expected to regulate MTF did not change as MTF expression decreased (Figure 20A and 20C). Based on eigen centrality, it was confirmed that lncRNAs with higher centrality than MTF were not affected by MTF. Therefore, it was confirmed that the most influential lncRNA in the J4 cluster regulates the lower genes. This suggests that there are lncRNAs that ultimately regulate MTF, and that lncRNAs may also regulate downstream EMT genes (Figure 21).

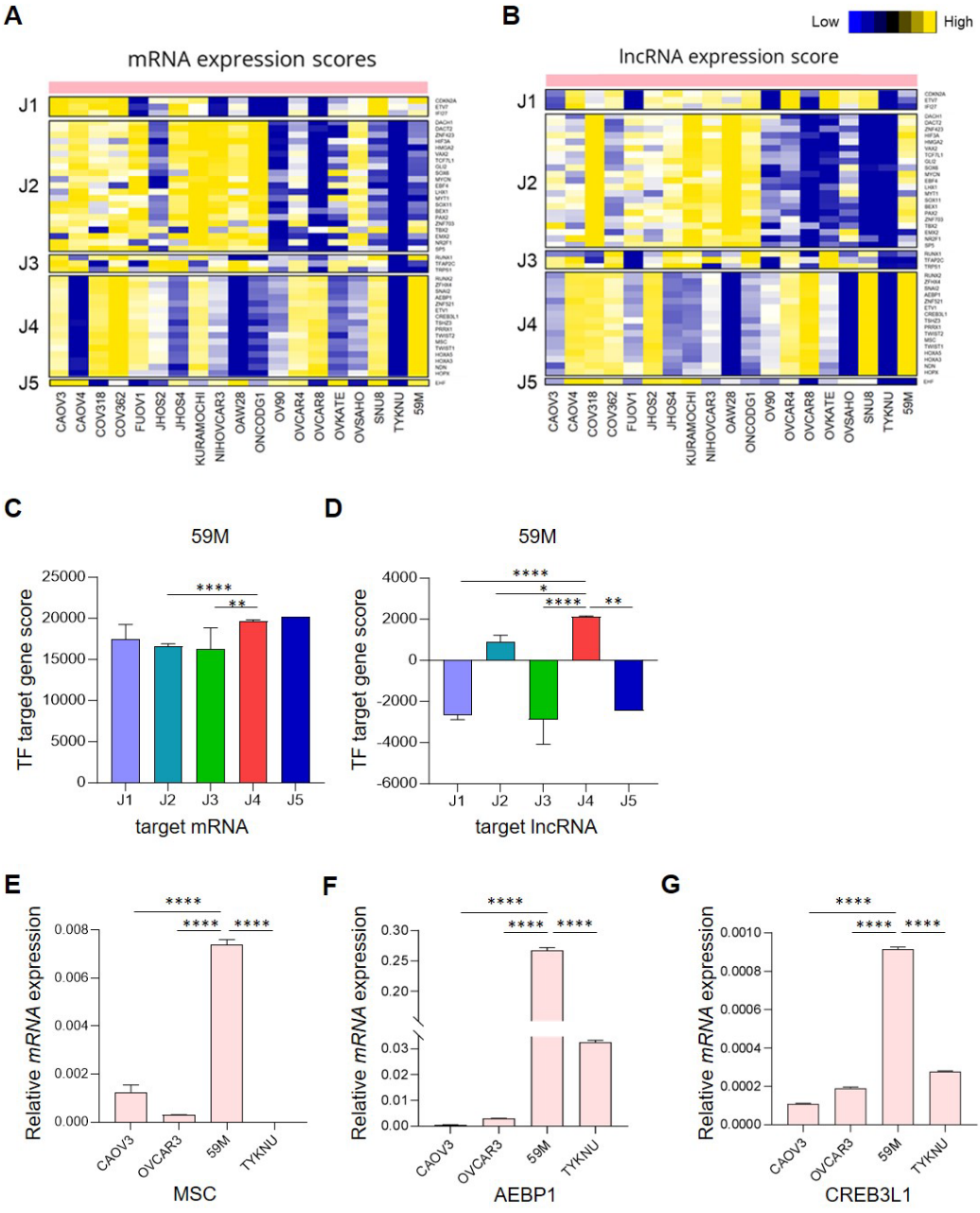


Figure 18. Screening of cell lines reflecting cluster characteristics. (A) ssGSEA scores of selected transcription factor target genes and **(B)** ssGSEA scores of lncRNAs and transcription factor target genes. In the 59M cell line, **(C)** the ssGSEA scores of transcription factor target genes were calculated for each cluster, as well as the **(D)** ssGSEA scores of lncRNAs that positively correlate with transcription factor target genes. Relative RNA-expression ratios of master transcription factors **(E)** MSC in HGS_OvCa cell line. Data are presented as mean±SD and p-values were calculated by ANOVA and all experiments were repeated in triplicate. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. The results of Not Significance were not separately indicated.

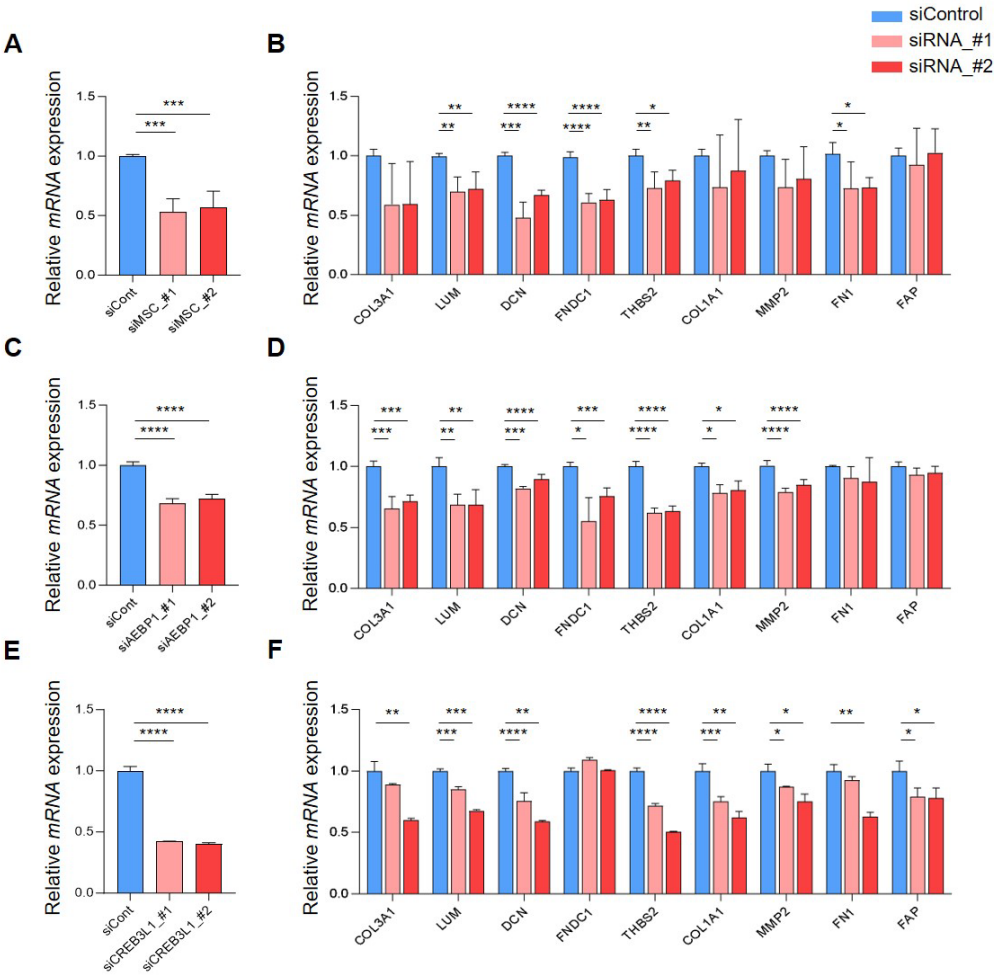


Figure 19. Verification of common gene expression of master transcription factors. 59M cells were treated with control siRNA (siControl) or three master transcription factor siRNAs for 48 hours. q-PCR results confirming the expression of **(A, C, E)** MSC, AEBP1, CREB3L1, **(B, D, F)** common mRNA target genes in transfected cells. Data are presented as mean±SD and p-values were calculated by ANOVA and all experiments were repeated in triplicate. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. The results of Not Significance were not separately indicated.

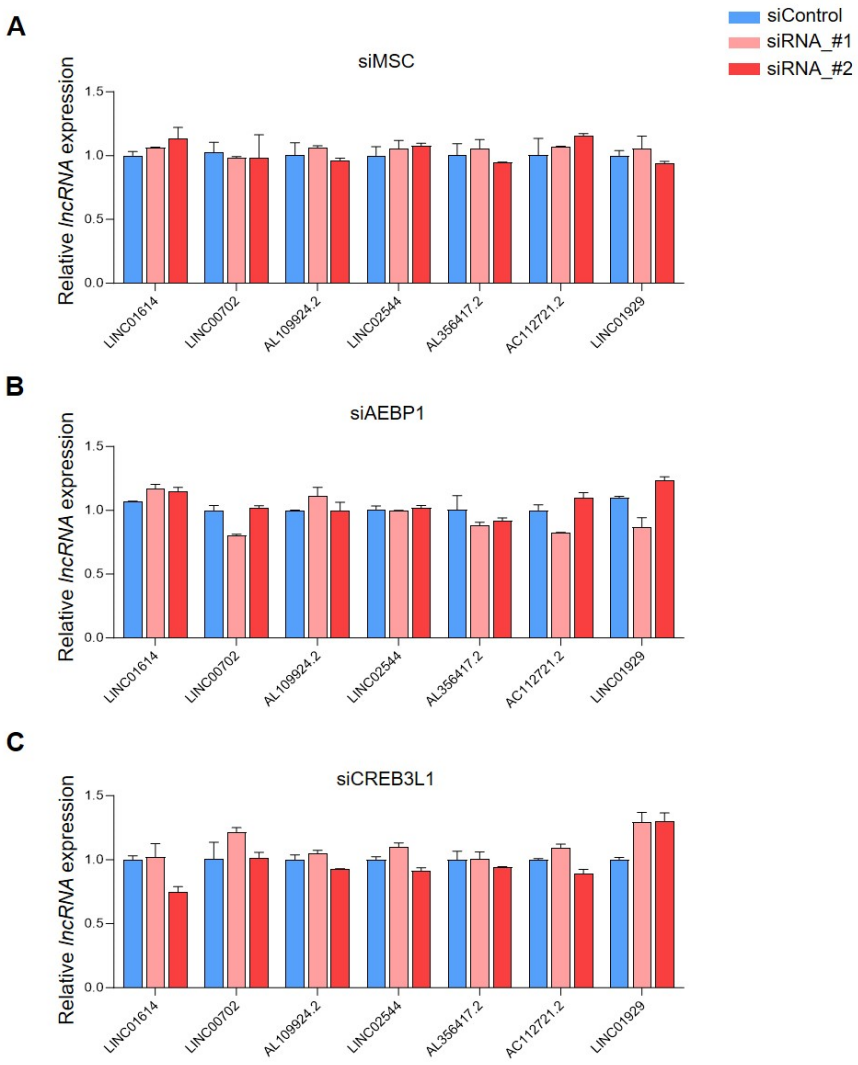


Figure 20. Validation of lncRNAs regulating master transcription factors. 59M cells were treated with control siRNA (siControl) or three master transcription factor siRNAs for 48 hours. **(A and C)** q-PCR results confirming the expression of common lncRNAs in transfected cells. Data are presented as mean±SD and p-values were calculated by ANOVA and all experiments were repeated in triplicate. The results of Not Significance were not separately indicated.

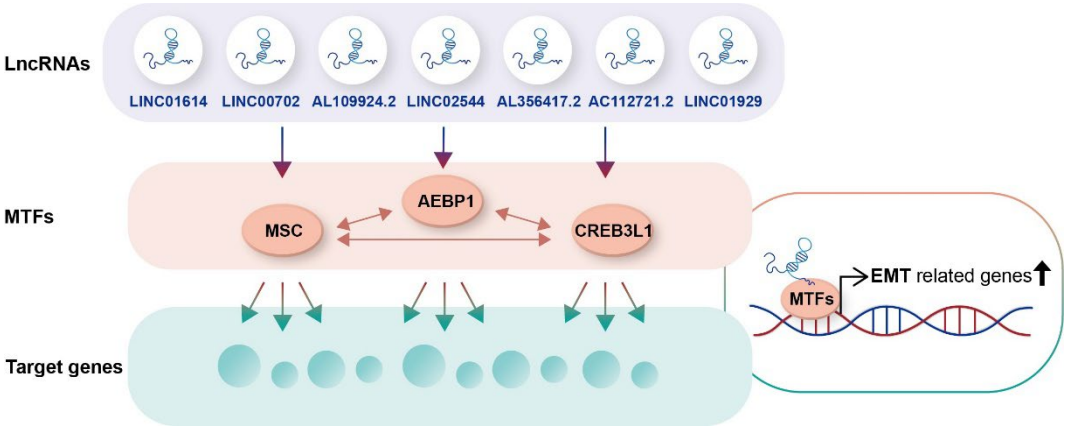


Figure 21. A proposed model of EMT gene regulation through lncRNAs, which regulate master transcription factors (MTFs) in HGS_OvCa.

IV. DISCUSSION

Ovarian cancer, particularly high-grade serous ovarian carcinoma (HGS_OvCa), is one of the most lethal human malignancies, and ovarian cancer patients usually present at an advanced disease stage and relapse frequently[79,80]. New targets are being discovered to overcome these therapeutic limitations, and one of them is lncRNAs. LncRNAs are being investigated as new diagnostic and therapeutic targets in various types of human cancers, and recent studies suggest that lncRNAs play an important role in regulating tumor progression, metastasis, estrogen response, and drug resistance in ovarian cancer[81-83]. In addition, aberrant expression of lncRNAs can provide important information for diagnosis, treatment, and prognosis of patients. However, the role of lncRNAs in HGS_OvCa is still in its infancy compared to miRNAs and requires further study.

To determine whether lncRNAs are involved in the carcinogenesis process, I performed CNMF clustering using HGSC RNA-seq data. In the process of confirming the function of lncRNAs, the characteristics of each cluster were defined using an analysis tool that indirectly predicted the function of lncRNAs by matching the nucleotide sequence of the coding gene. Research related to lncRNAs has been actively conducted since 2010[84], but there is no gene set provided such as hallmark, GO, or Kegg gene set used for functional analysis. Therefore, like “FuncPred”, functions are indirectly inferred by matching coding genes, or functions are suggested through *cis*-acting by checking neighboring genes. In order to review the functions identified in this way, the function of the coding gene was checked with gprofiler, and as a result, it was confirmed that it was consistent with FuncPred. Through this process, HGS_OvCa was classified into five major clusters according to functional characteristics: "Immune group", "EMT group", "Estrogen response group", "EMT-androgen response group", and "Differentiated group."

To confirm where the difference in function of each cluster originated, each factor that plays a large role in the carcinogenesis process was analyzed. First, as a result of checking the DNA mutation profile, it was confirmed that there was no difference in the top mutation genes for each cluster. Second, the same difference was not confirmed in the result of

somatic copy number alternation analysis. In ovarian high-grade serous carcinoma (HGSC), somatic copy number alternation analysis confirmed amplification in the 1q21.3 chromosomes, and lncRNA present in the chromosome was investigated. Expression of the lncRNA FAL1 (Focally Amplified Long Noncoding on Chromosome 1) identified through this process is closely related to E2F1 upregulation, suggesting that FAL1 affects carcinogenesis through the cell cycle[85,86]. However, in each cluster classified, characteristic amplification and deletion chromosome regions, as in previous studies, were not identified. Third, in addition to the rule that miRNA and lncRNA act only on mRNA, the lncRNA-miRNA interaction was confirmed based on the study that they interact with each other to further regulate their effects in the transcriptome. Depending on these interactions, miRNA regulates the function of lncRNA or acts as miRNA decoy to suppress miRNA target mRNA [87,88]. However, miRNA expression medians were similar for each cluster, and no significant difference between clusters was identified. LncRNAs are widely involved in epigenetic regulatory mechanisms such as DNA methylation and are known to be involved in the development and progression of malignant tumors[89]. For example, the p53-induced lncRNA TP53TG1 exhibits promoter hypermethylation in gastric and colon cancer[90]. Another example is the tumor suppressor lncRNA Growth Arrest-Specific transcript 5 (GAS5), which has been shown to be downregulated in gastric cancer through promoter hypermethylation[91]. As a result of confirming DNA methylation by cluster based on previous studies, a value like the average beta-value of ovarian cancer was confirmed.

As a result of identifying four important factors in the carcinogenesis process, no factors that could clearly explain the difference in function of each cluster were identified. Except for the carcinogenesis-related factors analyzed above, I paid attention to the association between transcription factors and lncRNAs as factors related to other carcinogenesis processes. Identification of transcription factor drivers in solid tumors has expanded considerably in recent years. Overexpression of the transcription factor ETS translocation variant 1 (ETV1) occurs via chromosomal translocation and has been reported to cause prostate cancer[92,93]. In addition, various mechanisms (transcription factors accessing DNA,

mRNA synthesis, processing, stability, and translation) regulating gene expression are influenced by lncRNAs[94,95]. Through these preceding studies, ARACNe and Viper were used to identify the transcription factors of each cluster. It was confirmed that the functional characteristics defined for each cluster and the function of the target gene set of transcription factors were similar. Although the J2 and the J4 clusters have the same EMT function, different transcription factors were selected.

The process of EMT remains incompletely resolved to date, making it a difficult process to target therapeutically[96]. Therefore, it is important to identify important molecules that regulate EMT during carcinogenesis. The difference between the J2 and J4 clusters is androgen response, and the role of androgen receptors in tumor development and tumor progression in female breast and ovarian cancer has been previously reported[97]. The ER, PR, and AR pathways are involved in regulating signaling pathways such as cell-proliferation, apoptosis, epithelial to mesenchymal transition, and cell migration and invasion[98,99]. In women, androgens are produced by the ovaries, adrenal glands, and peripheral conversion of androgen precursors (DHEA)[100,101]. The percentage of testosterone of ovarian origin is higher in postmenopausal women. As a result of checking each age in the classified clusters, all cluster except for the J3 group (52.00 ± 8.80) were included in the standard age of postmenopausal women (premenopausal is < 55 years, postmenopausal is ≥ 55 years)[102], and the oldest cluster in the cluster was the J4 cluster (66.00 ± 9.40). There was no difference between the groups of the J2 cluster and the J4 cluster (p-value = 0.124), but functional analysis suggests that androgen response brings about a difference between the two groups. In addition, through data confirming the correlation with transcription factors selected for each cluster, it was confirmed that the J2 cluster contains many mRNAs and lncRNAs, but the proportion of mRNAs and lncRNAs that are positively correlated with target genes of transcription factors is smaller than that of the J4 cluster. Through this result, it was found that transcription factors were strongly activated in the J4 cluster. Further analysis confirming the centrality of each cluster confirmed that the transcription factors, lncRNAs and mRNAs contained in the J4 cluster

have higher eigen centrality than the other clusters. In other words, the fact that the J4 cluster contains many factors with high influence supports the contents of the previously confirmed results.

Through the results of the centrality analysis, it was confirmed that the centrality was greatly different within the transcription factors. There is a group of transcription factors that regulate oncogenes and disease processes, and it was confirmed that there is a hierarchy in the regulatory activities of these transcription factors, confirming previous studies that identified the upper regulators of the hierarchy as master transcription factors[103]. Master transcription factors have also been defined as those that are not regulated or influenced by other genes or regulators. Through this definition, three MTFs (MSC, AEBP1, CREB3L1) with high eigen centrality among transcription factors and not affected by other transcription factors were selected. In previous studies, MECOM, PAX8, and SOX17 were suggested as master transcription factors involved in the development of high-grade serous ovarian cancer[33,34,36,37]. However, the relevant transcription factors do not appear to play a role in defining the characteristics of each cluster identified through the analysis. Selected as master transcription factors, MSCs are members of the helix-loop-helix (HLH) family of TFs and were first reported in mouse skeletal muscle precursors[104]. Recent studies have shown that MSCs, along with LEF1, have functions related to EMT-related extracellular matrix (ECM) organization and cell-ECM interactions, and these two transcription factors appear to be specific transcription factors identified only in LUAD[72]. In conclusion, MSCs have been reported to promote malignant progression of lung cancer through the EMT process. The role of AEBP1 in promoting carcinogenesis has been recently investigated by several research groups, and it has been found to promote tumorigenesis through the NF-κB pathway and EMT process in colon cancer, gastric cancer[105,106]. In addition, recent studies have reported that CREB3L1 participates in cancer initiation and progression and can serve as a promising clinical biomarker for cancer patients[107]. Similar to the functions of other MTFs identified above, they induce cell invasion and metastasis through induction of EMT. It was interesting to note that the function of the cluster identified above coincided

with the role of the selected master transcription factor.

In addition, it was confirmed that lncRNAs commonly included in the selected MTFs had high centrality, and it was interesting that lncRNAs could have the possibility of regulating MTFs. Therefore, I confirmed whether the lncRNA regulates MTF while verifying whether the expression of the common lncRNA was reduced when MTF was silencing. Interestingly, the three MTF target genes showed a significant decrease as MTF decreased, but the common lncRNA did not change. This result is consistent with the definition of master regulator mentioned above, and therefore, it was judged that the corresponding lncRNA regulates MTF. Previous studies have reported that lncRNAs are specifically transcribed and act as signaling molecules to regulate the transcription of downstream genes. For instance, in cases of DNA damage, the LncRNA PANDA, activated by the interaction between p53 and cyclin-dependent kinase inhibitor 1A (CDKN1A, p21), has been reported to increase the survival time of tumor cells by targeting the nuclear transcription factor Y subunit alpha (NF-YA)[108]. Therefore, it seems that lncRNAs that regulate MTF also play a role as signal molecules. Among lncRNAs, LINC01614, which has the highest centrality, is known to promote cancer development in lung cancer and breast cancer. Additionally, most of the genes identified as being co-expressed with LINC01614 were associated with EMT in human cancers[109-111].

Currently, several examples of lncRNAs have been described as potential clinical biomarkers for predicting response to therapy or for prognosis in breast cancer, such as HOTAIR, H19, and DSCAM-AS1[112]. Although their clinical utility has not yet been clearly demonstrated, the use of lncRNAs as predictive biomarkers in response to treatment has advantages over protein- and mRNA-based biomarkers as they reveal tissue- and stage-specific expression[113,114]. For example, CRISPR/Cas9 silencing of NEAT1 or MALAT1 has been reported to inhibit cancer cell metastasis. A patent silencing UCA1 using CRISPR/Cas9 inhibited the growth of cancer cells (CN106399306B)[115]. In addition, observational clinical trials of lncRNA WRAP53 and UCA-1 for hepatocellular carcinoma are in progress (NCT05088811), and observational clinical trials of lncRNA MFI2-AS1

are also underway with 260 patients with Kidney cancer (NCT04946266)[116]. Therefore, this study confirmed that it can be classified according to lncRNA expression using HGS_OvCa RNA-seq data, and that each classified cluster has a distinct molecular biological function. In addition, master transcription factors (MTFs) regulating transcription factors have been identified as potential triggers for these traits, and seven lncRNAs regulating MTFs have been proposed. Seven lncRNAs were selected from the J4 cluster, which is associated with a poor prognosis in HGS_OvCa. It is anticipated that patients with increased expression of these specific lncRNAs will also show increased expression of EMT-related genes, a characteristic of the J4 cluster. As the selected lncRNA acts as an upstream regulator influencing downstream genes, its potential as a therapeutic agent can be confirmed through lncRNA manipulation. Similar to the lncRNA therapeutics mentioned in the clinical trials above, gene silencing methods such as lncRNA CRISPR/Cas9 and siRNA are necessary to identify changes in relevant master transcription factors and downstream genes. It is expected that the role that the selected lncRNA plays in the field of precision medicine will become clearer through future verification results.

V. CONCLUSION

In summary, my study identified functional features of each cluster classified by lncRNA expression. Each cluster was classified into "Immune group", "EMT group", "Estrogen response group", "EMT-Androgen response group", and "Differentiation group" through functional analysis. Afterward, transcription factors representing the characteristics of each cluster were presented, and among them, the J4 cluster, which has a stronger transcription factor activity than other clusters, was noted. Among the J4 cluster, master transcription factors "MSC", "AEBP1", and "CREB3L1" that regulate transcription factors were identified. In addition, seven lncRNAs with stronger influence than the selected master transcription factors were identified, and it was confirmed that the selected lncRNAs regulate master transcription factors and EMT-related genes. This study suggests that the seven lncRNAs regulating MTFs contribute to identifying the transcription factor regulation mechanism of lncRNAs in HGS_OvCa and can be indicators for personalized medicine.

VI. REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019;69:7-34.
2. Banerjee S, Kaye SB. New strategies in the treatment of ovarian cancer: current clinical perspectives and future potential. Clin Cancer Res 2013;19:961-8.
3. Peres LC, Cushing-Haugen KL, Kobel M, Harris HR, Berchuck A, Rossing MA, et al. Invasive Epithelial Ovarian Cancer Survival by Histotype and Disease Stage. J Natl Cancer Inst 2019;111:60-8.
4. Romero I, Bast RC, Jr. Minireview: human ovarian cancer: biology, current management, and paths to personalizing therapy. Endocrinology 2012;153:1593-602.
5. Carcangiu M, Kurman RJ, Carcangiu ML, Herrington CS. WHO classification of tumours of female reproductive organs: International Agency for Research on Cancer; 2014.
6. Meinhold-Heerlein I, Fotopoulou C, Harter P, Kurzeder C, Mustea A, Wimberger P, et al. The new WHO classification of ovarian, fallopian tube, and primary peritoneal cancer and its clinical implications. Arch Gynecol Obstet 2016;293:695-700.
7. Reid BM, Permuth JB, Sellers TA. Epidemiology of ovarian cancer: a review. Cancer Biol Med 2017;14:9-32.
8. Kurman RJ, Shih Ie M. Pathogenesis of ovarian cancer: lessons from morphology and molecular biology and their clinical implications. Int J Gynecol Pathol 2008;27:151-60.
9. Punzon-Jimenez P, Lago V, Domingo S, Simon C, Mas A. Molecular Management of High-Grade Serous Ovarian Carcinoma. Int J Mol Sci 2022;23.
10. Bowtell DD, Bohm S, Ahmed AA, Aspuria PJ, Bast RC, Jr., Beral V, et al. Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. Nat Rev Cancer 2015;15:668-79.
11. Prat J. Ovarian carcinomas: five distinct diseases with different origins, genetic alterations, and clinicopathological features. Virchows Arch 2012;460:237-49.
12. Pedersen JS, Bejerano G, Siepel A, Rosenbloom K, Lindblad-Toh K, Lander ES, et al. Identification and classification of conserved RNA secondary structures in the human genome. PLoS Comput Biol 2006;2:e33.
13. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell 2011;43:904-14.
14. Sherstyuk VV, Medvedev SP, Zakian SM. Noncoding RNAs in the Regulation of Pluripotency and Reprogramming. Stem Cell Rev Rep 2018;14:58-70.
15. Quinodoz S, Guttman M. Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. Trends Cell Biol 2014;24:651-63.
16. Wu DD, Chen X, Sun KX, Wang LL, Chen S, Zhao Y. Role of the lncRNA ABHD11-AS(1) in the tumorigenesis and progression of epithelial ovarian cancer

through targeted regulation of RhoC. Mol Cancer 2017;16:138.
17. Lin X, Yang F, Qi X, Li Q, Wang D, Yi T, et al. LncRNA DANCR promotes tumor growth and angiogenesis in ovarian cancer through direct targeting of miR-145. Mol Carcinog 2019;58:2286-96.
18. Yang M, Zhai Z, Guo S, Li X, Zhu Y, Wang Y. Long non-coding RNA FLJ33360 participates in ovarian cancer progression by sponging miR-30b-3p. Onco Targets Ther 2019;12:4469-80.
19. Pei CL, Fei KL, Yuan XY, Gong XJ. LncRNA DANCR aggravates the progression of ovarian cancer by downregulating UPF1. Eur Rev Med Pharmacol Sci 2019;23:10657-63.
20. Liu D, Li H. Long non-coding RNA GEHT1 promoted the proliferation of ovarian cancer cells via modulating the protein stability of HIF1alpha. Biosci Rep 2019;39.
21. Zhang H, Yao B, Tang S, Chen Y. LINK-A Long Non-Coding RNA (lncRNA) Participates in Metastasis of Ovarian Carcinoma and Upregulates Hypoxia-Inducible Factor 1 (HIF1alpha). Med Sci Monit 2019;25:2221-7.
22. Zhu D, Huang X, Liang F, Zhao L. LncRNA miR503HG interacts with miR-31-5p through multiple ways to regulate cancer cell invasion and migration in ovarian cancer. J Ovarian Res 2020;13:3.
23. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869-74.
24. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci U S A 2001;98:13790-5.
25. Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin Cancer Res 2008;14:5198-208.
26. Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. Nature 2011;474:609-15.
27. Zhang Z, Huang K, Gu C, Zhao L, Wang N, Wang X, et al. Molecular Subtyping of Serous Ovarian Cancer Based on Multi-omics Data. Sci Rep 2016;6:26001.
28. Libermann TA, Zerbini LF. Targeting transcription factors for cancer gene therapy. Curr Gene Ther 2006;6:17-33.
29. Darnell JE, Jr. Transcription factors as targets for cancer therapy. Nat Rev Cancer 2002;2:740-9.
30. Garraway LA, Sellers WR. Lineage dependency and lineage-survival oncogenes in human cancer. Nat Rev Cancer 2006;6:593-602.
31. Eliades P, Abraham BJ, Ji Z, Miller DM, Christensen CL, Kwiatkowski N, et al. High MITF Expression Is Associated with Super-Enhancers and Suppressed by CDK7 Inhibition in Melanoma. J Invest Dermatol 2018;138:1582-90.
32. Leach DA, Panagopoulos V, Nash C, Bevan C, Thomson AA, Selth LA, et al. Cell-lineage specificity and role of AP-1 in the prostate fibroblast androgen receptor

cistrome. Mol Cell Endocrinol 2017;439:261-72.
33. Bleu M, Mermet-Meillon F, Apfel V, Barys L, Holzer L, Bachmann Salvy M, et al. PAX8 and MECOM are interaction partners driving ovarian cancer. Nat Commun 2021;12:2442.
34. Lin L, Shi K, Zhou S, Cai MC, Zhang C, Sun Y, et al. SOX17 and PAX8 constitute an actionable lineage-survival transcriptional complex in ovarian cancer. Oncogene 2022;41:1767-79.
35. Reddy J, Fonseca MAS, Corona RI, Nameki R, Segato Dezem F, Klein IA, et al. Predicting master transcription factors from pan-cancer expression data. Sci Adv 2021;7:eabf6123.
36. Chaves-Moreira D, Mitchell MA, Arruza C, Rawat P, Sidoli S, Nameki R, et al. The transcription factor PAX8 promotes angiogenesis in ovarian cancer through interaction with SOX17. Sci Signal 2022;15:eabm2496.
37. Shi K, Yin X, Cai MC, Yan Y, Jia C, Ma P, et al. PAX8 regulon in human ovarian cancer links lineage dependency with epigenetic vulnerability to HDAC inhibitors. Elife 2019;8.
38. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a Shared Vision for Cancer Genomic Data. N Engl J Med 2016;375:1109-12.
39. Deng M, Bragelmann J, Kryukov I, Saraiva-Agostinho N, Perner S. FirebrowseR: an R client to the Broad Institute's Firehose Pipeline. Database (Oxford) 2017;2017.
40. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;483:603-7.
41. Bradbury M, Borras E, Perez-Benavente A, Gil-Moreno A, Santamaria A, Sabido E. Proteomic Studies on the Management of High-Grade Serous Ovarian Cancer Patients: A Mini-Review. Cancers (Basel) 2021;13.
42. Rojas V, Hirshfield KM, Ganesan S, Rodriguez-Rodriguez L. Molecular Characterization of Epithelial Ovarian Cancer: Implications for Diagnosis and Treatment. Int J Mol Sci 2016;17.
43. Shih IM, Wang Y, Wang TL. The Origin of Ovarian Cancer Species and Precancerous Landscape. Am J Pathol 2021;191:26-39.
44. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res 2019;47:D766-D73.
45. Christophe Leys CL, Olivier Klein, Philippe Bernard, Laurent Licata. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology 2013;49:764-6.
46. Xu T, Le TD, Liu L, Su N, Wang R, Sun B, et al. CancerSubtypes: an R/Bioconductor package for molecular cancer subtype identification, validation and visualization. Bioinformatics 2017;33:3131-3.
47. Rohlf FJ, Fisher DR. Tests for Hierarchical Structure in Random Data Sets.

Systematic Biology 1968;17:407-12.
48. Fatima Batool CH. Clustering with the Average Silhouette Width. Computational Statistics & Data Analysis 2021;158:107190.
49. Abiodun M. Ikotun AEE, Laith Abualigah, Belal Abuhaija, Jia Heming. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. Information Sciences 2023;622:178-210.
50. Agarwal GG. Statistics for surgeons - understanding survival analysis. Indian J Surg Oncol 2012;3:208-14.
51. Alboukadel Kassambara MK, Przemyslaw Biecek. survminer: Drawing Survival Curves using 'ggplot2'. 0.4.9 ed; 2021. p.R package.
52. Perron U, Provero P, Molineris I. In silico prediction of lncRNA function using tissue specific and evolutionary conserved expression. BMC Bioinformatics 2017;18:144.
53. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res 2019;47:W191-W8.
54. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018;28:1747-56.
55. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 2011;12:R41.
56. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
57. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 2010;11:587.
58. Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat Genet 2016;48:838-47.
59. Lachmann A, Giorgi FM, Lopez G, Califano A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. Bioinformatics 2016;32:2233-5.
60. Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res 2018;46:D380-D6.
61. Yevshin I, Sharipov R, Kolmykov S, Kondrakhin Y, Kolpakov F. GTRD: a database on gene transcription regulation-2019 update. Nucleic Acids Res 2019;47:D100-D5.
62. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell 2018;172:650-65.
63. Pedersen TL. tidygraph: A Tidy API for Graph Manipulation. 1.2.3 ed; 2023.

64. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498-504.
65. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 2009;462:108-12.
66. Gregory R. Warnes BB, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz, Bill Venables, Tal Galili. gplots: Various R Programming Tools for Plotting Data. 2.17.0 ed; 2015.
67. Fransen M, Lismont C. Redox Signaling from and to Peroxisomes: Progress, Challenges, and Prospects. Antioxid Redox Signal 2019;30:95-112.
68. Therachiyil L, Anand A, Azmi A, Bhat A, Korashy HM, Uddin S. Role of RAS signaling in ovarian cancer. F1000Res 2022;11:1253.
69. Domcke S, Sinha R, Levine DA, Sander C, Schultz N. Evaluating cell lines as tumour models by comparison of genomic profiles. Nat Commun 2013;4:2126.
70. Lord CJ, Ashworth A. Targeted therapy for cancer using PARP inhibitors. Curr Opin Pharmacol 2008;8:363-9.
71. Drew Y, Calvert H. The potential of PARP inhibitors in genetic breast and ovarian cancers. Ann N Y Acad Sci 2008;1138:136-45.
72. Zhang S, Li M, Ji H, Fang Z. Landscape of transcriptional deregulation in lung cancer. BMC Genomics 2018;19:435.
73. Cheon DJ, Tong Y, Sim MS, Dering J, Berel D, Cui X, et al. A collagen-remodeling gene signature regulated by TGF-beta signaling is associated with metastasis and poor survival in serous ovarian cancer. Clin Cancer Res 2014;20:711-23.
74. Becker-Weimann S, Xiong G, Furuta S, Han J, Kuhn I, Akavia UD, et al. NFkB disrupts tissue polarity in 3D by preventing integration of microenvironmental signals. Oncotarget 2013;4:2010-20.
75. Chan CP, Kok KH, Jin DY. CREB3 subfamily transcription factors are not created equal: Recent insights from global analyses and animal models. Cell Biosci 2011;1:6.
76. Chiu YK, Lin IY, Su ST, Wang KH, Yang SY, Tsai DY, et al. Transcription factor ABF-1 suppresses plasma cell differentiation but facilitates memory B cell formation. J Immunol 2014;193:2207-17.
77. Gerhard GS, Hanson A, Wilhelmsen D, Piras IS, Still CD, Chu X, et al. AEBP1 expression increases with severity of fibrosis in NASH and is regulated by glucose, palmitate, and miR-372-3p. PLoS One 2019;14:e0219764.
78. Ye J. Transcription factors activated through RIP (regulated intramembrane proteolysis) and RAT (regulated alternative translocation). J Biol Chem 2020;295:10271-80.
79. Kurman RJ, Shih Ie M. Molecular pathogenesis and extraovarian origin of epithelial ovarian cancer--shifting the paradigm. Hum Pathol 2011;42:918-31.

80. Manzanedo I, Pereira F, Serrano A, Perez-Viejo E, Martinez-Torres B, Carrion L, et al. The use of cisplatin plus doxorubicin or paclitaxel in hyperthermic intraperitoneal chemotherapy (HIPEC) for stage IIIC or IV epithelial ovarian cancer: a comparative study. Clin Transl Oncol 2019;21:1357-63.
81. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet 2011;12:861-74.
82. Braga EA, Fridman MV, Moscovtsev AA, Filippova EA, Dmitriev AA, Kushlinskii NE. LncRNAs in Ovarian Cancer Progression, Metastasis, and Main Pathways: ceRNA and Alternative Mechanisms. Int J Mol Sci 2020;21.
83. Salamini-Montemurri M, Lamas-Maceiras M, Barreiro-Alonso A, Vizoso-Vazquez A, Rodriguez-Belmonte E, Quindos-Varela M, et al. The Challenges and Opportunities of LncRNAs in Ovarian Cancer Research and Clinical Use. Cancers (Basel) 2020;12.
84. Volders PJ, Anckaert J, Verheggen K, Nuytens J, Martens L, Mestdagh P, et al. LNCipedia 5: towards a reference set of human long non-coding RNAs. Nucleic Acids Res 2019;47:D135-D9.
85. Hu X, Feng Y, Zhang D, Zhao SD, Hu Z, Greshock J, et al. A functional genomic approach identifies FAL1 as an oncogenic long noncoding RNA that associates with BMI1 and represses p21 expression in cancer. Cancer Cell 2014;26:344-57.
86. Jeong S, Lee J, Kim D, Seol MY, Lee WK, Jeong JJ, et al. Relationship of Focally Amplified Long Noncoding on Chromosome 1 (FAL1) lncRNA with E2F Transcription Factors in Thyroid Cancer. Medicine (Baltimore) 2016;95:e2592.
87. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. Genes Dev 2009;23:1494-504.
88. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet 2009;10:155-9.
89. Huang W, Li H, Yu Q, Xiao W, Wang DO. LncRNA-mediated DNA methylation: an emerging mechanism in cancer and beyond. J Exp Clin Cancer Res 2022;41:100.
90. Diaz-Lagares A, Crujeiras AB, Lopez-Serra P, Soler M, Setien F, Goyal A, et al. Epigenetic inactivation of the p53-induced long noncoding RNA TP53 target 1 in human cancer. Proc Natl Acad Sci U S A 2016;113:E7535-E44.
91. Sun M, Jin FY, Xia R, Kong R, Li JH, Xu TP, et al. Decreased expression of long noncoding RNA GAS5 indicates a poor prognosis and promotes cell proliferation in gastric cancer. BMC Cancer 2014;14:319.
92. Tomlins SA, Bjartell A, Chinnaiyan AM, Jenster G, Nam RK, Rubin MA, et al. ETS gene fusions in prostate cancer: from discovery to daily clinical practice. Eur Urol 2009;56:275-86.
93. Qian C, Li D, Chen Y. ETS factors in prostate cancer. Cancer Lett 2022;530:181-9.
94. Sun Q, Hao Q, Prasanth KV. Nuclear Long Noncoding RNAs: Key Regulators of Gene Expression. Trends Genet 2018;34:142-57.
95. Wei JW, Huang K, Yang C, Kang CS. Non-coding RNAs as regulators in epigenetics (Review). Oncol Rep 2017;37:3-9.

96. Takai M, Terai Y, Kawaguchi H, Ashihara K, Fujiwara S, Tanaka T, et al. The EMT (epithelial-mesenchymal-transition)-related protein expression indicates the metastatic status and prognosis in patients with ovarian cancer. J Ovarian Res 2014;7:76.
97. Rahim B, O'Regan R. AR Signaling in Breast Cancer. Cancers (Basel) 2017;9.
98. Elattar A, Warburton KG, Mukhopadhyay A, Freer RM, Shaheen F, Cross P, et al. Androgen receptor expression is a biological marker for androgen sensitivity in high grade serous epithelial ovarian cancer. Gynecol Oncol 2012;124:142-7.
99. Vassilomanolakis M, Koumakis G, Barbounis V, Hajichristou H, Tsousis S, Efremidis A. A phase II study of flutamide in ovarian cancer. Oncology 1997;54:199-202.
100. Arlt W. Androgen therapy in women. Eur J Endocrinol 2006;154:1-11.
101. Labrie F. DHEA, important source of sex steroids in men and even more in women. Prog Brain Res 2010;182:97-148.
102. Schock H, Zeleniuch-Jacquotte A, Lundin E, Grankvist K, Lakso HA, Idahl A, et al. Hormone concentrations throughout uncomplicated pregnancies: a longitudinal study. BMC Pregnancy Childbirth 2016;16:146.
103. Sikdar S, Datta S. A novel statistical approach for identification of the master regulator transcription factor. BMC Bioinformatics 2017;18:79.
104. Mathas S, Janz M, Hummel F, Hummel M, Wollert-Wulf B, Lusatis S, et al. Intrinsic inhibition of transcription factor E2A by HLH proteins ABF-1 and Id2 mediates reprogramming of neoplastic B cells in Hodgkin lymphoma. Nat Immunol 2006;7:207-15.
105. Xing Y, Zhang Z, Chi F, Zhou Y, Ren S, Zhao Z, et al. AEBP1, a prognostic indicator, promotes colon adenocarcinoma cell growth and metastasis through the NF-kappaB pathway. Mol Carcinog 2019;58:1795-808.
106. Liu JY, Jiang L, Liu JJ, He T, Cui YH, Qian F, et al. AEBP1 promotes epithelial-mesenchymal transition of gastric cancer cells by activating the NF-kappaB pathway and predicts poor outcome of the patients. Sci Rep 2018;8:11955.
107. Lin Z, Wu Y, Xiao X, Zhang X, Wan J, Zheng T, et al. Pan-cancer analysis of CREB3L1 as biomarker in the prediction of prognosis and immunotherapeutic efficacy. Front Genet 2022;13:938510.
108. Puvvula PK, Desetty RD, Pineau P, Marchio A, Moon A, Dejean A, et al. Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit. Nat Commun 2014;5:5323.
109. White NM, Cabanski CR, Silva-Fisher JM, Dang HX, Govindan R, Maher CA. Transcriptome sequencing reveals altered long intergenic non-coding RNAs in lung cancer. Genome Biol 2014;15:429.
110. Liu AN, Qu HJ, Yu CY, Sun P. Knockdown of LINC01614 inhibits lung adenocarcinoma cell progression by up-regulating miR-217 and down-regulating FOXP1. J Cell Mol Med 2018;22:4034-44.
111. Wang D, Zhang H, Fang X, Cao D, Liu H. Pan-cancer analysis reveals the role of

long non-coding RNA LINC01614 as a highly cancer-dependent oncogene and biomarker. Oncol Lett 2020;20:1383-99.
112. Arriaga-Canon C, Contreras-Espinosa L, Aguilar-Villanueva S, Bargallo-Rocha E, Garcia-Gordillo JA, Cabrera-Galeana P, et al. The Clinical Utility of lncRNAs and Their Application as Molecular Biomarkers in Breast Cancer. Int J Mol Sci 2023;24.
113. Yan X, Hu Z, Feng Y, Hu X, Yuan J, Zhao SD, et al. Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers. Cancer Cell 2015;28:529-40.
114. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet 2015;47:199-208.
115. Goyal A, Myacheva K, Gross M, Klingenberg M, Duran Arque B, Diederichs S. Challenges of CRISPR/Cas9 applications for long non-coding RNA genes. Nucleic Acids Res 2017;45:e12.
116. Ito M, Miyata Y, Okada M. Current clinical trials with non-coding RNA-based therapeutics in malignant diseases: A systematic review. Transl Oncol 2023;31:101634.

ABSTRACT (IN KOREAN)

lncRNA 발현에 따른 장액성 난소암의 분류와 분자 생물학적 특성 분석

<지도교수 조 영 석>

연세대학교 대학원 의학과

정 선 향

난소암(OvCa)은 가장 치명적인 부인과 악성 종양으로, 매년 전 세계적으로 130,000명 이상이 사망한다. 가장 흔한 조직학적 아형인 장액성 난소암(HGS_OvCa)은 진단된 상피난소암 사례의 70%를 차지하며 특별한 증상이 없기 때문에 암 발생 후 복부 또는 복강 외부로 전이된 단계에서 처음 진단된다. 화학요법과 병행하는 수술이 치료의 기준으로 사용되지만 치료받은 환자의 75%는 약물 내성과 재발로 인해 낮은 생존율을 보인다. 이러한 치료적 한계를 극복하기 위해 새로운 표적이 발굴되고 있으며 그 중 하나로 제시되고 있는 것은 lncRNA이다. LncRNA는 다양한 암 종의 새로운 진단 및 치료 표적으로 연구되고 있으며 최근 연구에 따르면 lncRNA는 난소암에서 종양 진행, 전이, 에스트로겐 반응 및 약물 내성을 조절하는 데 중요한 역할을 한다. 본 연구의 목적은 장액성 난소암 RNA-seq 데이터를 이용한 군집 분류 결과를 바탕으로 lncRNA 발현에 의한 기능 분석을 통해 특성을 확인하는 것이며, 이러한 특성을 반영하는 요소를 식별하여 제시하는 것이다. 연구 목적에 따라 TCGA database의 367명 장액성 난소암 환자 데이터를 이용하여 CNMF 군집화를 수행하였고, 임상 분석을 통해 예후를 반영하는 군집을 선택했다. 각 군집의 기능 분석 결과를 토대로 "면역 그룹", "EMT 그룹", "Estrogen 반응 그룹", "EMT-Androgen 반응 그룹", "분화 그룹"으로 분류했다. 각 군집의 특성에 영향을 미치는 요인을 확인하기 위해 DNA 돌연변이, 체세포 복제수 변이, miRNA 및 DNA 메틸화 발현 분석을 진행했다.

또한, lncRNA와 mRNA를 조절하는 전사인자를 군집별로 분류하였다. J4 군집 중에서 고유 중심성을 기준으로 MSC, AEBP1, CREB3L1를 마스터 전사인자(Master Transcription factor, MTF)로 선택했다. 추가적으로, 선택된 마스터 전사인자보다 더 강한 영향을 미치는 7개의 lncRNA (LINC01614, LINC00702, AL109924.2, LINC02544, AL356417.2, AC112721.2, LINC01929)를 확인하였으며 해당 lncRNA가 마스터 전사 인자를 조절하고 EMT 관련 유전자를 조절하는 것을 실험 결과를 통해 확인했다. 이 연구를 통해 MTFs를 조절하는 7개의 lncRNA는 장액성 난소암에서 lncRNA의 전사인자 조절 메커니즘을 규명하는데 기여하고 개인 맞춤형 의학의 지표가 될 수 있음을 시사한다.

핵심되는 말: 난소암, 장액성 난소암, long noncoding RNA, biomarker, multi-omics data