Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Screening and identification of MSX1 for predicting progestin resistance in endometrial cancer

linlin yang
Shanghai Jiao Tong University School of Medicine https://orcid.org/0000-0002-4260-1934
Yunxia Cui
1Department of Gynecologic Oncology, The International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. 2Shanghai Municipal Key Clinical Specialty, Shanghai, China. 3Shanghai Key Laboratory of E
Ting Huang
1Department of Gynecologic Oncology, The International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. 2Shanghai Municipal Key Clinical Specialty, Shanghai, China. 3Shanghai Key Laboratory of E
Xiao Sun
1Department of Gynecologic Oncology, The International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. 2Shanghai Municipal Key Clinical Specialty, Shanghai, China. 3Shanghai Key Laboratory of E
Yudong Wang (wangyudong6688@126.com)
1Department of Gynecologic Oncology, The International Peace Maternity and Child Health Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. 2Shanghai Municipal Key Clinical Specialty, Shanghai, China. 3Shanghai Key Laboratory of E

Research

Keywords: Bioinformatic analysis, Progestin resistance, MSX1, Endometrial carcinoma

Posted Date: May 3rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-25541/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

Background Progestin resistance is a critical obstacle for endometrial conservative therapy. Therefore, the studies to acquire a more comprehensive understanding of the mechanisms are very important. However, the pivotal roles of essential molecules are still unexplored.

Methods We downloaded GSE121367 from the GEO database. The “limma” R language package was applied to identify differentially expressed genes (DEGs). We conducted Gene Set Enrichment Analysis (GSEA) and Gene Set Variation Analysis (GSVA) analysis. Protein–protein interaction was constructed by STRING and visualized in Cytoscape. The tumor immune microenvironment was explored by TISIDB database. Methylation validation and overall survival analysis was conducted by TCGA database. In addition, the upstream modulators of hub genes were predicted by miRTarBase and Network Analyst database.

Results A total of 3282 DEGs were identified and they were mostly enriched in cell adhesion pathway. We screened out ten hub genes including CDH1, JAG1, PTGES, EPCAM, CNTNAP2, TBX1, MSX1, KRT19, OAS1 and DAB2 among different groups, whose genomic alteration rates were low based on the current endometrial carcinoma sample sets. Has-miR-335-5p, has-miR-124-3p, MAZ and TFDP1 were the most prominent upstream regulators. The methylation status of CDH1, JAG1, EPCAM and MSX1 were decreased, corresponding to their high protein expression, which also predicted better overall survival. The homeobox protein of MSX1 showed significantly tissue specificity and better prognostic value.

Conclusions Our study identified the gene of MSX1 promised to be the specific indicator. This would shed new light on the underlying biological mechanism to overcome progestin resistance of endometrial cancer.

Background

Endometrial carcinoma (EC), which results from aberrant regeneration in terms of excessive growth of endometrial glands [1], accounts for 4.4% of carcinoma cases among women in 2018 [2] with more than 60,000 cases estimated in the United States in 2019 [3]. As for endometrial precancerous lesions including atypical hyperplasia or endometrial intraepithelial neoplasia and well-differentiated cancer, hysterectomy would not be a feasible and effective optimal choice for them and conservative treatment to preserve fertility for young patients is becoming significantly essential. While progestin remedy is commonly applied, approximately 30% of such patients don’t respond to the therapy, which causes poor effect for fertility preservation [4]. Till now, there is no effective solution to detect or predict which group of patients may respond to the progestin treatment.

A more comprehensive exploration of the precise molecular targets of progestin resistance would facilitate further improvements in disease diagnosis and would probe new biomarkers, continuous researches including ours have been carried out in the last decade to address the problem [5–7]. At present, the microarray technology and bioinformatics processing are considered to be promising tools for genomic analysis and could be well applied to identify genetic or epigenetic alterations in carcinogenesis and drug resistance, which become a necessary complement to experimental research [8]. Considering recent developments in open-access datasets like the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA), the exploration of key genes and detection of functional pathways have been implemented in EC [9]. The GEO and TCGA database contains thousands of clinical information and gene sequencing data, allowing for well-rounded analysis of various cancers. The gene expression profiling interactive analysis (GEPIA), acts as a web server for gene expression profiling, survival analysis and correlative analyses on the basis of different tumor characteristics such as grades or stages [10]. Epigenetics which covers fields of aberrant DNA methylation, dysregulated noncoding RNA and altered post-translational histone modification refers to heritable changes in gene expression which is not associated with an alteration in DNA sequence but plays an essential role in carcinogenesis and resistance detection [11]. Aberrant DNA methylation is most widely explored and may become an effective detection indicator [12]. Up to now, there was no relevant analysis of bioinformatics focused on progestin resistance of EC and the exploration of methylation marker of resistant genes was needed.

In this research, bioinformatics analysis was applied to reveal the differential expressed genes (DEGs) that leaded to progestin resistance based on microarray datasets from GEO databases and screen out significant hub genes. Gene-related microRNAs (miRNAs), transcription factors (TFs), methylation status and survival analysis as well as biological functions and pathways were also integrated to explore the mechanisms and potential therapeutic value of these DEGs in resistance by constructing networks. Tumor Immune Estimation Resource (TIMER), Gene Set Enrichment Analysis (GSEA) and Gene Set Variation Analysis (GSVA) were utilized to detect underlying biological mechanisms. Our results may help understand the pathogenesis of progestin resistance. Moreover, it may provide insight regarding the novel treatment for EC.

Methods

Microarray data and Data procession

The Gene Expression Omnibus (GEO) is a public repository for data storage. In the present study, the gene expression profiling data sets (GSE121367) was obtained from GEO database. It included endometrial cancer cell line Ishikawa and IshikawaPR which were established from Ishikawa cell as an acquired medroxyprogesterone acetate (MPA) resistant subline. Normalized data of GSE121367 was downloaded from GEO database and further processed by the “limma” R language package to identify differentially expressed genes (DEGs) between IshikawaPR and Ishikawa cell lines. P value < 0.05 and $|\log \text{fold change(FC)}| > 2$ were set as criteria to screen DEGs. The Cancer Genome Atlas (TCGA) database of EC was used to verify the

expression status and survival function of hub genes. Subsequently, Web-based software OmicShare (http://www.omicshare.com/tools) and Heml (http://heml.biocuckoo.org/down.php) were used to draw volcano plot and heatmap, respectively.

Functional and pathway enrichment analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID, http://davidd.ncifcrf.gov/) is an online program offering systematic and integrative functional annotation tools for researchers to explore biological meaning behind large list of genes [13]. In this study, the DAVID database and Metascape (http://metascape.org/) were introduced to perform both Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the top 250 DEGs [14]. $P < 0.05$ was set as the cut-off criterion.

Data analysis of Gene Set Enrichment Analysis (GSEA) and Gene Set Variation Analysis (GSVA)

In order to explore biological pathways of different groups, GSEA software (https://www.broadinstitute.org/gsea/index.jsp) was used. The annotated gene sets of c5.all.v7.0.symbols.gmt and h.all.v7.0.symbols.gmt were downloaded from the website and considered as the reference gene sets. The number of permutations was 1,000. Other parameters were set to default. Significant difference at P-value < 0.05 was defined as the cutoff criteria. Normalized enrichment score (NES) and false discovery rate (FDR) were applied to determine the statistical differences. The differential results were visualized by Enrichment Map plug-in of Cytoscape[15]. Furthermore, the “GSVA” R package was utilized to explore the pathways most associated with hub genes[16]. On the basis of the median expression of hub gene, 91 EC samples were divided into two groups (high expression and low expression). $P < 0.01$ was defined as statistically significant.

Protein–protein interaction (PPI) Network Construction and Hub Genes Screening

Firstly, online database Search Tool for the Retrieval of Interacting Genes (STRING, http://stringdb.org) was employed to explore the functional interactions between DEG-encoded proteins and build the PPI network [17]. PPI pairs with combined score ≥ 0.4 were considered as the threshold value. Subsequently, the PPI network was visualized by Cytoscape software.16 [18] and the degree of connectivity was also analyzed. Then the network relationship file was downloaded and the top 10 hub genes were identified by the analysis tool of its plug-in (degrees ranking of cytoHubba) [19].

Validation of the hub genes

Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn) is a web-based server to conduct out Hub genes expression analysis, correlation analysis and patient survival analysis [10]. Survival analyses of hub genes were conducted by log-rank tests and Kaplan–Meier survival curves were plotted. Then the mutation and DNA copy-number alterations of hub genes were investigated in cBioPortal (https://www.cbioportal.org/), the methylation status of hub genes was validated in Ualcan (http://ualcan.path.uab.edu/), which were based on TCGA analysis. Furthermore, the RNA expression level of hub genes in different carcinoma tissues were shown on basis of TCGA database and the protein level difference of hub genes were displayed by immunohistochemistry (IHC) on basis of the Human Protein Atlas database (HPA, https://www.proteinatlas.org/) [20].

Prediction of Relevant MicroRNAs and Transcriptional Factors (TFs) of Hub Genes

For the hub DEGs identified from the PPI network, the related miRNAs were predicted by miRTarBase (http://mirtarbase.mbc.nctu.edu.tw/php/download.php), which is the database aiming to provide hundreds of published experimentally validated miRNA–gene interactions [21].The Network Analyst (https://www.networkanalyst.ca/faces/home.xhtml) is designed to support integrative analysis of gene expression data through statistical, visual, and network-based approaches [22]. In this study, Network Analyst was introduced to predict hub generated TFs. A list of the hub genes were enrolled into the input area and proceeded step by step, finally the gene-related TFs as well as TFs–gene interactions pairs were presented. Then the results were visualized using the Cytoscape software.16 and the correlations were also evaluated based on GEPIA.

Results

Identification of Aberrantly Expressed Genes

After data preprocessing and quality evaluation, we obtained the expression matrices from the cell samples in the set GSE121367 and further processed with GEO2R tool. Results showed that a total number of 3282 commonly DEGs were screened out from the dataset GSE121367 with 1819 up-regulated genes (logFC > 2) and 1463 down-regulated genes (logFC < 2). The volcano plot is shown in Fig. 1a. The red plots represent the up-regulated genes, green plots represent down-regulated genes and black plots represent genes without differentially expression based on the cut-off criteria of P-value < 0.05 and |log fold change (FC)| > 2. The differential results of each comparison group were shown in Supplementary file 1. The top 250 of DEGs with significant fold change were also drawn with heatmap (Fig. 1b). Red pane shows high expression level and green pane shows low expression level.

Gene Functional Enrichment Analysis

To illustrate the biological classification of DEGs, GO term enrichment analysis was carried out using DAVID(Supplementary file 2). Three categories of GO terms including biological process (BP), cellular component (CC), and molecular function (MF) results were presented in Fig. 1c-e and Table 1–2. Results suggested that changes in BP of top 250 key genes were significantly enriched in “negative regulation of DNA binding”, “type I interferon signaling pathway” and “neuron migration” (Fig. 1c), As for CC term, these genes showed enrichment in “anchored component of membrane”, “extracellular space” and “multivesicular body” (Fig. 1d), Besides, MF term indicated enrichment predominantly at “protein dimerization activity”, “Notch binding” and “transporter activity” (Fig. 1e). To further analyze the DEG-enriched pathways, KEGG pathway analysis was subsequently conducted（Supplementary file 3）. As shown in Fig. 1f, it covered “Cell adhesion molecules pathway” and “Endocytosis pathway”. This functional investigation identified that these DEGs had close association with changes of DNA binding and cell adhesion pathway. Furthermore, we also analyzed the pathway of differential genes by the website of Metascape (Fig. 1g-h) which revealed that differential genes were enriched in “mesenchymal cell differential”, “cell-cell adhesion mediated by cadherin” and “negative regulation of DNA binding and cell proliferation”.

Table 1
GO enrichment analysis of DEGs

Category	Term	Count	P Value
GOTERM_BP_DIRECT	GO:0021983 ~ pituitary gland development	5	2.52E-04
GOTERM_BP_DIRECT	GO:0043392 ~ negative regulation of DNA binding	5	2.52E-04
GOTERM_BP_DIRECT	GO:0060337 ~ type I interferon signaling pathway	6	7.45E-04
GOTERM_BP_DIRECT	GO:0010628 ~ positive regulation of gene expression	11	7.56E-04
GOTERM_BP_DIRECT	GO:0001764 ~ neuron migration	7	0.001169
GOTERM_CC_DIRECT	GO:0031225 ~ anchored component of membrane	8	0.000235
GOTERM_CC_DIRECT	GO:0005615 ~ extracellular space	29	0.000690
GOTERM_CC_DIRECT	GO:0005771 ~ multivesicular body	4	0.001670
GOTERM_CC_DIRECT	GO:0005576 ~ extracellular region	31	0.002465
GOTERM_CC_DIRECT	GO:0016324 ~ apical plasma membrane	10	0.004640
GOTERM_MF_DIRECT	GO:0046983 ~ protein dimerization activity	7	0.006223
GOTERM_MF_DIRECT	GO:0008201 ~ heparin binding	7	0.008454
GOTERM_MF_DIRECT	GO:0005112 ~ Notch binding	3	0.016286
GOTERM_MF_DIRECT	GO:0046872 ~ metal ion binding	34	0.016772
GOTERM_MF_DIRECT	GO:0005215 ~ transporter activity	7	0.023963

Table 2
Pathway enrichment analysis of DEGs

Category	Term	Count	P Value
KEGG_PATHWAY	hsa04514:Cell adhesion molecules (CAMs)	7	0.0046
KEGG_PATHWAY	hsa04612:Antigen processing and presentation	5	0.0096
KEGG_PATHWAY	hsa05168:Herpes simplex infection	7	0.0153
KEGG_PATHWAY	hsa04144:Endocytosis	8	0.0166
KEGG_PATHWAY	hsa04145:Phagosome	6	0.0246

Data Processing and Gene Set Enrichment Analysis (GSEA)

Although differential expression of individual genes could play a critical role in mechanistic aspects of cellular regulation, many compounds and genes are regulated complicatedly. For the sake of categorizing such modules of cellular regulation, bioinformatics approaches for “gene set enrichment” (GSEA) statistics have been developed [23]. The consequences of GSEA analysis revealed that 428 gene sets were upregulated in the IshikawaPR cell line with P-value < 0.05, among which 145 gene sets were significantly enriched at nominal P-value < 0.01. A total of 116 gene sets were upregulated in the Ishikawa cell line with P-value < 0.05, among which 34 gene sets were significantly enriched at nominal P-value < 0.01 (Supplementary file 4). As is shown in Table 3, pathways including interferon gamma response, TNF-a signaling via NF-KB, epithelial mesenchymal transition, interleukin1 beta production and negative regulation of response to drug were significantly enriched in IshikawaPR cell line (Fig. 2b). While in Ishikawa cell line (Table 4), the consequences showed that pathways about mesenchymal to epithelial transition, negative regulation of insulin secretion, apical junction assembly and plasma membrane receptor complex compounds were highly enriched (Fig. 2c). All the differential results of

gene sets were visualized by Enrichment Map plug-in of Cytoscape (Fig. 2a). Altogether, these data suggested that when Ishikawa cells were stimulated and selected by MPA for almost 10 months, the functions of cell signal transduction such as nuclear receptor activity and cytokine biosynthetic process including interferon gamma, interleukin1 production and epithelial cell polarity were dramatically changed.

Table 3
Upregulated gene sets in the IshikawaPR cell line

Gene Sets	SIZE	ES	NES	NOM p-value	FDR q-value
Interferon gamma response	196	0.55	2.07	0.00	0.00
TNF-a signaling via NF-KB	196	0.46	1.72	0.00	0.01
Epithelial mesenchymal transition	195	0.43	1.62	0.00	0.01
Hypoxia	191	0.39	1.47	0.00	0.04
Complement	193	0.44	1.67	0.00	0.01
Negative regulation of regulated secretory pathway	23	0.68	1.73	0.00	0.27
Chronic inflammatory response	18	0.69	1.72	0.01	0.27
Interleukin1 production	90	0.49	1.69	0.00	0.30
Interferon gamma mediated signaling pathway	87	0.51	1.74	0.00	0.30
Negative regulation of response to drug	25	0.63	1.68	0.00	0.30

Table 4
Upregulated gene sets in the Ishikawa cell line

Gene Sets	SIZE	ES	NES	NOM p-value	FDR q-value
ATP dependent microtubule motor activity plus end directed	26	-0.65	-1.76	0.00	0.62
Mesenchymal to epithelial transition	20	-0.69	-1.75	0.00	0.58
Phospholipid catabolic process	38	-0.59	-1.75	0.00	0.55
Respiratory chain complex IV	15	-0.72	-1.73	0.01	0.56
Transcytosis	18	-0.69	-1.72	0.01	0.52
Positive regulation of protein localization to cell periphery	60	-0.54	-1.69	0.00	0.59
Negative regulation of insulin secretion	36	-0.55	-1.59	0.01	0.66
Apical junction assembly	58	-0.48	-1.52	0.01	0.88
Cell cell adhesion via plasma membrane adhesion molecules	253	-0.36	-1.39	0.01	0.94
Plasma membrane receptor complex	184	-0.36	-1.37	0.01	0.94

PPI Network Construction of DEGs and Verification of Hub genes

The top 250 DEGs ($P < 0.05$) of GSE121367 were used to construct PPI network by the database of STRING and visualized in Cytoscape software (Fig. 3a). The red color of a node reflects the upregulated gene and the green means the downregulated gene. The size of the node indicates the connectivity degree and the width of edge displays the combined score. PPI network analysis had been studied by using the threshold value of confidence > 0.4 and connectivity degree ≥ 10. In this network, it contained 159 nodes and 244 edges. The plug-in of cytoHubba in Cytoscape was used to screen hub genes, then a significant submodule was obtained (Fig. 3b), from which we chose the hub genes with high scores. Finally, 10 common hub genes (CDH1, JAG1, PTGES, EPCAM, CNTNAP2, TBX1, MSX1, KRT19, OAS1 and DAB2) were identified in the subnetwork (Table 5 and Table 6). Next, we observed the mutation and DNA copy number alterations of 10 key genes (Fig. 3c). As is shown in the OncoPrint tab, it demonstrated a visual summary of the different alterations of ten hub genes across all sets of uterine corpus endometrial carcinoma samples based on a query of the ten genes. Each row represents a gene, and each column represents a tumor sample. Red bars indicate gene amplifications, blue bars are deep deletions, grey bars are no alterations and green squares are missense mutation. The genomic alteration rates of hub genes were < 10% in all enrolled endometrial cancer cases. Furthermore, 10 hub genes were validated in database TCGA to compare gene expression between endometrial carcinoma samples and normal samples (Fig. 3d). The genes of CDH1, EPCAM, MSX1, KRT19 and OAS1 were overexpressed in tumor tissues, while genes including JAG1, TBX1 and DAB2 were downregulated in cancer tissues. There were no differences in the expression of PTGES and CNTNAP2 in cancer and normal samples (results not shown). In addition, we explored the expression profiles of ten hub genes in other cancers by GEPIA database. As shown in Fig. 3e, above ten genes were all expressed in other tissues, however, significantly expressed in paired endometrial tissues.

Table 5
Identification of hub genes by cytoHubba.

name	Betweenness	Bottle Neck	Closeness	Clustering Coefficient	Degree	DMNC	Ec Centricity	EPC	MCC	MNC	Radiality	Stress
CDH1	7440.58	76	66.299	0.085	27	0.227	0.100	43.793	96	17	11.806	17516
JAG1	1707.27	14	52.316	0.321	8	0.428	0.100	39.736	26	6	11.457	3744
PTGES	510.00	1	38.399	1.000	2	0.308	0.082	21.129	2	2	10.558	500
EPCAM	423.05	6	52.267	0.378	10	0.339	0.090	41.186	58	10	11.312	1750
CNTNAP2	2582.66	10	38.916	0.000	5	0.000	0.100	15.668	5	1	10.609	4672
TBX1	1156.20	12	52.016	0.333	7	0.454	0.100	39.344	20	5	11.438	3108
MSX1	944.50	4	43.892	0.167	4	0.308	0.100	25.883	4	2	11.039	1976
KRT19	405.98	5	49.634	0.389	9	0.334	0.090	39.797	46	9	11.198	1452
OAS1	81.05	1	39.275	0.500	8	0.408	0.100	29.613	58	8	10.419	450
DAB2	751.752	2	45.945	0.267	6	0.379	0.112	32.237	10	4	11.191	1378

Table 6
The information of ten hub genes.

Gene name	logFC	P Value
CDH1(cadherin 1)	-5.41	2.16E-10
JAG1(jagged 1)	-5.44	2.35E-10
PTGES(prostaglandin E synthase)	7.77	2.42E-11
EPCAM(epithelial cell adhesion molecule)	-5.39	2.66E-10
CNTNAP2(contactin associated protein-like 2)	-7.10	1.76E-11
TBX1(T-box 1)	7.61	2.71E-10
MSX1(msh homeobox 1)	-5.45	5.79E-10
KRT19(keratin 19)	-6.60	4.00E-10
OAS1(2'-5'-oligoadenylate synthetase 1)	5.83	1.24E-10
DAB2(disabled homolog 2, mitogen-responsive phosphoprotein)	9.00	3.02E-10

Gene-Associated MicroRNAs Network Analysis

To explore the potential upstream regulator of hub genes, predicted miRNAs of hub gene were analyzed by miRTarBase database. Main miRNAs with interactions of more than two genes were listed in Table 7. Moreover, Cytoscape was used to construct the hub gene-relevant miRNAs network (Fig. 4a). There were a total number of 8 genes, 118 miRNAs and 128 genemiRNA pairs contained in the network. Some miRNAs were found to play a critical role in regulating essential genes. Has-miR-335-5p was predicted to regulate PTGES, OAS1, KRT19 and DAB2, has-miR-26b-5p may regulate DAB2, CDH1 and JAG1. Furthermore, genes of PTGES and JAG1 were regulated by has-miR-124-3p, whose high expression may be associated with worse survival (Supplementary Fig. 1), suggesting that it may be involved in tumor resistance and may become a prognostic indicator for endometrial cancer.

Table 7
The main related MicroRNAs of hub genes

MicroRNAs	Genes	Count
has-miR-335-5p	PTGES, OAS1, KRT19, DAB2	4
has-miR-26b-5p	DAB2, CDH1, JAG1	3
has-miR-9500	MSX1, PTGES	2
has-miR-124-3p	PTGES, JAG1	2
has-miR-129-5p	DAB2, CDH1	2
has-miR-199a-5p	CDH1, JAG1	2
has-miR-193b-3p	CDH1, KRT19	2

Core Transcriptional Factors Mediation Network Analysis of Hub Genes

To identify the transcriptional regulation of the hub genes and assess the effect of TFs on the expression of the hub genes, gene-TFs regulation network was performed by using a Network Analyst network-based service. Totally, 143 TFs were included in the network, constructing 203 gene–TFs interaction pairs (Fig. 4b). In this network, MAZ was considered as the key TF to regulate five hub genes: CDH1, EPCAM, KRT19, MSX1, and TBX1. In addition, TFDP1 plays a second important role in regulating CDH1, CNTNAP2, KRT19 and MSX1 (Table 8). Furthermore, we explored the correlation of hub genes and core TFs of MAZ and TFDP1in endometrial carcinoma using TCGA datasets, respectively. From these results, we found that MAZ and TFDP1 had positive correlations with CDH1, EPCAM, MSX1, KRT19, OAS1 and had negative correlations with JAG1, TBX1 and DAB2 (Fig. 5a-b). Additionally, we found that MAZ most positively related with EPCAM and negatively related to DAB2. While TFDP1 was positively associated with gene CDH1 and negatively associated with gene JAG1.

Table 8
The main related TFs of hub genes

TFs	Genes	Count
MAZ	CDH1, EPCAM, KRT19, MSX1, TBX1	5
TFDP1	CDH1, CNTNAP2, KRT19, MSX1	4
PPARG	CDH1, EPCAM, KRT19	3
NR2F1	CDH1, KRT19, TBX1	3
DMAP1	CDH1, KRT19, OAS1	3
EZH2	CDH1, MSX1, TBX1	3
ELF3	CDH1, EPCAM, KRT19	3
CHD1	CDH1, KRT19, PTGES	3
BCOR	EPCAM, PTGES, TBX1	3
E2F5	KRT19, PTGES, TBX1	3
JUND	EPCAM, KRT19, TBX1	3
SMARCA5	KRT19, PTGES, TBX1	3

Methylation Status and Expression Validation of Hub Genes in HPA

The initiation of cancer resistance was controlled by both genetic and epigenetic events. Epigenetic changes also make an important impact on occurrence of drug resistance. Therefore, we decided to detect the methylation status of hub genes. As is performed in Fig. 6a, the genes of CDH1, JAG1, EPCAM and MSX1 were aberrantly methylated, which were in consistent with their expression respectively, on the basis of Ualcan website (Fig. 6b). In addition, Immunohistochemistry (IHC) staining obtained from the HPA database showed the dysregulation of the expression of hub genes (Fig. 6c), among which CDH1, EPCAM and MSX1 were upregulated in endometrial carcinoma samples, while the expression of JAG1 were downregulated.

Tissue Specificity Analysis and Prognostic Value Evaluation of Hub Genes

To further evaluate the expression level of hub genes in different human carcinoma tissues, we explored the HPA website. From Fig. 7a-d, it indicated that the selected four hub genes had various RNA expression levels in different cancer tissues including glioma, thyroid cancer, lung cancer, colorectal cancer, head and neck cancer, stomach cancer, liver cancer, pancreatic cancer, renal cancer, urothelial cancer, prostate cancer, testis cancer, breast

cancer, cervical cancer, endometrial cancer, ovarian cancer and melanoma. Moreover, CDH1 and EPCAM displayed moderate expression level in carcinoma of endometrium, while JAG1 had relatively low level. Nevertheless, MSX1 displayed the highest expression level in endometrial cancer, which showed high tissue specificity. Meanwhile, we explored the prognostic values of the four essential genes further, which were obtained in GEPIA and displayed in Fig. 7e-h. Overall survival for endometrial cancer patients was analyzed in correspondence with the low or high expression of each gene. As is shown, high mRNA expression of CDH1 (P = 0.01) was associated with better overall survival for endometrial cancer patients, along with EPCAM (P = 0.045), JAG1 (P = 0.02), MSX1 (P = 0.001).

Validation of hub gene MSX1 based on the TCGA dataset

The relationship between MSX1 expression and different pathological grade was measured, which suggested that mRNA expression of MSX1 were significantly correlated with tumor grades (Fig. 8a). To further investigate the potential functions of MSX1, we performed GSVA on the TCGA data. As shown in Fig. 8b, genes in high expression groups of MSX1 were enriched in “positive regulation of intrinsic apoptotic signaling pathway by p53 class mediator” and “epithelial to mesenchymal transition” pathways. Furthermore, we enrolled survival and follow-up data from TCGA cohort. It suggested that high expression of MSX1 were significantly associated with favorable prognosis in EC patients (Fig. 8c). We also used the tumor-immune system interactions (TISIDB) online database to detect the expression and prognostic value of MSX1 in other types of tumors (Fig. 8d). It showed better prediction function in endometrial cancer (Fig. 8e). In addition, we explored association of hub genes’ expression with immune infiltration, the results showed that no or weak associations were observed between MSX1 and infiltration of lymphocytes (Supplementary Fig. 2), while we found MSX1 was positively associated with immunostimulator NT5E (Fig. 8f-g).

Discussion

The majority of women in reproductive period with endometrial precancerous lesion and well-differentiated endometrial neoplasm have an intense willing to preserve fertility. However, when closely following up younger patients with the progestin conservative therapy, clinicians observed that more than 30% of them responded poorly to the treatment [4]. The major account for such a high failure rate is that the potential molecular mechanisms of drug resistance remain unclear. Our team has been keen on this research theme for several years [5–7, 24, 25]. On the basis of our previous work, we recently conducted bioinformatics research replying on the dataset from GEO website. By comparing endometrial adenocarcinoma cell line Ishikawa with its counterpart IshikawaPR (MPA resistant cell line), we systematically analyzed their significant different genes, relevant molecular and pathways as well as their DNA copy number and the status of methylation in order to identify the essential candidate genes that promoted the carcinogenesis and progestin resistance.

In our present research, among the 24384 DEGs, we utilized the top 250 genes to conduct Functional and pathway enrichment analysis and GSEA analysis, which may provide novel insights for clarifying pathogenesis of progestin resistant. As was exhibited in DAVID, the genes were enriched in biological processes (BP) of negative regulation of DNA binding, type I interferon signaling pathway, neuron migration and axonogenesis involved in innervation, which made complementary remarks to previous points that EGF/EGFR [26] and insulin [27] signaling pathways may lead to progestin resistance. Meanwhile, KEGG pathway showed that cell adhesion molecules pathway and endocytosis signals were involved. Furthermore, the results of GSEA reported that nuclear receptor activity, chronic inflammatory response and endothelial cell proliferation pathway were enriched in MPA resistant cells, which were significantly different from MPA sensitive cells, suggesting that prolonged progestin treatment may result in changes of cell membranes and nuclear receptors activity, affect signaling transduction and induce peripheral inflammation response and neurological development. As was said, intestinal bacteria can induce a chronic subclinical inflammatory process, leading to insulin resistance [28], confirming the association between inflammation and resistance, while how inflammatory cytokines evoked progestin resistance remained uncovered and the role neurological factors played in drug resistance provided a new perspective for us to explore.

PPI network of DEGs demonstrated the functional correlations, in which hub genes were screened out. Afterwards, we investigated the mutation status of essential genes, all mutation rates were less than 10%, suggesting that their expressions were regulated by other factors, other than genomic mutations. Then, the protein expressions were testified in TCGA database, genes with statistically significant differences were enumerated. while PTGES and CNTNAP2 had no remarkable difference between normal and tumor samples, so the results were not shown. Moreover, the relevant microRNA and transcription factors were detected and has-miR-335-5p, has-miR-124-3p, MAZ and TFDP1 played a vital role. As was reported, has-miR-335-5p inhibits invasion and metastasis of thyroid cancer cells [29], breast cancer cells [30] and non-small cell lung cancer [31]. In endometrial stromal sarcomas (ESS), has-miR-335-5p was more highly expressed in patients with tumor metastasis and relapse [32]. In this study, high expression of has-miR-124-3p predicted worse survival, which was consistent with its effect on hepatocellular carcinoma [33]. MAZ was thought to act as a therapeutic target for aerobic glycolysis and the progression of neuroblastoma [34] and prostate cancer bone metastasis [35]. TFDP1 was involved in colon cancer stemness and cell cycle progression [36] and in the endometrium of women with deep infiltrating endometriosis (DIE) [37]. The relevant microRNA and interactions between hub genes and core TFs MAZ and TFDP1 had already been verified, suggesting they may participate in the formation of progestin resistance, while the detailed regulated mechanisms between TFs and hub genes needed to be further confirmed both in vitro and vivo.

Epigenetic processes, such as DNA methylation, are known to regulate specific gene expression [38]. Therefore, in the current study, we researched the methylation status of key genes and presented the significantly different genes such as CDH1, JAG1, EPCAM and MSX1, which corresponded to their

expressions based on TCGA datasets and HPA website. According to the RNA expressions of four hub genes in different carcinoma organs depending on TCGA database, both CDH1 and EPCAM showed moderate expression level, while the level of JAG1 was low. Meanwhile, MSX1 was reported to demonstrate the highest expression in endometrial neoplasm, showing its high tissue specificity. Additionally, existing researches reported that MSX1 inhibited the growth and metastasis of breast cancer cells and was frequently silenced by promoter methylation [39], and that MSX1 induced G0/G1 arrest and apoptosis in cervical cancer [40] and that epigenetic regulation of MSX1 associated with platinum-resistant disease in high-grade serous epithelial ovarian cancer [41]. However, the expression and function of MSX1 had not been explored in endometrial resistant lesions. Therefore, although all of the four genes were associated with patients' overall survival, due to its high tissue specificity, MSX1 may become a promising tissue specific marker for endometrial cancer initiation and predict the therapy effect of progestin.

We further verified the expression of MSX1 in a larger TCGA cohort. The results confirmed our findings that MSX1 had significant prediction value in EC and may make an impact on cancer progression through the P53 pathway. We also explored associations of MSX1's expression with the tumor microenvironment and found the interaction between MSX1 and immunostimulator NT5E [42]. However, a series of experimental verification about our found mechanism need to be investigated and its immune-reinforcing effect needs further detection.

In conclusion, a comprehensive analysis of the hub genes based on the GEO dataset will likely shed new light on progestin therapy. Our study highlighted a novel understanding of the potential biological mechanism in progestin resistance and identified the homeobox gene MSX1 as a biomarker to detect the sensitivity and efficacy of progestin treatment.

Conclusions

In summary, by using comprehensive bioinformatics analysis, we have identified DEGs and demonstrated for the first time that MSX1 is likely one of the main molecular indicator of progestin resistance in endometrial cancer. If validated in a larger cohort, MSX1 may become a useful target to detect progestin therapy effect, which is beneficial for younger patients who want to preserve fertility.

Abbreviations

EC: endometrial cancer; GEO: Gene Expression Omnibus; DEGs: differentially expressed genes; STRING: Search Tool for the Retrieval of Interacting Genes Database; PPI: protein–protein interaction; MCODE: Molecular Complex Detection; GSEA: gene set enrichment analysis; IHC: Immunohistochemistry.

Declarations

Acknowledgements

Not applicable.

Authors' contributions

Linlin Yang designed the experiments and supervised the completion of this work. Yunxia Cui and Ting Huang downloaded the information of relevant miRNA and TFs. Xiao Sun and Yudong Wang reviewed and edited the manuscript. All authors have read and approved the final manuscript.

Funding

This study was supported by Shanghai Municipal Key Clinical Specialty (No. shslczdzk06302), National Natural Science Foundation of China (No. 81172477, 81402135), the Project of the Science and Technology Commission of Shanghai Municipality (No. 17441907400) and Shanghai Jiao Tong University Medicine-Engineering Fund (No. YG2017MS41).

Availability of data and materials

All data analyzed during this study are included in this published article.

Ethics approval and consent to participate

Not required.

Consent for publication

This article is original and has not already been published in another journal.

Conflict of Interests

The authors declare that they have no competing interests

References

1. Syed SM, Kumar M, Ghosh A, Tomasetig F, Ali A, Whan RM, et al. Endometrial Axin2(+) Cells Drive Epithelial Homeostasis, Regeneration, and Cancer following Oncogenic Transformation. Cell stem cell. 2020;26(1):64-80.e13.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018;68(6):394-424.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA: a cancer journal for clinicians. 2019;69(1):7-34.
4. Perez-Medina T, Bajo J, Folgueira G, Haya J, Ortega P. Atypical endometrial hyperplasia treatment with progestogens and gonadotropin-releasing hormone analogues: long-term follow-up. Gynecologic oncology. 1999;73(2):299-304.
5. Wang J, Sun X, Zhang H, Wang Y, Li Y. MPA influences tumor cell proliferation, migration, and invasion induced by RANKL through PRB involving the MAPK pathway in endometrial cancer. Oncology reports. 2015;33(2):799-809.
6. Wang J, Liu Y, Wang L, Sun X, Wang Y. Clinical prognostic significance and pro-metastatic activity of RANK/RANKL via the AKT pathway in endometrial cancer. Oncotarget. 2016;7(5):5564-75.
7. Liu Y, Wang J, Ni T, Wang L, Wang Y, Sun X. CCL20 mediates RANK/RANKL-induced epithelial-mesenchymal transition in endometrial cancer cells. Oncotarget. 2016;7(18):25328-39.
8. Iizuka N, Oka M, Yamada-Okabe H, Nishida M, Maeda Y, Mori N, et al. Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection. Lancet (London, England). 2003;361(9361):923-9.
9. Zhang W, Gao L, Wang C, Wang S, Sun D, Li X, et al. Combining Bioinformatics and Experiments to Identify and Verify Key Genes with Prognostic Values in Endometrial Carcinoma. Journal of Cancer. 2020;11(3):716-32.
10. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic acids research. 2017;45(W1):W98-w102.
11. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012;150(1):12-27.
12. Ying J, Xu T, Wang Q, Ye J, Lyu J. Exploration of DNA methylation markers for diagnosis and prognosis of patients with endometrial cancer. Epigenetics. 2018;13(5):490-504.
13. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009;4(1):44-57.
14. Du J, Yuan Z, Ma Z, Song J, Xie X, Chen Y. KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. Molecular bioSystems. 2014;10(9):2441-7.
15. Huo X, Sun H, Liu Q, Ma X, Peng P, Yu M, et al. Clinical and Expression Significance of AKT1 by Co-expression Network Analysis in Endometrial Cancer. Frontiers in oncology. 2019;9:1147.
16. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC bioinformatics. 2013;14:7.
17. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic acids research. 2011;39(Database issue):D561-8.
18. 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 research. 2003;13(11):2498-504.
19. Yang G, Zhang Y, Yang J. A Five-microRNA Signature as Prognostic Biomarker in Colorectal Cancer by Bioinformatics Analysis. Frontiers in oncology. 2019;9:1207.
20. Li B, Pu K, Wu X. Identifying novel biomarkers in hepatocellular carcinoma by weighted gene co-expression network analysis. Journal of cellular biochemistry. 2019.
21. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic acids research. 2018;46(D1):D296-d302.
22. Xia J, Gill EE, Hancock RE. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. Nature protocols. 2015;10(6):823-44.
23. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(43):15545-50.
24. Meng Q, Sun X, Wang J, Wang Y, Wang L. The important application of thioridazine in the endometrial cancer. American journal of translational research. 2016;8(6):2767-75.
25. Gu S, Ni T, Wang J, Liu Y, Fan Q, Wang Y, et al. CD47 Blockade Inhibits Tumor Progression through Promoting Phagocytosis of Tumor Cells by M2 Polarized Macrophages in Endometrial Cancer. Journal of immunology research. 2018;2018:6156757.
26. Xu Y, Tong J, Ai Z, Wang J, Teng Y. Epidermal growth factor receptor signaling pathway involved in progestin-resistance of human endometrial carcinoma: In a mouse model. The journal of obstetrics and gynaecology research. 2012;38(12):1358-66.

27. Shen ZQ, Zhu HT, Lin JF. Reverse of progestin-resistant atypical endometrial hyperplasia by metformin and oral contraceptives. Obstetrics and gynecology. 2008;112(2 Pt 2):465-7.
28. Saad MJ, Santos A, Prada PO. Linking Gut Microbiota and Inflammation to Obesity and Insulin Resistance. Physiology (Bethesda, Md). 2016;31(4):283-93.
29. Luo L, Xia L, Zha B, Zuo C, Deng D, Chen M, et al. miR-335-5p targeting ICAM-1 inhibits invasion and metastasis of thyroid cancer cells. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie. 2018;106:983-90.
30. Jia Q, Ye L, Xu S, Xiao H, Xu S, Shi Z, et al. Circular RNA 0007255 regulates the progression of breast cancer through miR-335-5p/SIX2 axis. Thoracic cancer. 2020.
31. Du W, Tang H, Lei Z, Zhu J, Zeng Y, Liu Z, et al. miR-335-5p inhibits TGF-beta1-induced epithelial-mesenchymal transition in non-small cell lung cancer via ROCK1. Respiratory research. 2019;20(1):225.
32. Gonzalez Dos Anjos L, de Almeida BC, Gomes de Almeida T, Mourao Lavorato Rocha A, De Nardo Maffazioli G, Soares FA, et al. Could miRNA Signatures be Useful for Predicting Uterine Sarcoma and Carcinosarcoma Prognosis and Treatment? Cancers. 2018;10(9).
33. Long HD, Ma YS, Yang HQ, Xue SB, Liu JB, Yu F, et al. Reduced hsa-miR-124-3p levels are associated with the poor survival of patients with hepatocellular carcinoma. Molecular biology reports. 2018;45(6):2615-23.
34. Li H, Yang F, Hu A, Wang X, Fang E, Chen Y, et al. Therapeutic targeting of circ-CUX1/EWSR1/MAZ axis inhibits glycolysis and neuroblastoma progression. EMBO molecular medicine. 2019;11(12):e10835.
35. Yang Q, Lang C, Wu Z, Dai Y, He S, Guo W, et al. MAZ promotes prostate cancer bone metastasis through transcriptionally activating the KRas-dependent RalGEFs pathway. Journal of experimental & clinical cancer research : CR. 2019;38(1):391.
36. Morimoto Y, Mizushima T, Wu X, Okuzaki D, Yokoyama Y, Inoue A, et al. miR-4711-5p regulates cancer stemness and cell cycle progression via KLF5, MDM2 and TFDP1 in colon cancer cells. British journal of cancer. 2020;122(7):1037-49.
37. Jibrim RLM, de Carvalho CV, Invitti AL, Schor E. Expression of the TFDP1 gene in the endometrium of women with deep infiltrating endometriosis. Gynecological endocrinology : the official journal of the International Society of Gynecological Endocrinology. 2019;35(6):490-3.
38. Bai M, Yang L, Liao H, Liang X, Xie B, Xiong J, et al. Metformin sensitizes endometrial cancer cells to chemotherapy through IDH1-induced Nrf2 expression via an epigenetic mechanism. Oncogene. 2018;37(42):5666-81.
39. Yue Y, Yuan Y, Li L, Fan J, Li C, Peng W, et al. Homeobox protein MSX1 inhibits the growth and metastasis of breast cancer cells and is frequently silenced by promoter methylation. International journal of molecular medicine. 2018;41(5):2986-96.
40. Yue Y, Zhou K, Li J, Jiang S, Li C, Men H. MSX1 induces G0/G1 arrest and apoptosis by suppressing Notch signaling and is frequently methylated in cervical cancer. OncoTargets and therapy. 2018;11:4769-80.
41. Bonito NA, Borley J, Wilhelm-Benartzi CS, Ghaem-Maghami S, Brown R. Epigenetic Regulation of the Homeobox Gene MSX1 Associates with Platinum-Resistant Disease in High-Grade Serous Epithelial Ovarian Cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2016;22(12):3097-104.
42. Zhu J, Zeng Y, Li W, Qin H, Lei Z, Shen D, et al. CD73/NT5E is a target of miR-30a-5p and plays an important role in the pathogenesis of non-small cell lung cancer. Molecular cancer. 2017;16(1):34.

Figures

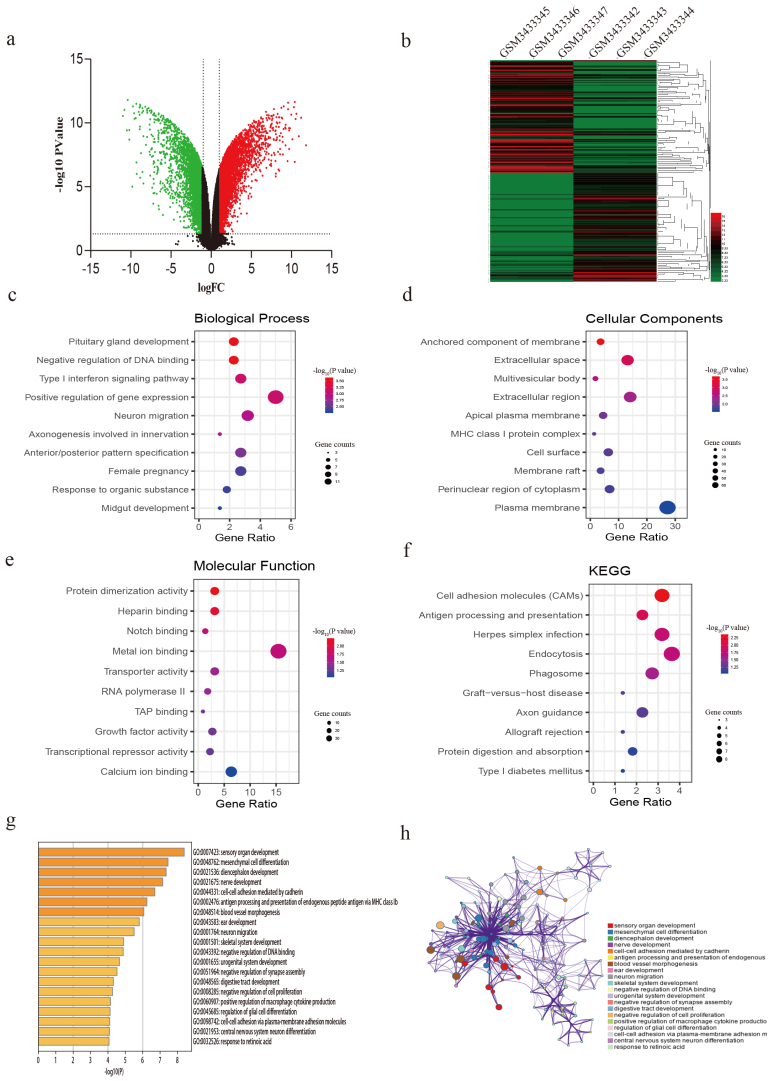

Figure 1

Identification of DEGs from GSE121367 dataset and functional enrichment analysis. a Volcano plot of the DEGs. Red dots and green dots represent the upregulated and downregulated genes respectively; black represent genes with no differential expression based on the threshold of P value <0.05 and |log FC| > 2.0. b Heatmap of top 250 DEGs. Gene expression levels were shown by color bar. Red color denotes high level and green color denotes low level. c Biological process (BP). d Cellular component (CC). e Molecular function (MF). f KEGG pathways. g Boxplot of enriched terms across DEGs, colored by P-values. h Network of enriched terms, colored by cluster ID, where nodes that share the same cluster ID are typically close to each other.

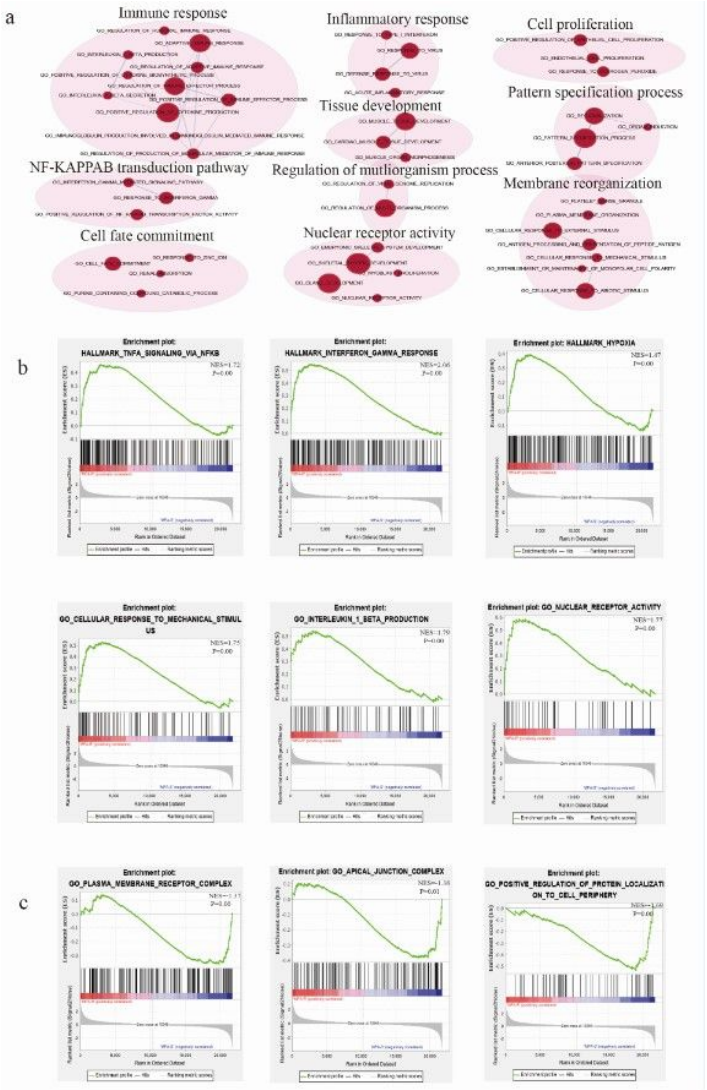

Figure 2

The results of GSEA analysis. a The pathway network of the group. The red dots represent upregulated pathways. b Significantly enriched gene sets in IshikawaPR cell line. c Significantly enriched gene sets in Ishikawa cell line. MPA-R represents cell line of IshikawaPR; MPA-S represents cell line of Ishikawa; NES, normalized enrichment score.

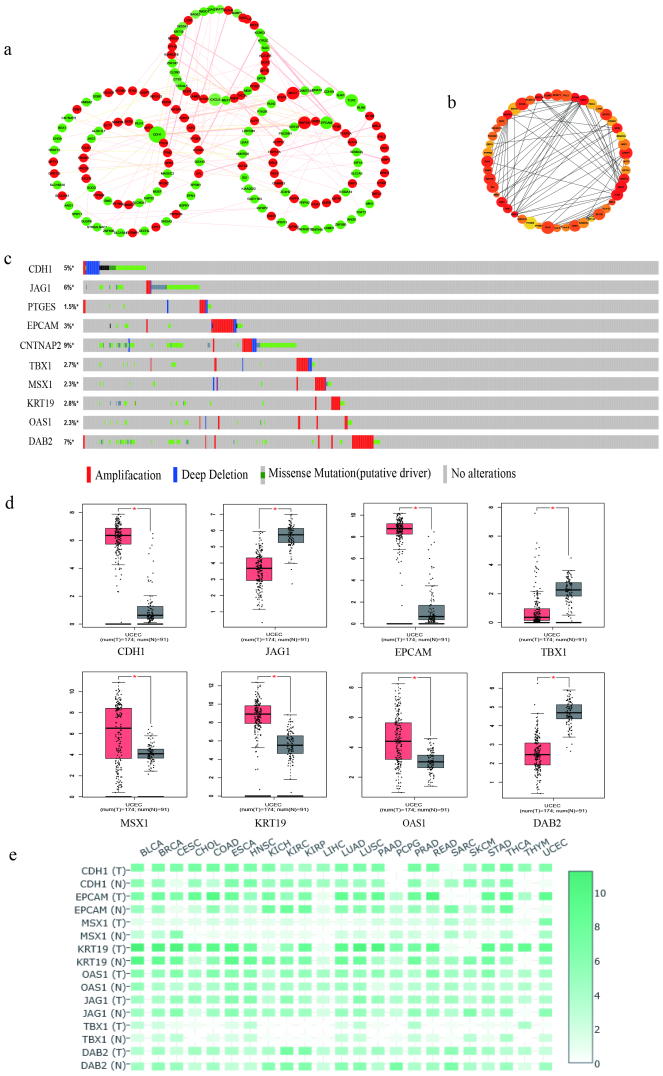

Figure 3

Identification and verification of hub genes. a The PPI network of top 250 DEGs. b The construction of submodule by the plug-in of CytoHubba in Cytoscape. c The OncoPrint tab showed a visual summary of the different alterations of ten hub genes by the website of cBioPortal. d The protein expression of hub genes in GEPIA. *P < 0.05 compared with normal endometrial tissues. e The expression heatmap of ten hub genes in human cancers.

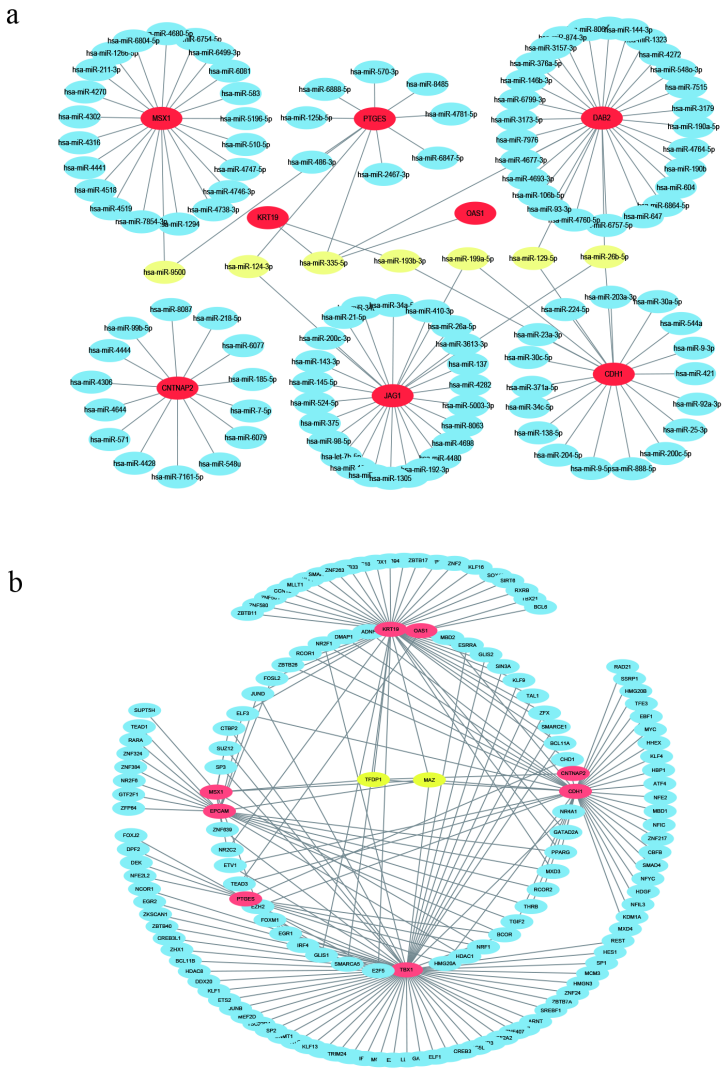

Figure 4

Hub gene-relevant MicroRNAs and Transcriptional Factors network analysis. a Hub gene-relevant miRNA network. Red nodes stand for hub genes, blue nodes stand for relevant miRNA and yellow nodes stands mainly relevant miRNA. b Hub gene-transcription factors (TFs) regulatory network. Red nodes represent hub genes, blue nodes represent TFs and yellow nodes represent major TFs.

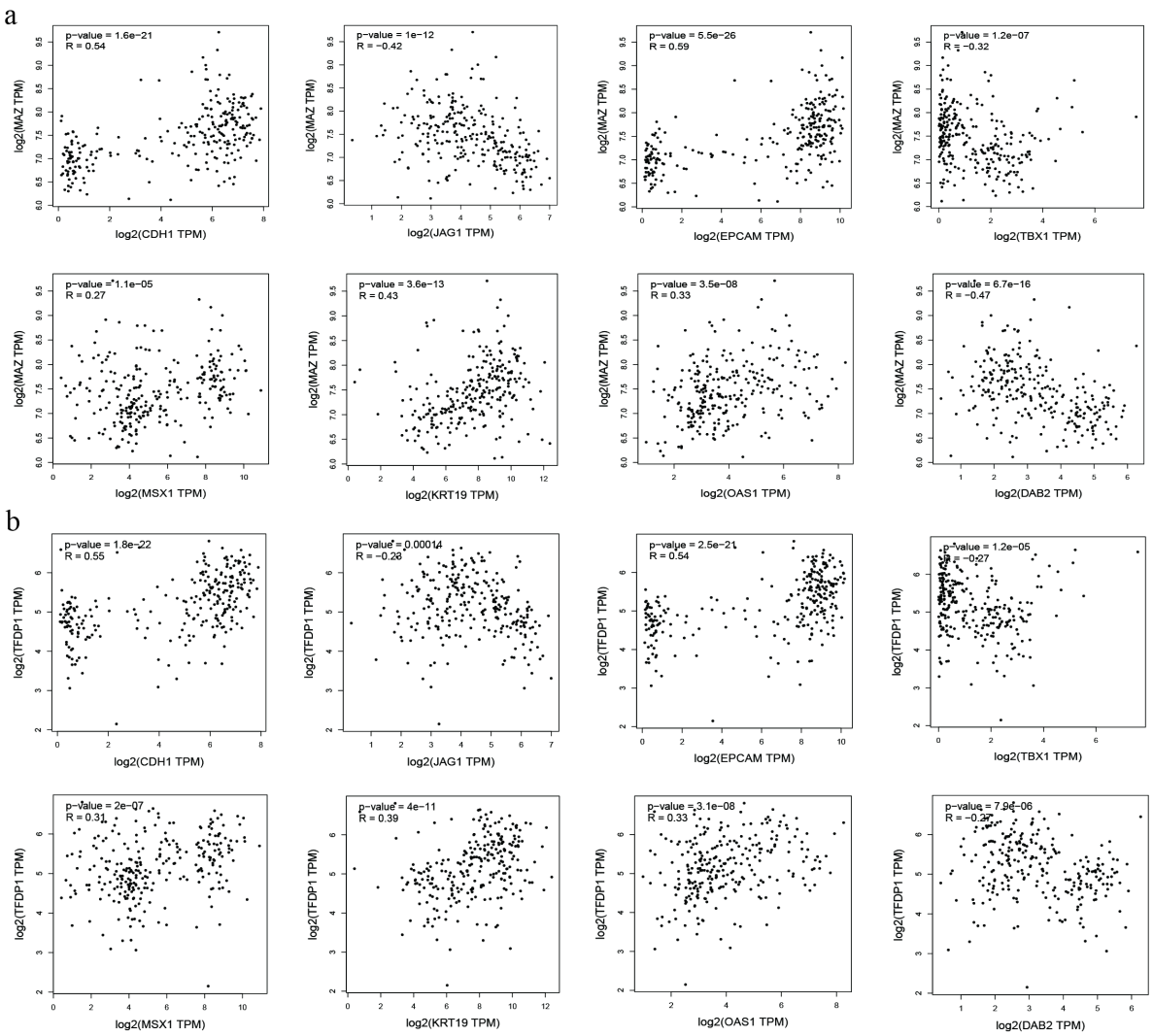

Figure 5

Correlation analyses of hub genes and core TFs. a Correlation analysis between hub genes and MAZ. b Correlation analysis between hub genes and TFDP1.

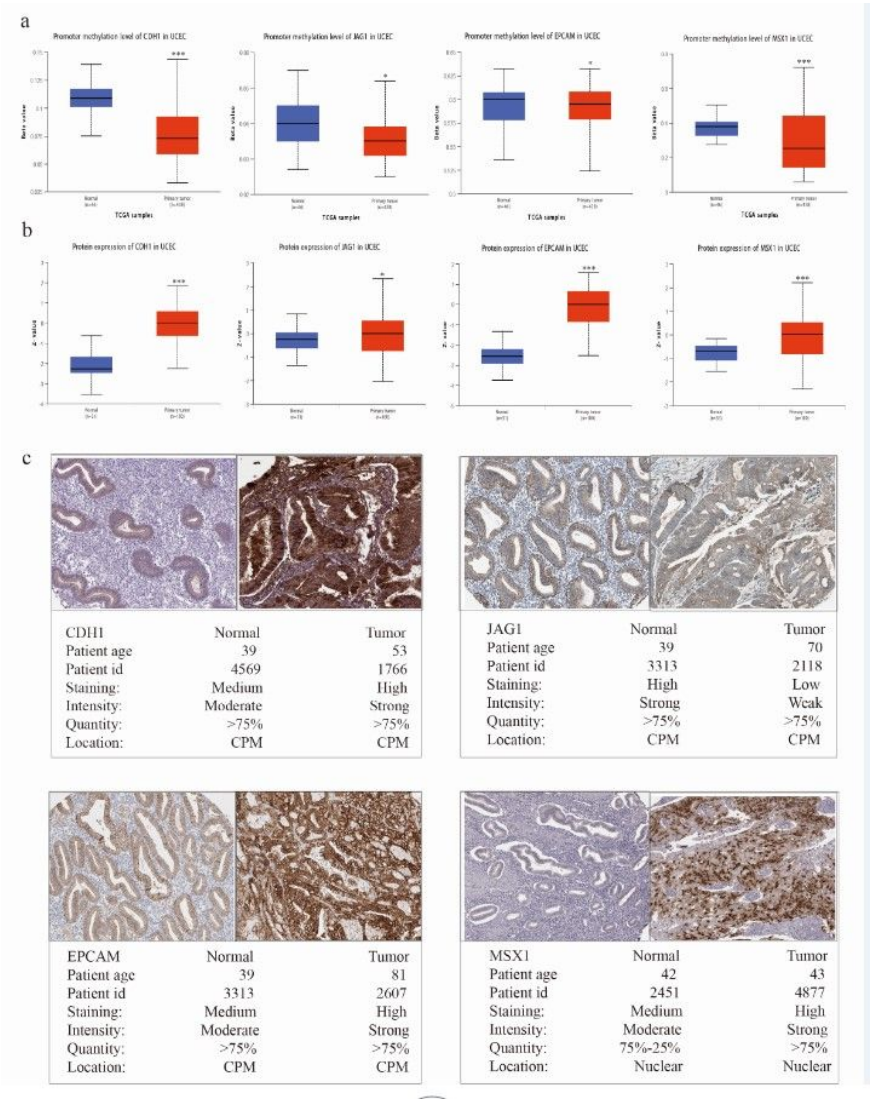

Figure 6

Gene methylation status and expression status of hub genes were further validated. a The methylation status of four hub genes. b The protein expressions of four genes were validated in Ualcan website. c The demonstration of Immunohistochemistry (IHC) staining of four genes by HPA website. *p < 0.05 compared with normal endometrial tissues; **P <0.01 compared with normal tissues; ***P<0.001 compared with normal tissues.

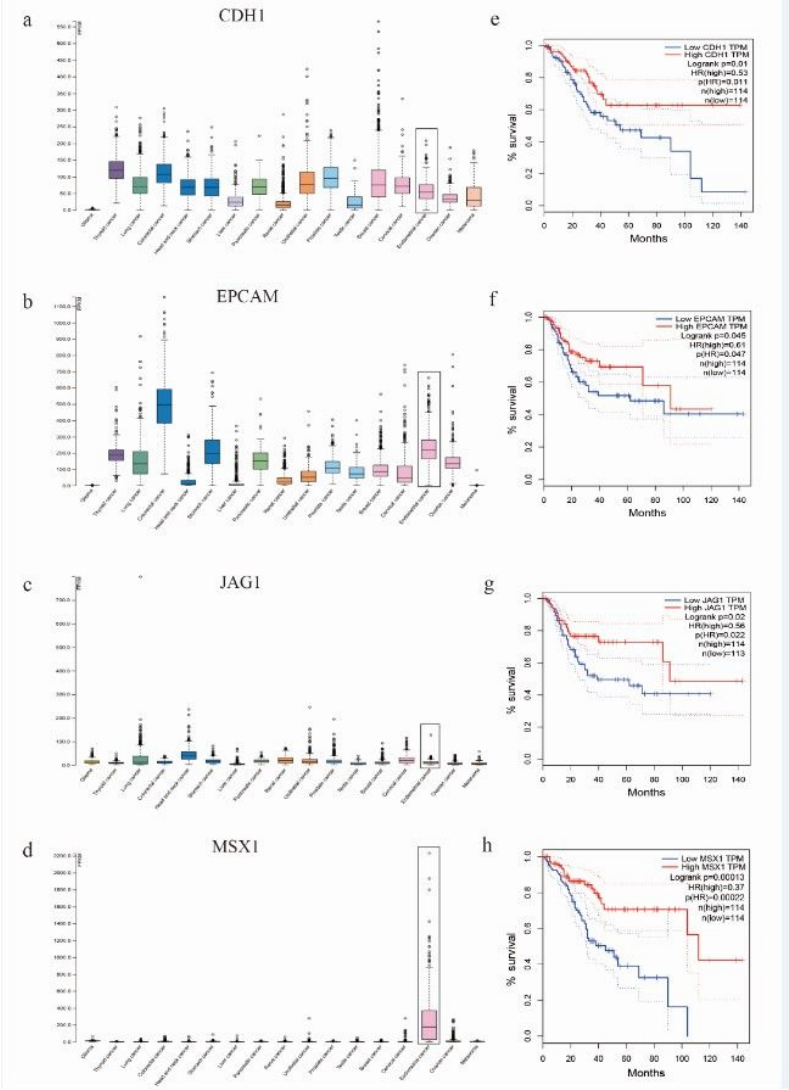

Figure 7

Tissue specificity analysis and overall survival evaluation of hub genes. a-d Basic expression of hub genes in different human cancer organs based on TCGA. e-h Survival analysis of hub genes in endometrial carcinoma by GEPIA.

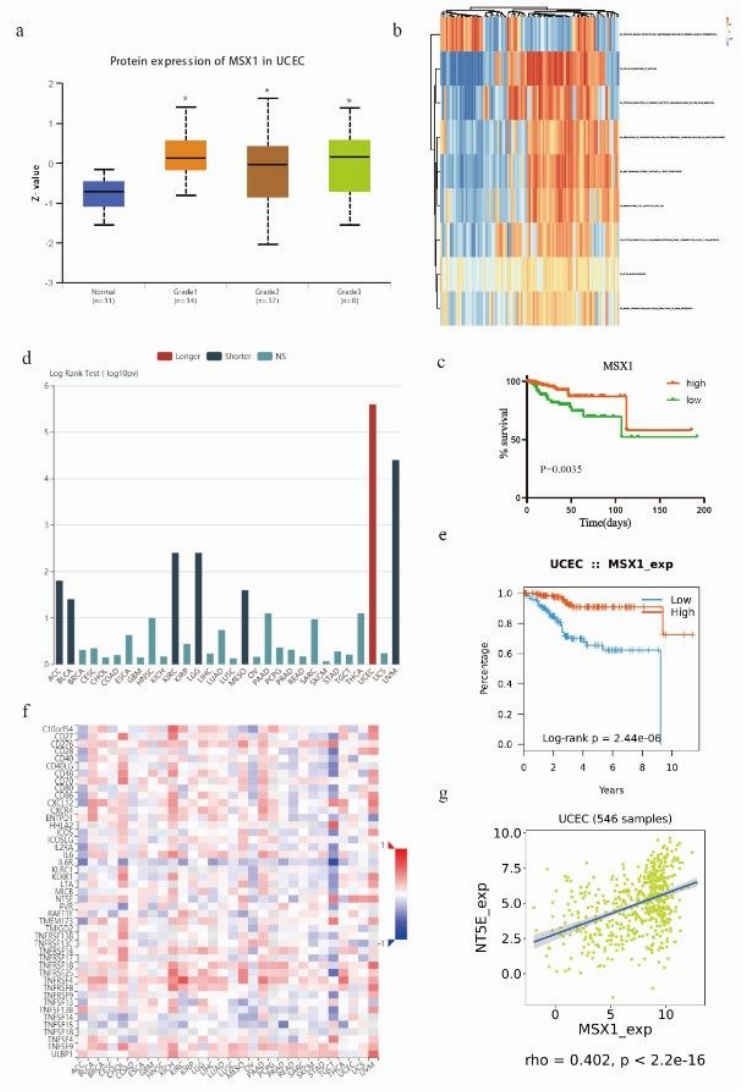

Figure 8

Validation of hub gene in the TCGA dataset. a Transcriptional expression of MSX1 was significantly correlated with EC stages. b GSVA-derived clustering heatmap of differentially expressed pathways for MSX1. c The survival value of MSX1 based on TCGA data. d The prognostic value of MSX1 in human tumors based on TCGA cohort. e High expression of MSX1 is related to the better prognosis of patients with EC. f Spearman correlations between expression of MSX1 and immunostimulators across human cancers. g Correlation between expression of MSX1 and immunostimulator NT5E.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfile2.xlsx
- SupplementaryFigure1.tif
- SupplementaryFigure2.png
- Supplementaryfile3.xlsx
- Supplementaryfile2.xlsx
- Supplementaryfile4.xlsx
- Supplementaryfilelegends.docx
- Supplementaryfile1.xlsx
- Supplementaryfile4.xlsx
- SupplementaryFigure1.tif

- Supplementaryfile1.xlsx
- SupplementaryFigure2.png
- Supplementaryfile3.xlsx
- Supplementaryfilelegends.docx