

PRIMARY RESEARCH

Open Access



Identification of molecular markers associated with the progression and prognosis of endometrial cancer: a bioinformatic study

JinHui Liu[†], Mingming Feng[†], SiYue Li[†], Sipei Nie, Hui Wang, Shan Wu, Jiangnan Qiu, Jie Zhang and WenJun Cheng^{*}

Abstract

Background: Endometrial cancer (EC) is one kind of women cancers. Bioinformatic technology could screen out relative genes which made targeted therapy becoming conventionalized.

Methods: GSE17025 were downloaded from GEO. The genomic data and clinical data were obtained from TCGA. R software and bioconductor packages were used to identify the DEGs. Clusterprofiler was used for functional analysis. STRING was used to assess PPI information and plug-in MCODE to screen hub modules in Cytoscape. The selected genes were copied with functional analysis. CMap could find EC-related drugs that might have potential effect. Univariate and multivariate Cox proportional hazards regression analyses were performed to predict the risk of each patient. Kaplan–Meier curve analysis could compare the survival time. ROC curve analysis was performed to predict value of the genes. Mutation and survival analysis in TCGA database and UALCAN validation were completed. Immunohistochemistry staining from Human Protein Atlas database. GSEA, ROC curve analysis, Oncomine and qRT-PCR were also performed.

Results: Functional analysis showed that the upregulated DEGs were strikingly enriched in chemokine activity, and the down-regulated DEGs in glycosaminoglycan binding. PPI network suggested that NCAPG was the most relevant protein. CMap identified 10 small molecules as possible drugs to treat EC. Cox analysis showed that BCHE, MAL and ASPM were correlated with EC prognosis. TCGA dataset analysis showed significantly mutated BHCE positively related to EC prognosis. MAL and ASPM were further validated in UALCAN. All the results demonstrated that the two genes might promote EC progression. The profile of ASPM was confirmed by the results from immunohistochemistry. ROC curve demonstrated that the mRNA levels of two genes exhibited difference between normal and tumor tissues, indicating their diagnostic efficiency. qRT-PCR results supported the above results. Oncomine results showed that DNA copy number variation of MAL was significantly higher in different EC subtypes than in healthy tissues. GSEA suggested that the two genes played crucial roles in cell cycle.

Conclusion: BCHE, MAL and ASPM are tumor-related genes and can be used as potential biomarkers in EC treatment.

Keywords: Endometrial cancer, Differentially expressed genes (DEGs), TCGA, GEO, PPI, CMap, Prognosis

*Correspondence: wenjunchengdoc@163.com

[†]JinHui Liu, Mingming Feng and SiYue Li contributed equally to this work
Department of Gynecology, The First Affiliated Hospital of Nanjing
Medical University, 300 Guangzhou Road, Nanjing 210029, Jiangsu, China



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Endometrial cancer (EC) is the fourth commonest malignancy in females [1]. In 2015, the American Cancer Society (ACS) predicted that the number of new EC cases was 54,870, and 10,170 of them died. This means that in the past 20 years, the mortality of EC has almost doubled. The average age of patients at diagnosis is 63. Among them, 90% are over 50 years old, and only about 20% can get diagnosed before menopause [2]. At present, no useful tool is available to screen EC, hence accurate and early diagnosis is critical for the treatment of EC patients [3–6]. The routine treatment options of EC include surgery, radiotherapy and chemotherapy. Recent years have seen the emergence but not the wide use of EC-related targeted therapy [7]. In recent years, based on the rapid development of high-throughput sequencing technology and increasingly complete public databases, The Cancer Genome Atlas (TCGA) database and Gene Expression Omnibus (GEO) database has collected a large number of clinical, pathological, and biological data from patients with malignancies [8, 9]. Through comprehensive analysis of data, we can more accurately predict the development trend and dig deeper into the mechanism of tumors, providing a reliable research direction for treatment programs [10, 11]. For example, Lin et al. found that a network of RBM8A expression regulated hepatocellular carcinoma [12]. This research screened DEGs using GEO and TCGA data firstly, secondly constructed PPI and co-expression network. Based on cox prognosis analysis, TCGA data, website validation and qRT-PCR, we determined the hub gene and pathways which might be related to EC [13].

Materials and methods

Data collection and analysis

The raw data on GSE17025 and TCGA were integrated. We obtained gene expression profiles from GEO database (<http://www.ncbi.nlm.nih.gov/geo/>). GSE17025 dataset [14] covered 91 tumor and 12 normal tissue samples. Affymetrix Human Genome U133 Plus 2.0 Array [15] processed raw data. Robust multi-array average (RMA) approach was performed for background correction and normalization [16]. The original GEO data was then converted into expression measures using affy R package [17]. TCGA dataset (<https://cancergenome.nih.gov/>) was downloaded, covering 35 normal tissue samples and 552 tumor tissues. TCGA (<https://cancergenome.nih.gov/>) provided the clinical, genomic and mutation data of UCEC and Illumina Hi-Seq RNA-Seq platform provided these RNA sequencing data. These data were downloaded on Feb 01, 2018. The research design was illustrated with a flow chart (Fig. 1).

Access to DEGs

Package “limma” [18] screened out the DEGs between EC and normal uterus samples. The cut-off criteria was set as The adjusted $p < 0.05$ and $|\log_2\text{fold change (FC)}| \geq 2$. For TCGA data, edgeR package was used for DEGs screening [19]. The cut-off criteria was set as $\text{FDR} < 0.01$ and $|\log_2\text{fold change (FC)}| \geq 2$. Adjusted p -value and FDR were for ruling out false-positives. Online Wayne diagram was for identifying the DEGs simultaneously found in GSE17025 and TCGA. Heatmap was completed by “heatmap” package in R 3.4.4 [20].

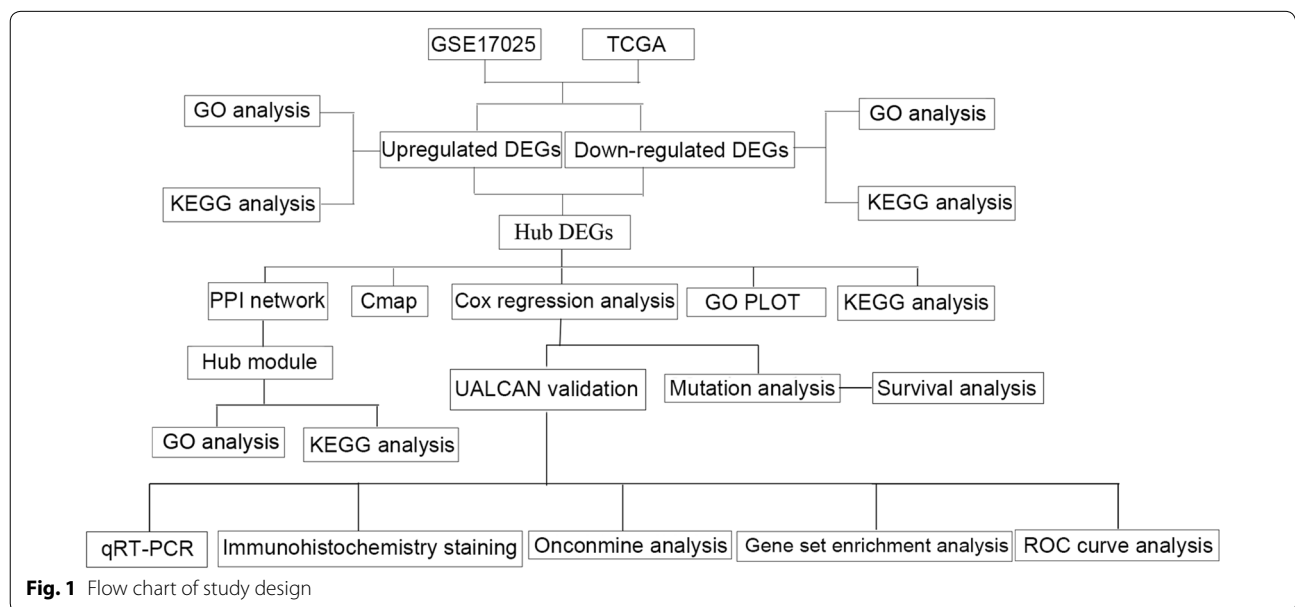


Fig. 1 Flow chart of study design

Functional enrichment analysis

DAVID database (<https://david.ncifcrf.gov/>) is a foundation for high-throughput gene functional analysis. The function and enriched pathways of the proteins encoded by the candidate genes were analyzed and these genes were annotated using DAVID database. GO annotations of the screened DEGs were performed using the DAVID online tool and clusterprofiler [21]. KEGG pathway analysis on DEGs was performed using clusterprofiler. p -value < 0.05 was set to be significant.

Construction and analysis of PPI network complex

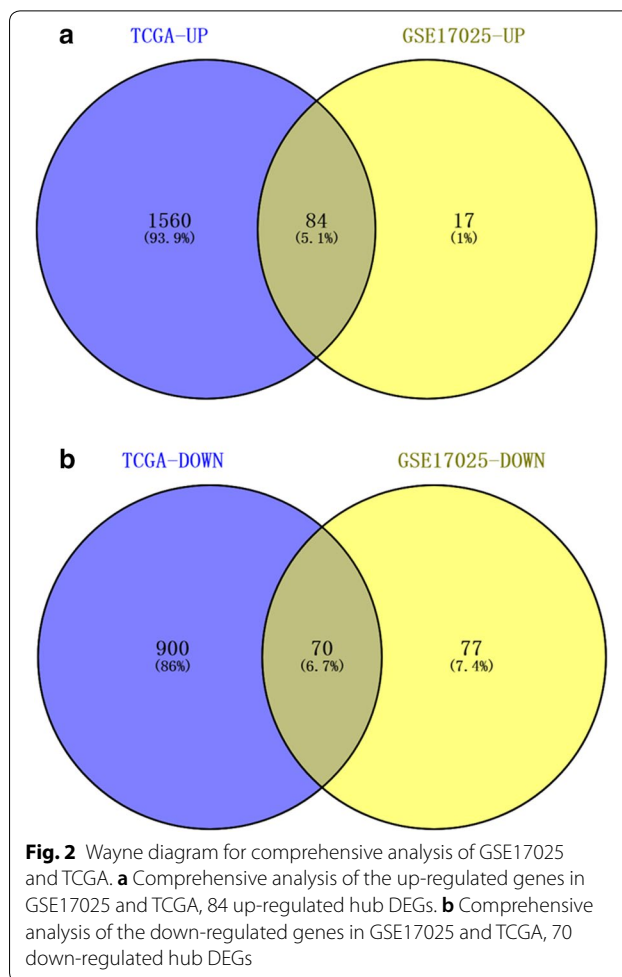
STRING (The Retrieval of Interacting Genes Database) (<http://www.string-db.org/>) provided PPI (protein–protein interaction) information [22]. We used STRING database to explore the interactions between DEGs and visualize the results using Cytoscape software. Cytoscape MCODE plug-in provided access to select hub modules of PPI network [23]. The criteria default parameters as follows: k -core = 2, degree cut-off = 2, max. depth = 100 and node score cut-off = 0.2. For genes in the hub module, we use clusterprofiler again for functional enrichment analysis.

Identification of potential small molecules

The EC gene signature was queried in CMap. CMap is a computer simulation method for predicting potential drugs that may affect the biological state encoded in gene expression signatures [23]. The DEGs probesets were used to query the CMap database. Finally, the enrichment score indicative of similarity was calculated, ranging from -1 to 1 . A positive connectivity score indicated that a drug could induce the signatored biology in human cell lines. Conversely, a negative connectivity score indicated that a drug could reverse the signatored biology in human cell lines, suggestive of its Possible treatment value. After rank ordering all instances, the connectivity scores were filtered by p -value. Tomograph of these relative molecular drugs was researched in Pubchem database (<https://pubchem.ncbi.nlm.nih.gov/>).

Construction of a prognostic signature

Univariate Cox proportional hazards regression analyses could provide some Prognosis-related information. With the cutoff of $p < 0.05$, DEGs were seemed to be Prognosis-related. For the top 10 significant prognosis-related genes, construction of multivariate Cox proportional hazards regression model would be helpful. Cox proportional hazards regression with a $p < 0.05$ was set for risk score of developing EC in each patient. According to the mean risk score, patients will be divided into low- and high-risk groups. Kaplan–Meier curve analysis will compare the survival of low-risk and high-risk groups. $p < 0.05$ was the significant cutoff. Receiver operating



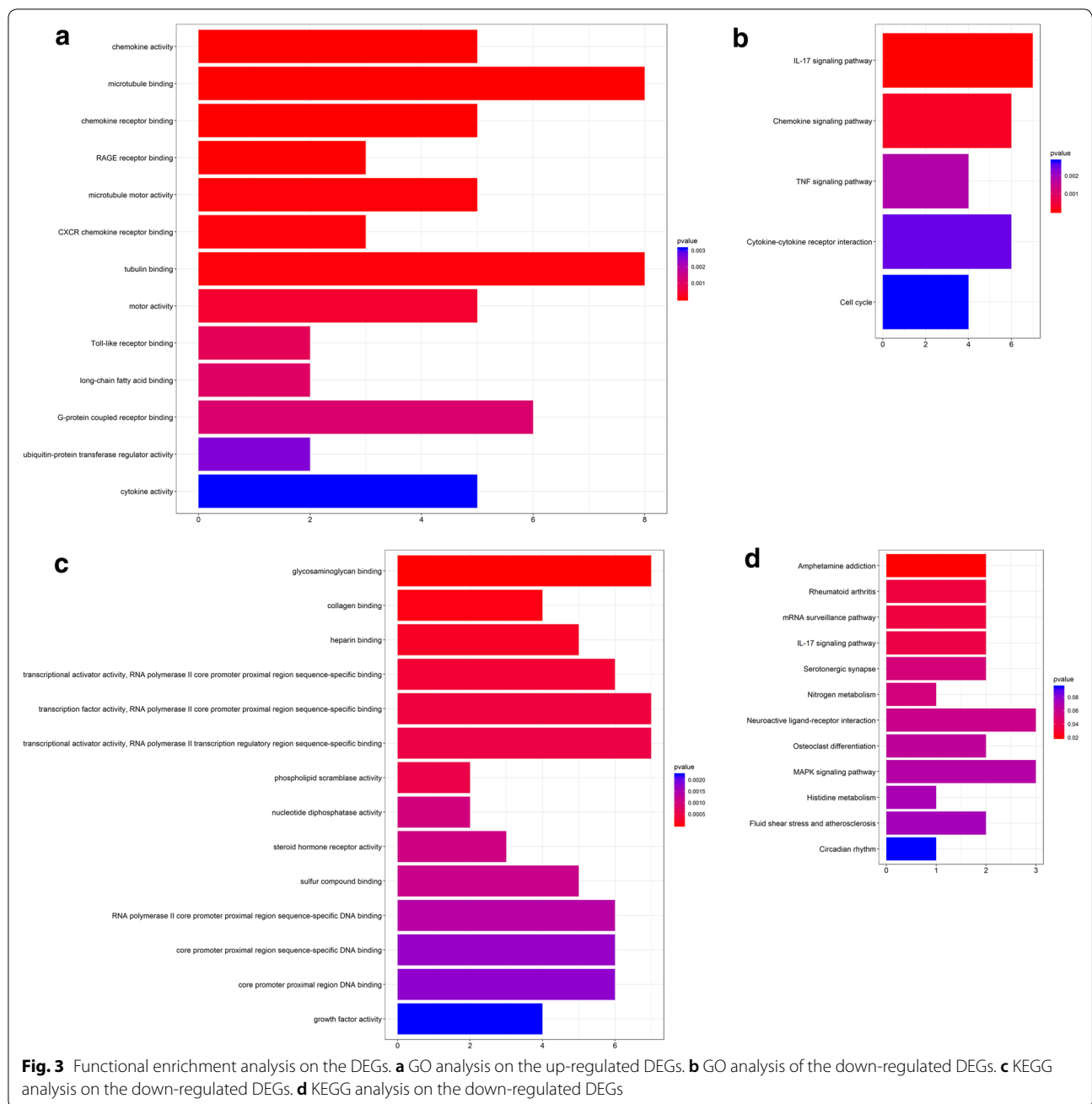
characteristic (ROC) curve analysis was also performed to estimate the five-year predictive value of the outcomes. The area under the ROC curve was calculated as a predictive value shown as sensitive and specificity.

Validation of hub genes

R package was for mutation analysis based on the TCGA dataset. The real hub gene was finally validated in UALCAN (<http://ualcan.path.uab.edu/analysis.html>) [24]. Oncomine 4.5 database (<http://www.oncomine.org>) was utilized to compare differential expression of common cancer types and their normal adjacent tissues. The HPA (Human Protein Atlas) (<http://www.proteinatlas.org/>) was used to validate real hub gene [25]. ROC curve analysis was performed in SPSS 23.0 to distinguish normal and cancer tissues.

Preparation for human EC samples

The study was approved by the Institutional Review Board of Nanjing Medical University. The tissue was removed from the EC patient showing informed consent and

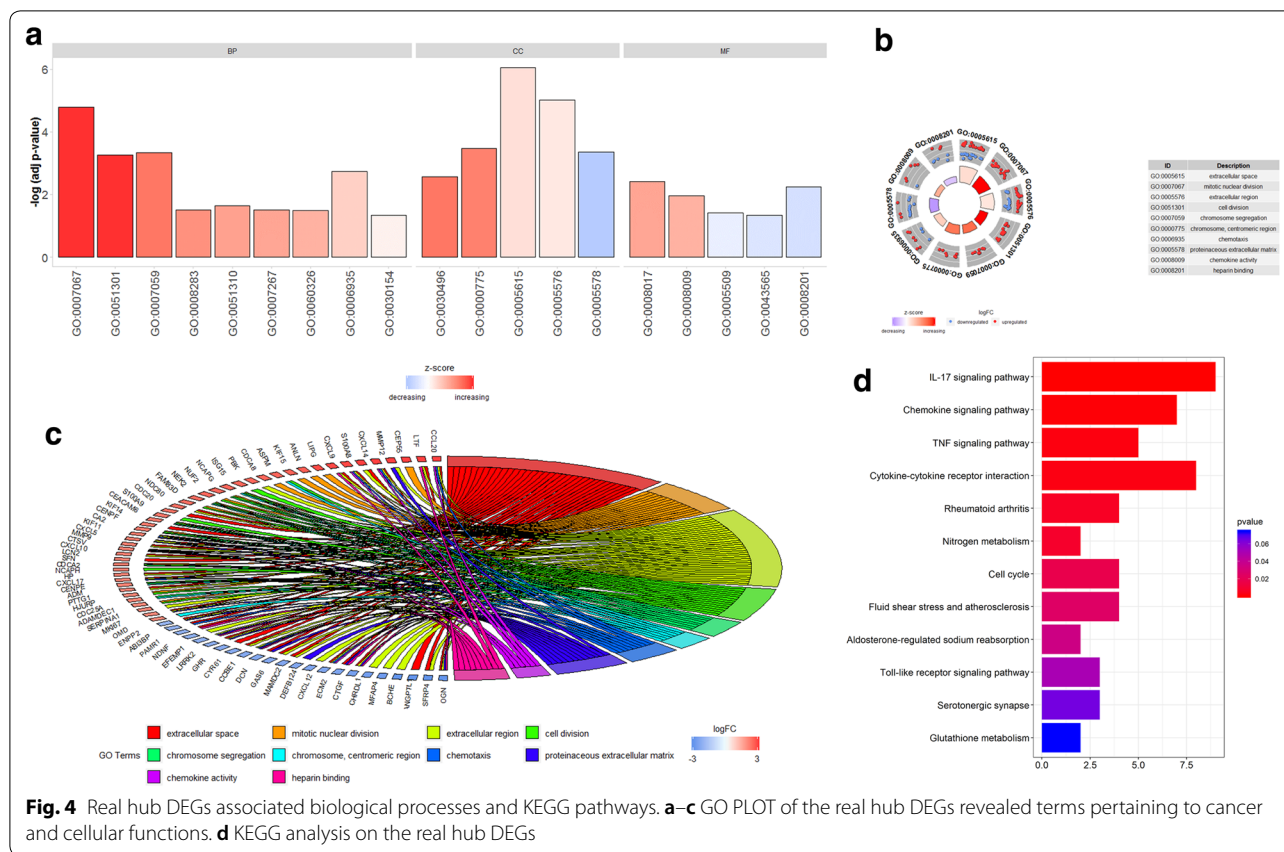


immediately stored in an environment of -80°C until use. From June 2018 to January 2019, 16 EC tissue samples and 16 normal uterus tissue samples were made in the Department of Gynecology and Obstetrics, the First Affiliated Hospital of Nanjing Medical University including.

Quantitative real-time RT-PCR (qRT-PCR) analysis

Total RNA was extracted by TRIzol reagent (Invitrogen); RNA 6000 Nano kit and Agilent Bioanalyzer 2100 were used to assess the integrity of the isolated RNA through

OD260/280 and OD260/230 ratios, PrimeScript[®] RT reagent kit was used to react RNA and synthesize single-stranded complementary DNA from RNA according to the manufacturer’s instructions. SYBR[®] Premix Ex Taq[™] Kit (TaKaRa DRR041) was utilized to perform real-time quantification. The cycle threshold (Ct) of each gene was recorded. The relative expression of MAL was calculated as follow: $2^{-\Delta\Delta\text{Ct}}$ ($\Delta\text{Ct} = \text{Ct}_{\text{target gene}} - \text{Ct}_{\text{internal control}}$). Forward Primer of MAL was “CGCTGCCCTCTTTTACCTCA”. Reverse Primer of MAL was “GAA



GCCGTCCTTGCATCGTGAT”. GAPDH was used as an endogenous control. Forward primer of GAPDH was “AGAAGGCTGGGGCTCATTG”. Reverse primer of GAPDH was “AGGGCCATCCACAGTCTTC”. Quantitative real-time RT-PCR was performed according to the manufacturer’s protocols.

Gene set enrichment analysis (GSEA)

According to the expression level of hub genes, EC samples from TCGA were divided into 2 different groups. In order to dig out the relative functions, GSEA (<http://software.broadinstitute.org/gsea/index.jsp>) was used to define the biological processes enriched in the gene rank derived from DEGs between the two groups [26]. Terms with FDR < 0.05 and enriched in all real hub genes were identified.

Statistical analysis

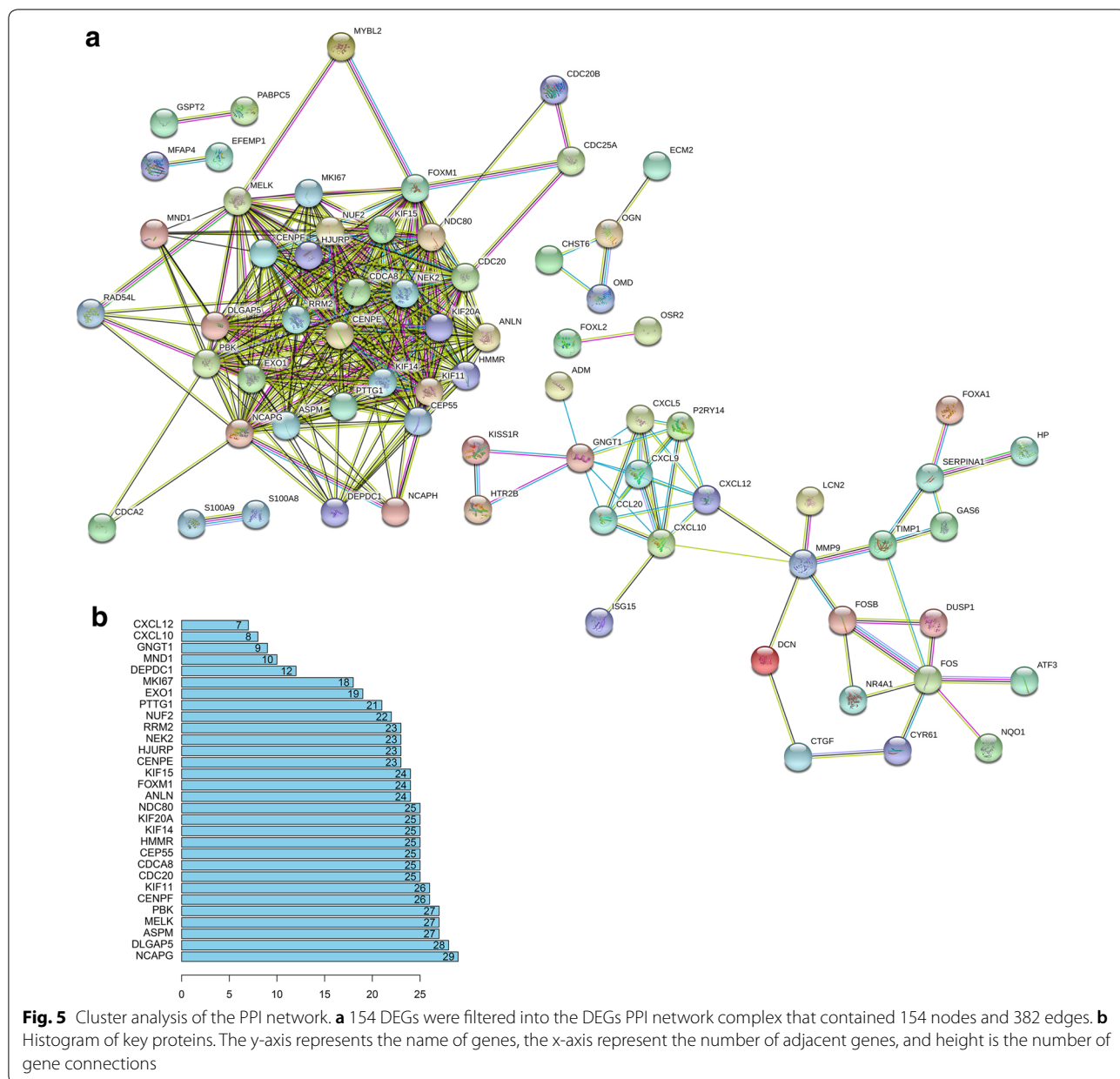
Two-tailed Student’s t-test was for calculating the difference between subgroups. All analyses were repeated three times. The represented data comes from three separate experiments. Statistical analysis was completed by SPSS 23.0 and R software 3.4.4. *p* < 0.05 was considered statistically significant.

Results

Identification of DEGs and the enriched processes

We identified the DEGs in GSE17025 using “limma” with adj. *p* < 0.05 and |log2fold change (FC)| ≥ 2. The top 200 genes in GSE17025 were displayed in the heatmap (Additional file 1: Figure S1). We identified the DEGs in TCGA using the edgeR package with FDR < 0.01 and |log2fold change (FC)| ≥ 2. All the genes in TCGA were displayed in the heatmap (Additional file 2: Figure S2). In GSE17025, we screened out 248 DEGs, including 101 up-regulated and 147 down-regulated. All of them were found in EC samples (Additional file 3: Figure S3). In TCGA, we screened out 2614 DEGs from EC samples, which contained 1644 up-regulated and 970 down-regulated. (Additional file 4: Figure S4).

Of the DEGs in GSE17025 and TCGA, we screened out 84 up-regulated hub genes (Fig. 2a) and 70 down-regulated hub genes (Fig. 2b). Clusterprofiler was for evaluating the enrichment of gene clusters in biological terms with a cutoff of *p* < 0.05. GO analysis demonstrated that the up-regulated hub genes were mostly enriched in chemokine activity, microtubule binding, chemokine receptor binding, RAGE receptor binding, microtubule motor activity, CXCR chemokine receptor binding and



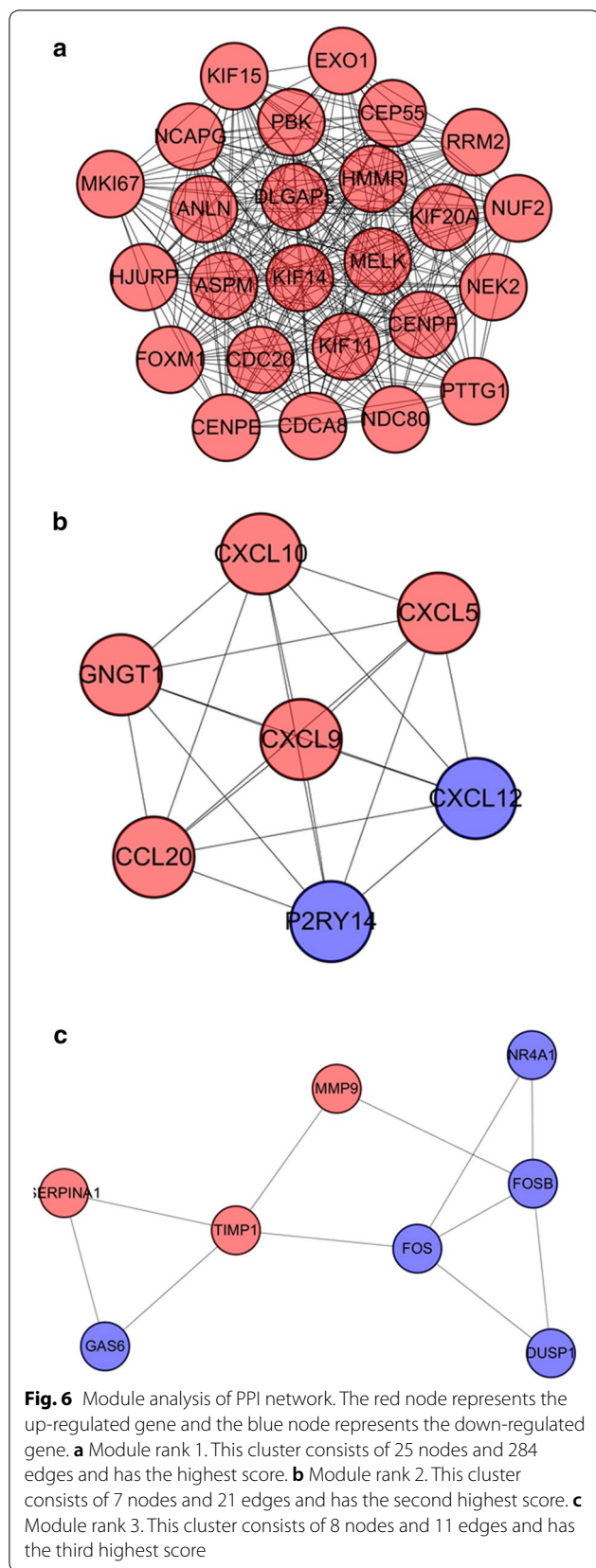
tubulin binding (Fig. 3a); the down-regulated hub genes were highly enriched in glycosaminoglycan binding and collagen binding (Fig. 3b). In KEGG analysis, the up-regulated hub genes were mostly enriched in IL-17 signaling pathway (Fig. 3c); the down-regulated hub genes were highly enriched in Amphetamine addiction (Fig. 3d). We further analyzed these DEGs with adjusted $p < 0.05$ and $|\log_{2}FC| \geq 2$, finding out 154 real DEGs.

Identification of hub DEGs in EC and functional analysis

To clarify the functions of these DEGs, we first explored the associated biological processes and

KEGG pathways in TCGA and GEO datasets. The most enriched GO terms pertaining to biological process (BP), cellular, component (CC) and molecular function (MF) are shown in Fig. 4a. The most enriched GO term in BP was “mitotic nuclear division” ($p < 0.05$), that in CC was “extracellular space” ($p < 0.05$), and that in MF was “microtubule binding” ($p < 0.05$) (Fig. 4b). We further obtained 10 significantly enriched GO terms with a p -value < 0.05 (Fig. 4c).

KEGG analysis showed that the DEGs were strikingly enriched in IL-17 signaling pathway, Chemokine signaling pathway and TNF signaling pathway (Fig. 4d).



PPI network and analysis on clusters

STRING mapped 154 DEGs into PPI network containing 154 nodes and 382 edges (Fig. 5a). A total of 30 prominent proteins were identified, with estrogen NCAPG being the most important protein contacting 29 nodes (Fig. 5b).

Then MCODE was used to find clusters in the network. Four clusters were calculated according to $k\text{-core}=2$. Among them, cluster 1 contained 25 nodes and 284 edges, with the highest score (Fig. 6a), cluster 2 contained 7 nodes and 21 edges (Fig. 6b), cluster 3 contained 8 nodes and 11 edges (Fig. 6c), cluster 4 contained 3 nodes and 3 edges. These results suggested that the 154 DEGs had effects on EC.

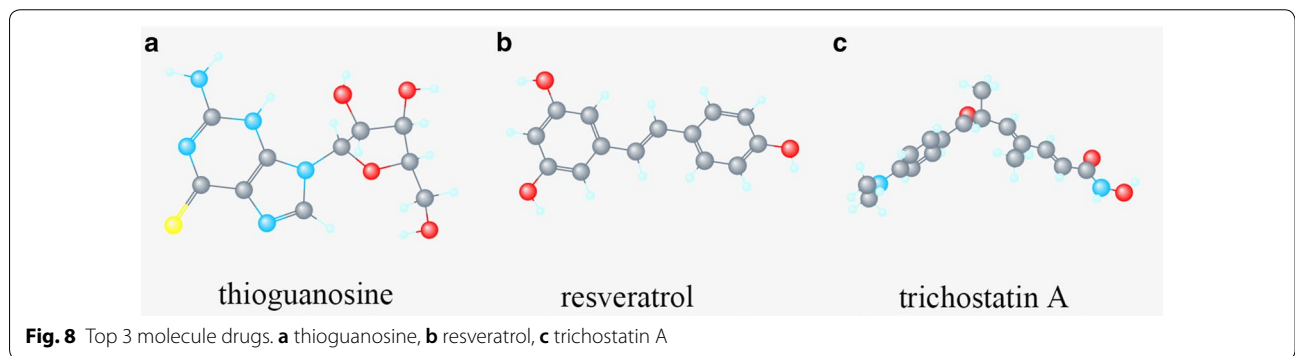
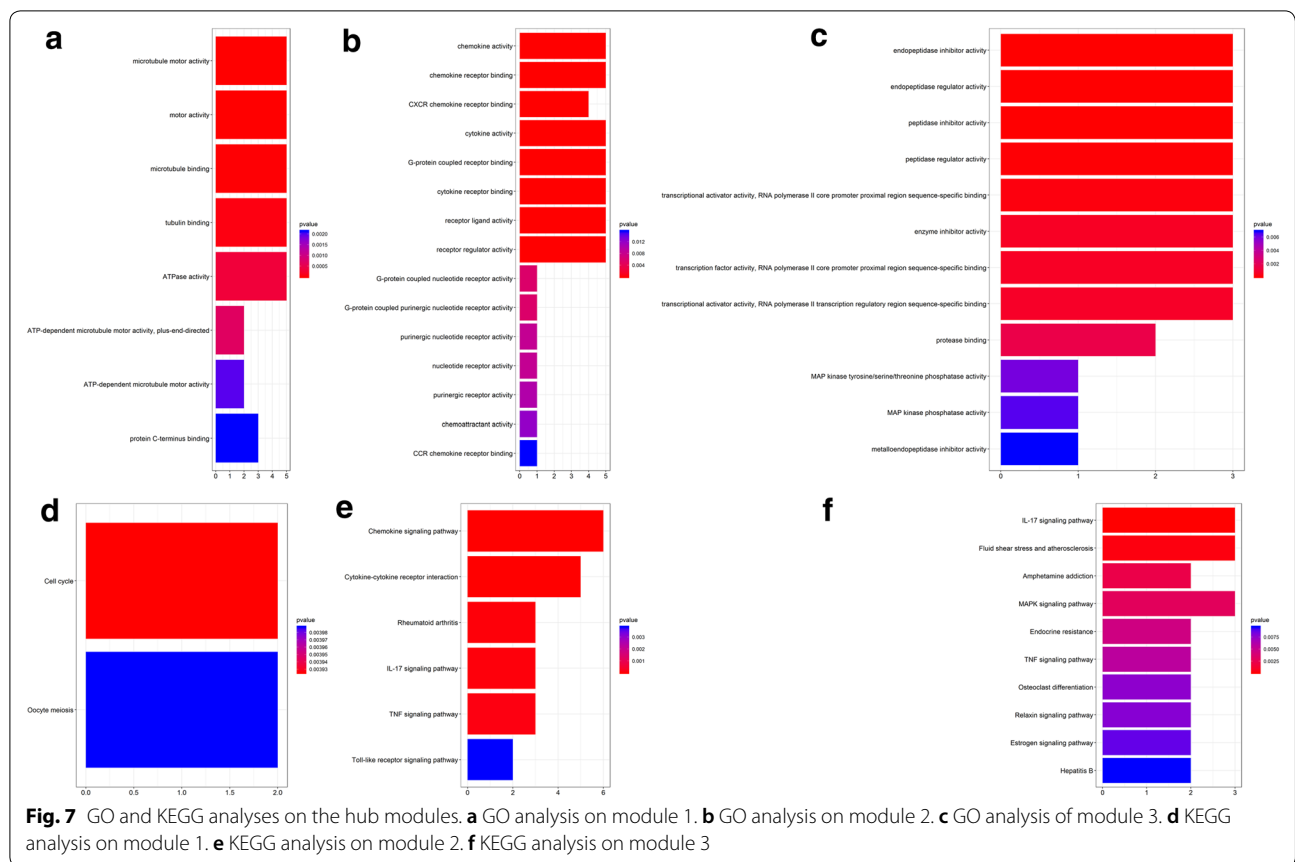
We performed the functional analysis for the top 3 clusters. In GO analysis, the DEGs of cluster 1 were mostly enriched in microtubule motor activity, motor activity and microtubule binding (Fig. 7a); the DEGs of cluster 2 in chemokine activity, chemokine receptor binding, CXCR chemokine receptor binding, cytokine activity, G-protein coupled receptor binding, cytokine receptor binding and receptor ligand activity (Fig. 7b); the DEGs of cluster 3 in endopeptidase inhibitor activity, endopeptidase regulator activity, peptidase inhibitor activity, peptidase regulator activity and transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding (Fig. 7c). KEGG analysis showed that the DEGs of cluster 1 were mostly enriched in cell cycle (Fig. 7d); the DEGs of cluster 2 in chemokine signaling pathway and cytokine–cytokine receptor interaction (Fig. 7e); the DEGs of cluster 3 in IL-17 signaling pathway (Fig. 7f).

Potential small molecule drugs

CMap compared EC samples with healthy controls to screen out small molecule drugs. Strong negative correlation was found between EC and thioguanosine, resveratrol, trichostatin A, 0175029-0000, trifluoperazine and LY-294002; strong positive correlation was found between EC and viomycin, adiphenine, clorsulon and heptaminol (Additional file 5: Figure S5). These drugs might have therapeutic effects on EC. The tomographies of the top 3 associated molecule drugs were investigated in Pubchem database (Fig. 8a–c).

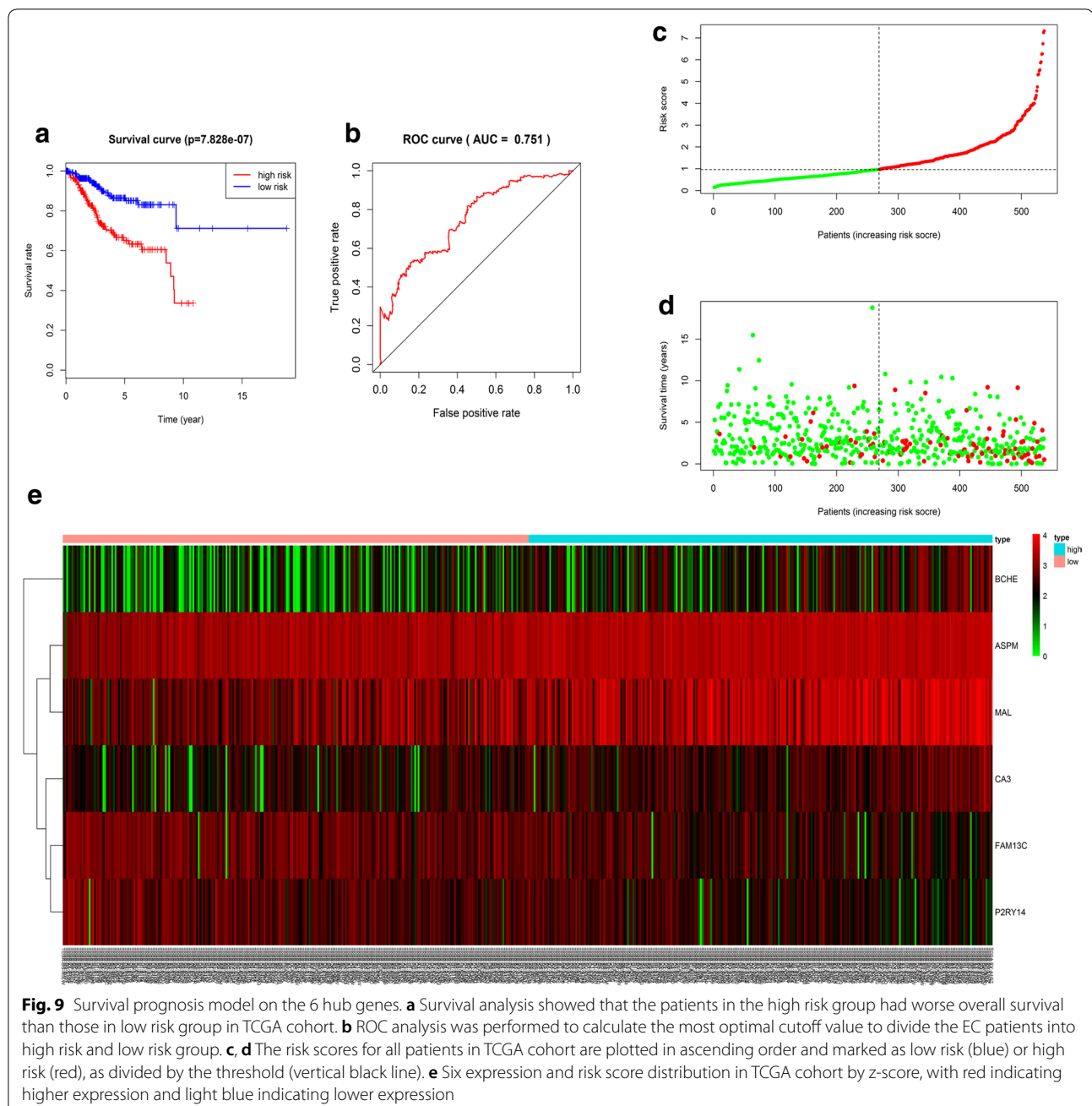
Identification of prognostic signature

Of the 154 DEGs, the univariate Cox proportional hazards regression analysis screened out the top 10 EC-relative genes, including MAL, BCHE, P2RY14, ASPM, CA3, FAM13C, FAM83D, CXorf57, SFN and CPED1 (Additional file 6: Figure S6); multivariate Cox proportional hazards regression analysis was further performed



on the 10 genes, which screened BCHE, FAM13C, CA3, P2RY14, ASPM and MAL (Additional file 7: Figure S7). The risk score for predicting overall survival was calculated as follows: Risk score = 0.126 * BCHE - 0.121 * FAM13C + 0.136 * CA3 - 0.139 * P2RY14 + 0.231 * ASPM + 0.0892 * MAL. According to the risk score, patients were divided into low- and high-risk groups. Survival analysis

showed that low-risk patients had longer overall survival than high-risk patients in TCGA cohort (Fig. 9a). The AUC of five-year survival ROC curve analysis was 0.751 (Fig. 9b). The distribution of risk score, survival status, and the expression of six genes of each patient were also analyzed (Fig. 9c-e). The expression level of the six genes in low- and high-risk groups was shown in Additional



file 8: Figure S8. Meanwhile, we analyzed the relationship between the different clinical parameters and the risk score based on six genes. The univariate and multivariate Cox proportional hazards regression showed that only tumor status together with the risk score based on six genes was independent prognostic indicator of EC (Table 1). The heatmap showed the expression levels of the six genes in high- and low-risk groups based on the

TCGA dataset. We observed significant between-group differences in tumor status, grade, histological type, age and stage ($p < 0.001$) (Additional file 9: Figure S9).

Hub gene validation

Based on TCGA dataset and using R language, we performed mutation analysis on BCHE, ASPM and MAL which exhibited significant prognostic value ($p < 0.05$). We

Table 1 Univariate analysis and multivariate analysis of the correlation of six-gene expression with OS among endometrial cancer patients

Parameter	Univariate analysis			Multivariate analysis		
	HR	95% CI	p value	HR	95% CI	p value
Age (≤ 60 vs > 60)	1.788	1.118–2.859	<i>0.015</i>	1.566	0.956–2.566	0.075
Stage (stage I & stage II vs stage III & stage IV)	4.070	2.70–6.205	<i>0.000</i>	1.604	0.993–2.591	0.053
Histological_type (endometrioid vs mix & serous)	2.997	1.972–4.553	<i>0.000</i>	0.840	0.489–1.442	0.527
Grade (G1 & G2 vs G3 & G4)	3.395	1.975–5.835	<i>0.000</i>	1.544	0.834–2.857	0.167
Tumor_status (with tumor vs tumor free)	11.042	7.05–17.300	<i>0.000</i>	7.795	4.682–12.978	<i>0.000</i>
RiskScore	1.543	1.369–1.739	<i>0.000</i>	1.195	1.022–1.398	<i>0.026</i>

Italic values indicate $p < 0.05$

HR hazard ratio, CI confidence interval

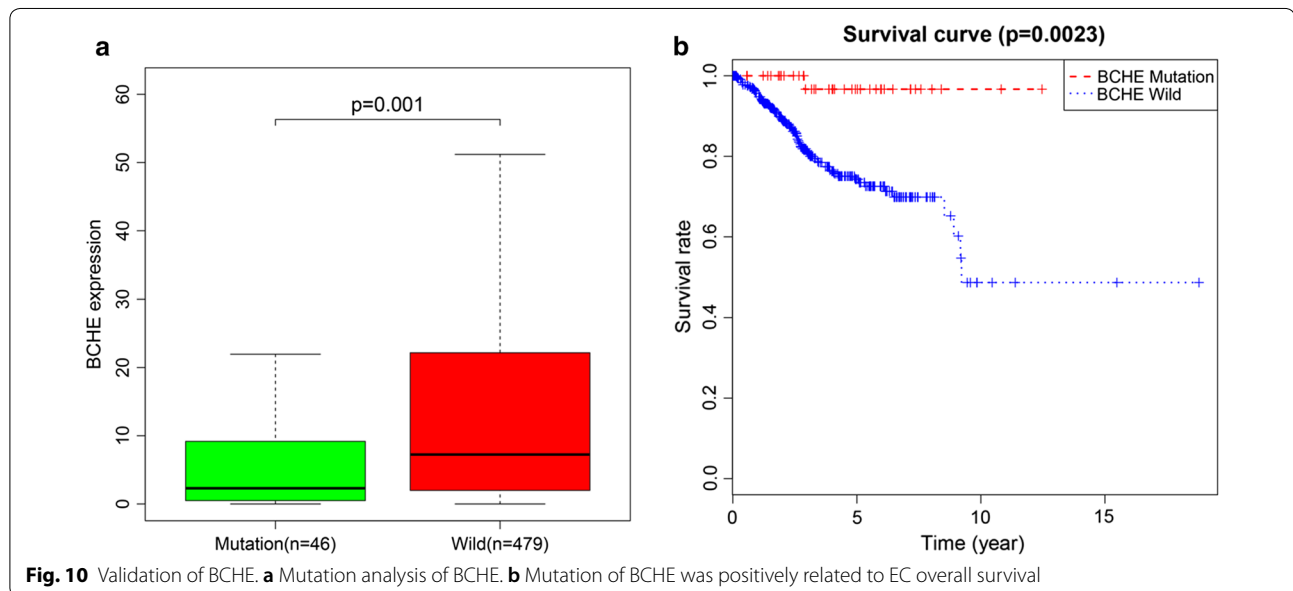
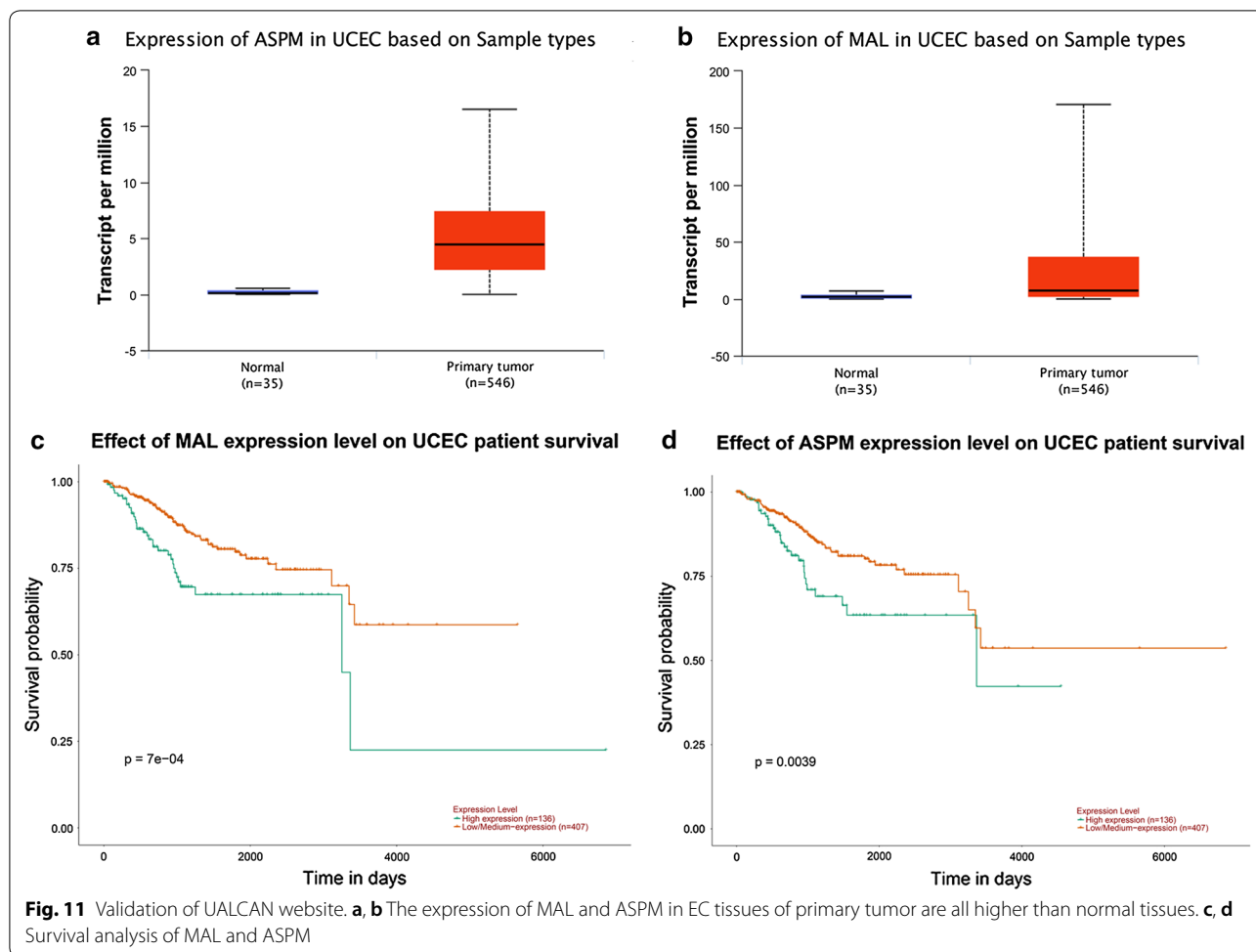


Fig. 10 Validation of BCHE. **a** Mutation analysis of BCHE. **b** Mutation of BCHE was positively related to EC overall survival

found that BCHE showed significant mutation (Fig. 10a). We further found that patients with BCHE mutation had a better prognosis (Fig. 10b), suggesting that BCHE mutation may be a protective factor for EC patients.

Using UALCAN, we found that MAL and ASPM expressed higher in tumor than in normal tissues (Fig. 11a, b), both negatively related to the overall survival of the EC patients (Fig. 11c, d). In addition, both had higher expression levels in EC tissues of different subtypes, such as serous carcinoma, endometrioid adenocarcinoma and mixed serous and endometrioid adenocarcinoma (Fig. 12a, b). Their expression levels also increased at different stages of EC (Fig. 12c, d). Finally, ROC curve analysis was for evaluating the capacity of

MAL and ASPM, so as to distinguish EC from normal tissues (Fig. 13). MAL was missing in the immunohistochemistry database. Immunohistochemistry staining showed the higher expression of ASPM in the tumor sample compared with the normal sample (Fig. 14). Data in the Oncomine 4.5 database revealed that DNA copy number variation (CNV) of MAL was significantly higher in different subtypes of EC tissues than in normal tissues ($p < 0.01$). Although the fold change of DNA CNV was within 2, MAL ranked within the top 5% (Fig. 15a–c). We further validate the expression of MAL in clinical tissues using qRT-PCR. Interestingly, the relative expression level of MAL was significantly elevated in tumor tissue than in normal tissue (Fig. 15d).



Gene set enrichment analysis (GSEA)

To identify the potential function of MAL and ASPM in EC, GSEA was conducted to search the enriched KEGG pathways. For ASPM, “cell cycle”, “DNA replication”, “oocyte meiosis”, “p53 signaling pathway”, “pancreatic cancer”, “progesterone mediated oocyte maturation”, “small cell lung cancer”, “ubiquitin mediated proteolysis” were enriched in eight gene sets (n=552) (Fig. 16). For MAL, “B cell receptor signaling pathway”, “bladder cancer”, “cell cycle”, “chronic myeloid leukemia”, “glycine serine and threonine metabolism”, “leukocyte transendothelial migration”, “pancreatic cancer”, “small cell lung cancer” were enriched in eight gene sets (n=552) (Fig. 17) (FDR < 0.05).

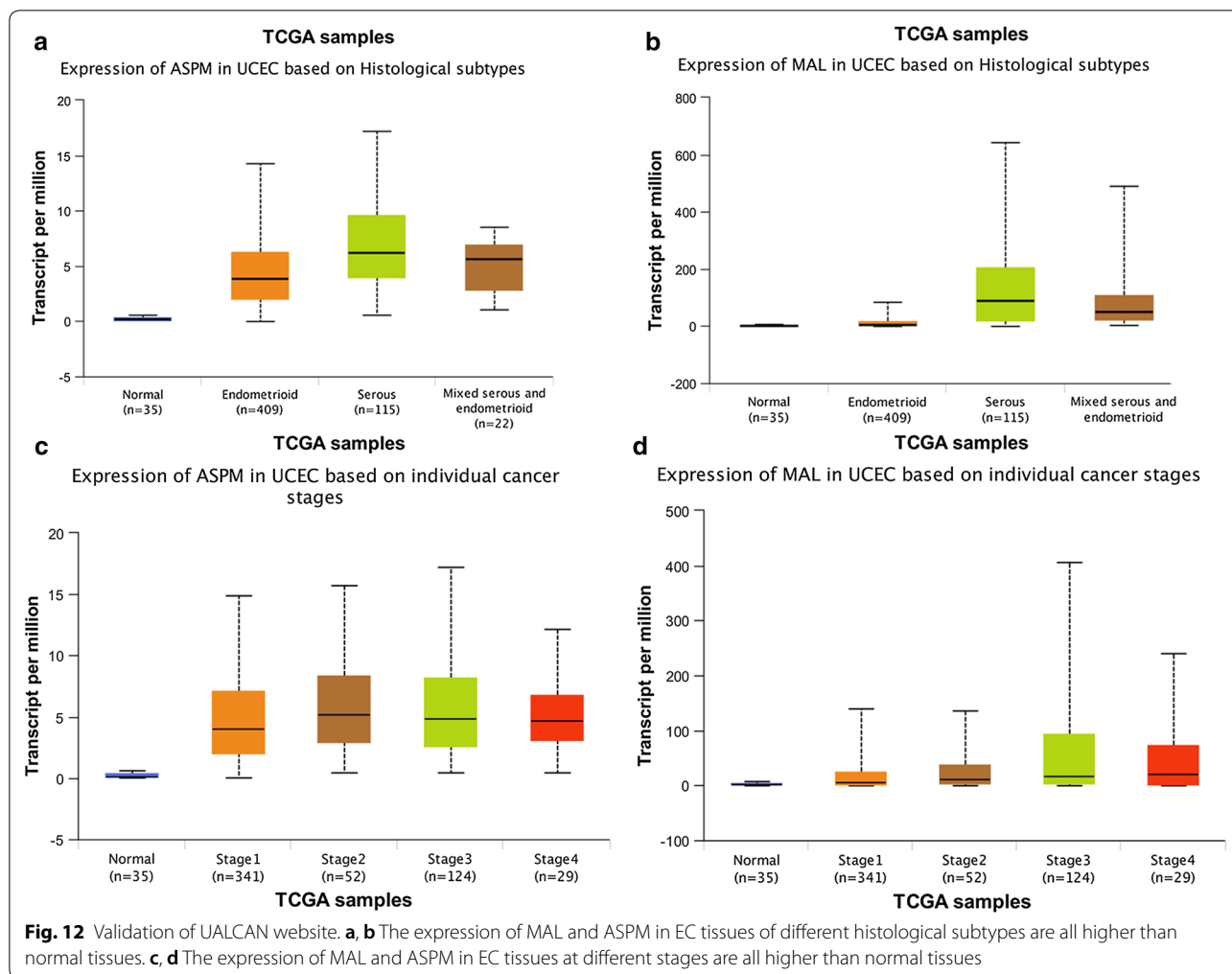
Discussion

In this study, we screened out EC-related DEGs based on GSE17025 and TCGA datasets. The up-regulated DEGs were strikingly enriched in chemokine activity and IL-17 signaling pathway, and the down-regulated DEGs in

glycosaminoglycan binding and Amphetamine addiction. We further screened out 154 hub DEGs, most of which were enriched in mitotic nuclear division, extracellular space, microtubule binding and IL-17 signaling pathway.

Liu et al. have found the chemokine activity in hepatocellular carcinoma metastasis [27]. IL-17 signaling pathway acts in the progression of lung cancer and liver cancer [28, 29]. Microtubule binding is involved in the development of colorectal cancer [30]. All these findings point out the possible correlation between some DEGs and EC development, which may provide a new direction for EC research.

We also found NCAPG was a functional protein in EC. NCAPG has shown its regulatory property in digestive tract tumors [31–33]. Some small EC-countering molecules have been identified. Among them, thioguanosine, resveratrol and trichostatin A have shown tight association with EC development. Thioguanosine can regulate the basal activity of leukemia cells [34] in. Resveratrol was proved to affect ovarian, breast and digestive tract

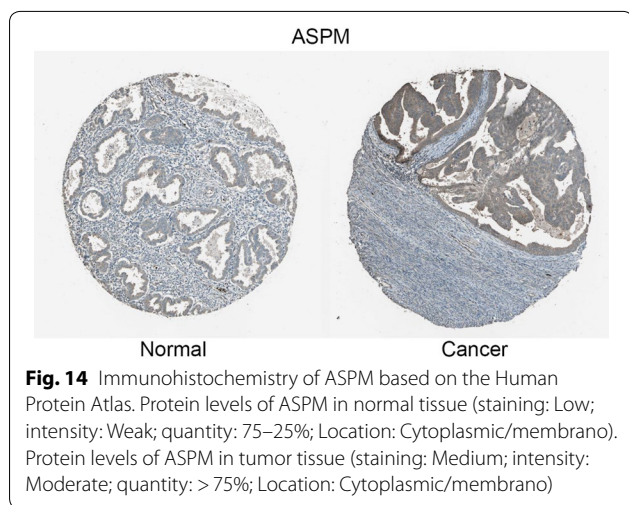
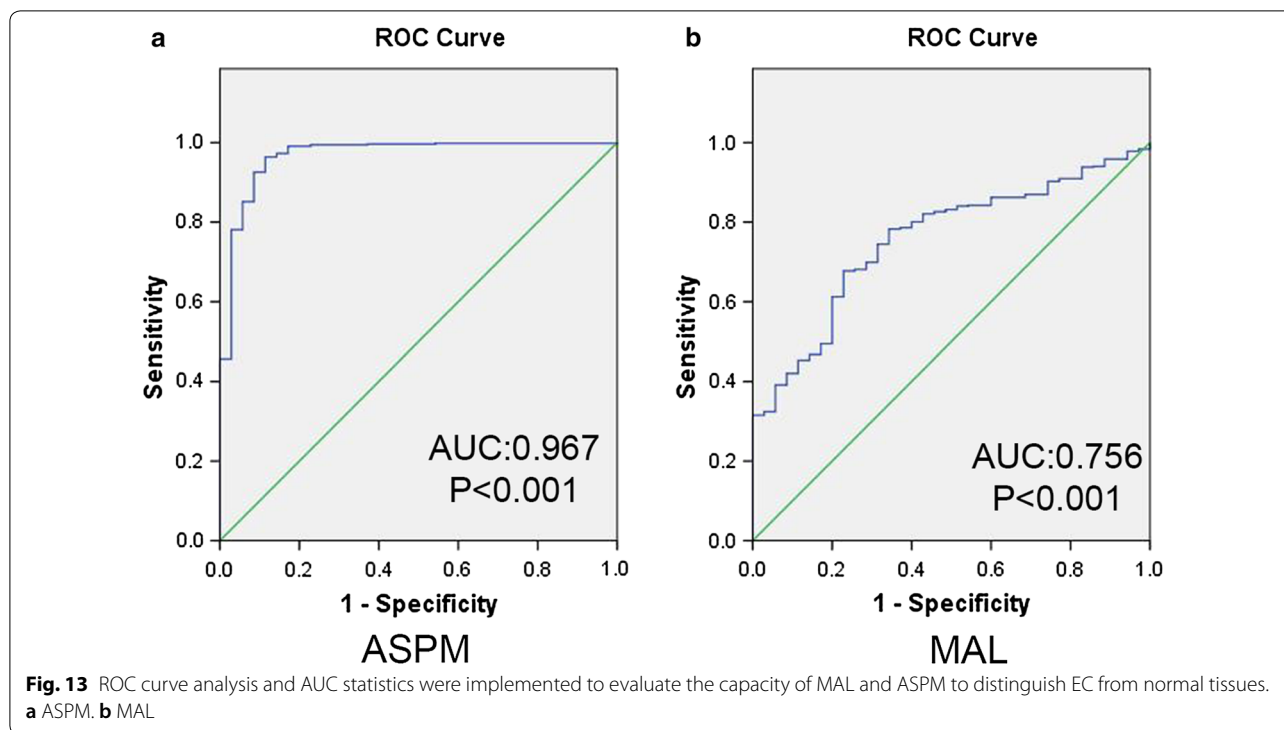


tumors [35–38]. Trichostatin A, as a histone deacetylase (HDAC) inhibitor, exhibits anticancer effects when used in combination with radiotherapy or chemotherapy [39, 40].

We found that ASPM, BCHE and MAL were highly EC-prognosis-related. BCHE mutation showed positive correlation with EC prognosis. Using UALCAN and TCGA datasets, we found the higher expression of MAL and ASPM in EC tissue than in normal tissue, and their expression levels were negatively related to the overall survival of the EC patients. In addition, MAL and ASPM had higher expression levels in EC tissues of different subtypes, such as serous carcinoma, endometrioid adenocarcinoma and mixed serous and endometrioid adenocarcinoma. Their expression levels also increased with EC stage. Finally, ROC curve analysis was conducted for evaluating the capacity of MAL and ASPM in distinguishing EC from normal tissues. Immunohistochemistry

staining demonstrated the higher expression of ASPM in EC samples, compared with that in the normal samples. Data in the Oncomine 4.5 database revealed that DNA copy number variation (CNV) of MAL was significantly higher in different subtypes of EC tissues than in the healthy tissues. We further validated the expression of MAL in clinical tissues using qRT-PCR. Interestingly, the relative expression level of MAL was significantly elevated in tumor tissues compared to the normal tissues. GSEA enrichment analysis showed that MAL and ASPM were mostly associated with cell cycle.

Butyrylcholinesterase (BChE) is a plasma enzyme that hydrolyzes ghrelin and bioactive esters. Its modulation in prostate cancer development has been proven [41]. Bernardi et al. found that amplification and deletion of BCHE was related to cholinesterase genes in sporadic breast cancer [42].

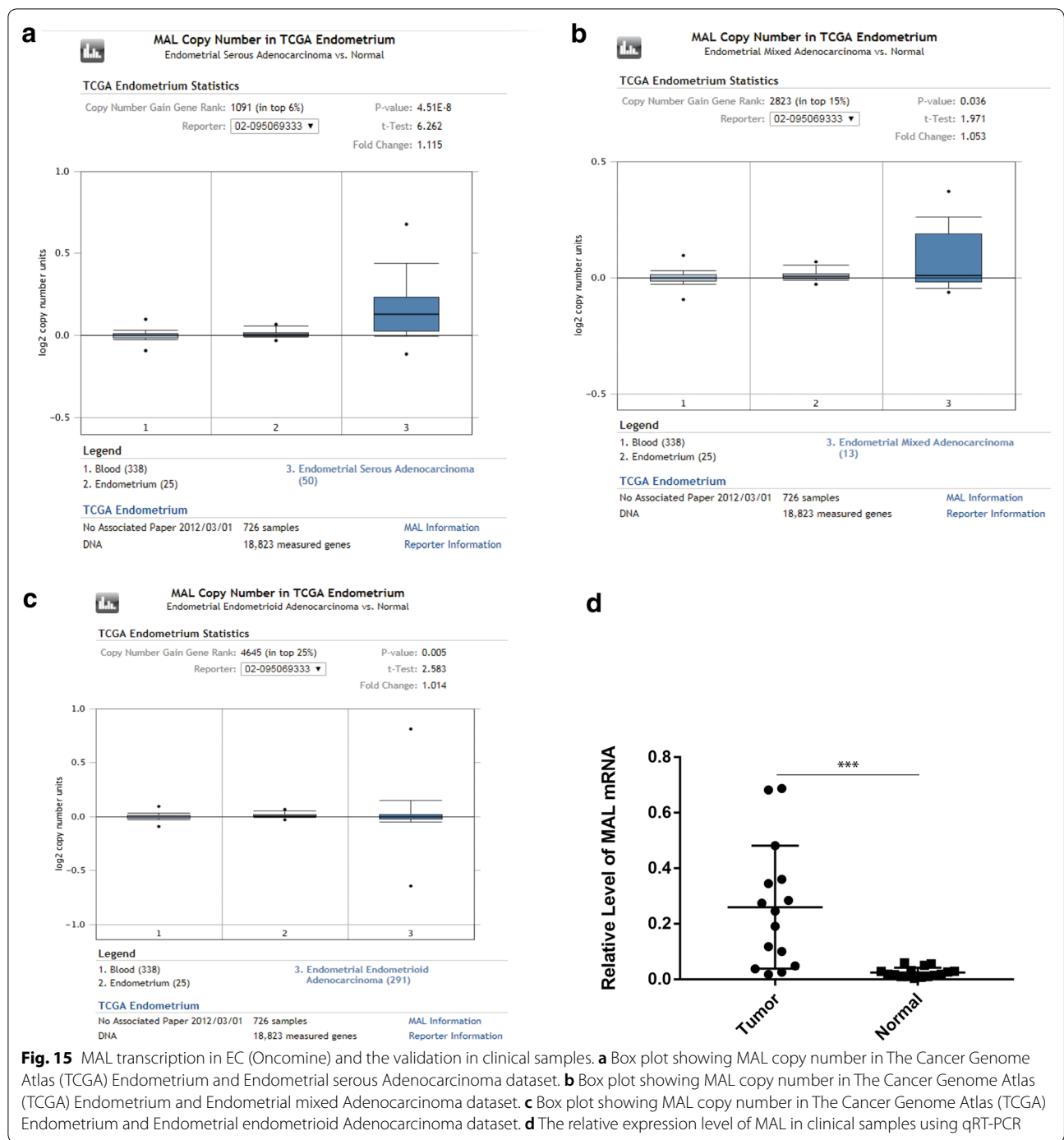


A study has confirmed that MAL serves as a bridge between TLR2/TLR4- and MyD88-mediated signaling to orchestrate downstream inflammatory responses and regulate intestinal homeostasis and colitis-associated colorectal cancer in mice [43]. Choi et al. found that MAL was significantly down-regulated and methylated in gastric cancer tissues [44]. Van Baars et al. found that MAL

methylation in cervical scrapes could indicate the status of underlying lesion. Zanotti et al. found that in high-grade serous ovarian carcinoma, MAL overexpression predicted chemoresistance and poor prognosis [45, 46]. In the present study, MAL as the hub gene also showed its aberrant expression in EC development, a finding that can guide the future exploration into EC mechanism.

In prostate and hepatocellular cancers, ASPM promoted the progression and exacerbated the prognosis, which means ASPM might be a novel marker for cancer progression [47, 48]. Researches showed that ASPM worsened the prognosis of ovarian cancer in many ways [49, 50]. ASPM can also increase the aggressiveness of pancreatic tumor [51]. In the present study, we verified the prognostic value of ASPM in EC, which broadens the landscape of ASPM research.

There are some highlights of our study. First of all, we obtained hub genes by taking the intersection between DEGs of TCGA and GEO, finding that BCHE, MAL and ASPM are tumor-related genes and can be used as potential biomarkers in EC treatment. Second, the prognostic model in our study can effectively predict EC patients' outcomes, which provide a new method to evaluate patients' prognosis. Third, Thioguanosine, resveratrol and trichostatin A can be used as antagonists against EC.



However, there are some limitations in this study. For example, our research was actually an analysis based on previous data; therefore, additional experimental studies in vivo and vitro are needed. In the future, based on the results of this study, we will design PCR, Western blotting and immunohistochemistry tests to explore the

molecular mechanisms. Second, the clinical sample size of PCR was not large enough. Finally, the therapeutic effects of candidate drugs targeting DEGs should be verified. In all, more well-designed studies should be carried out to support our findings.

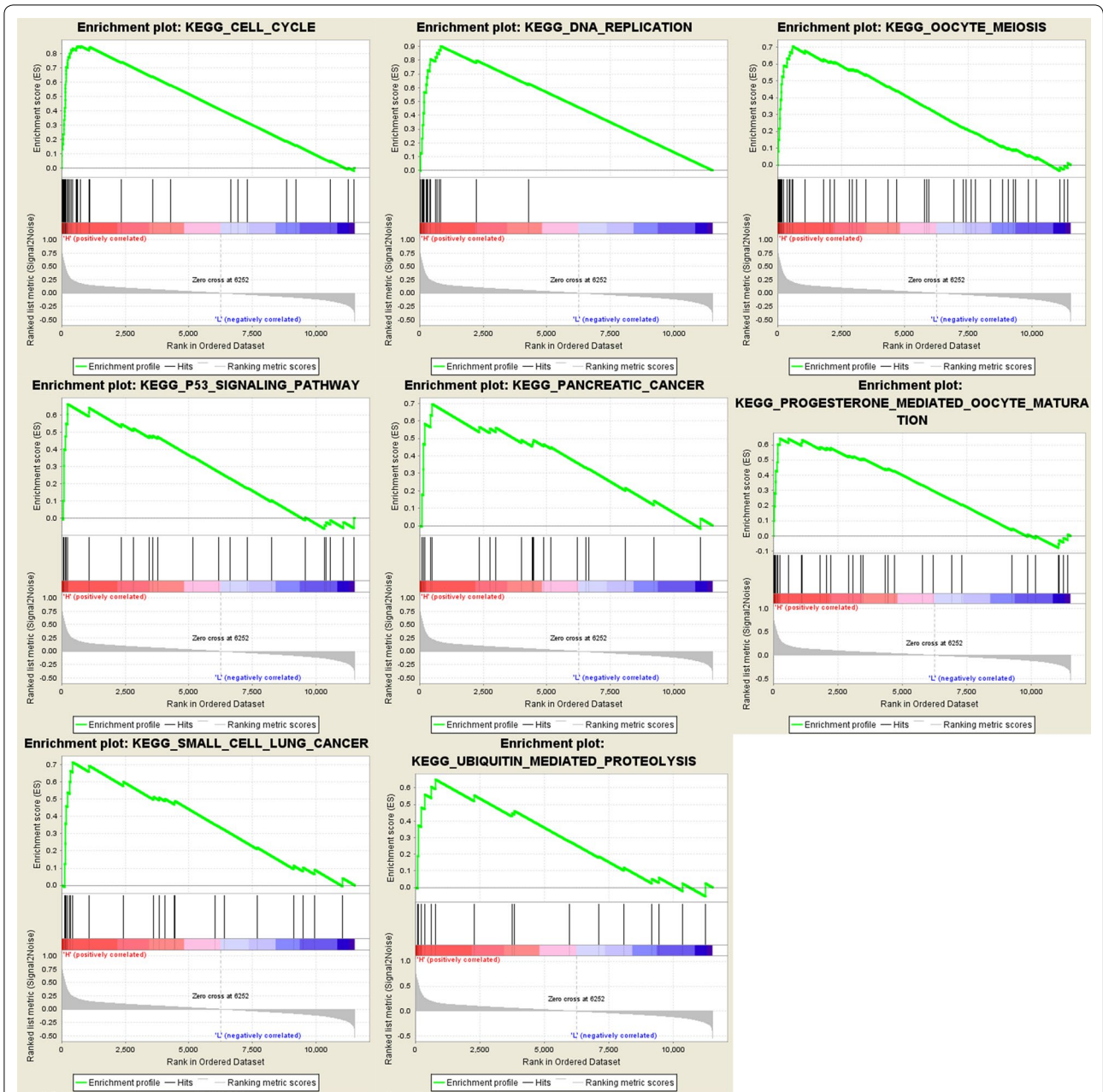
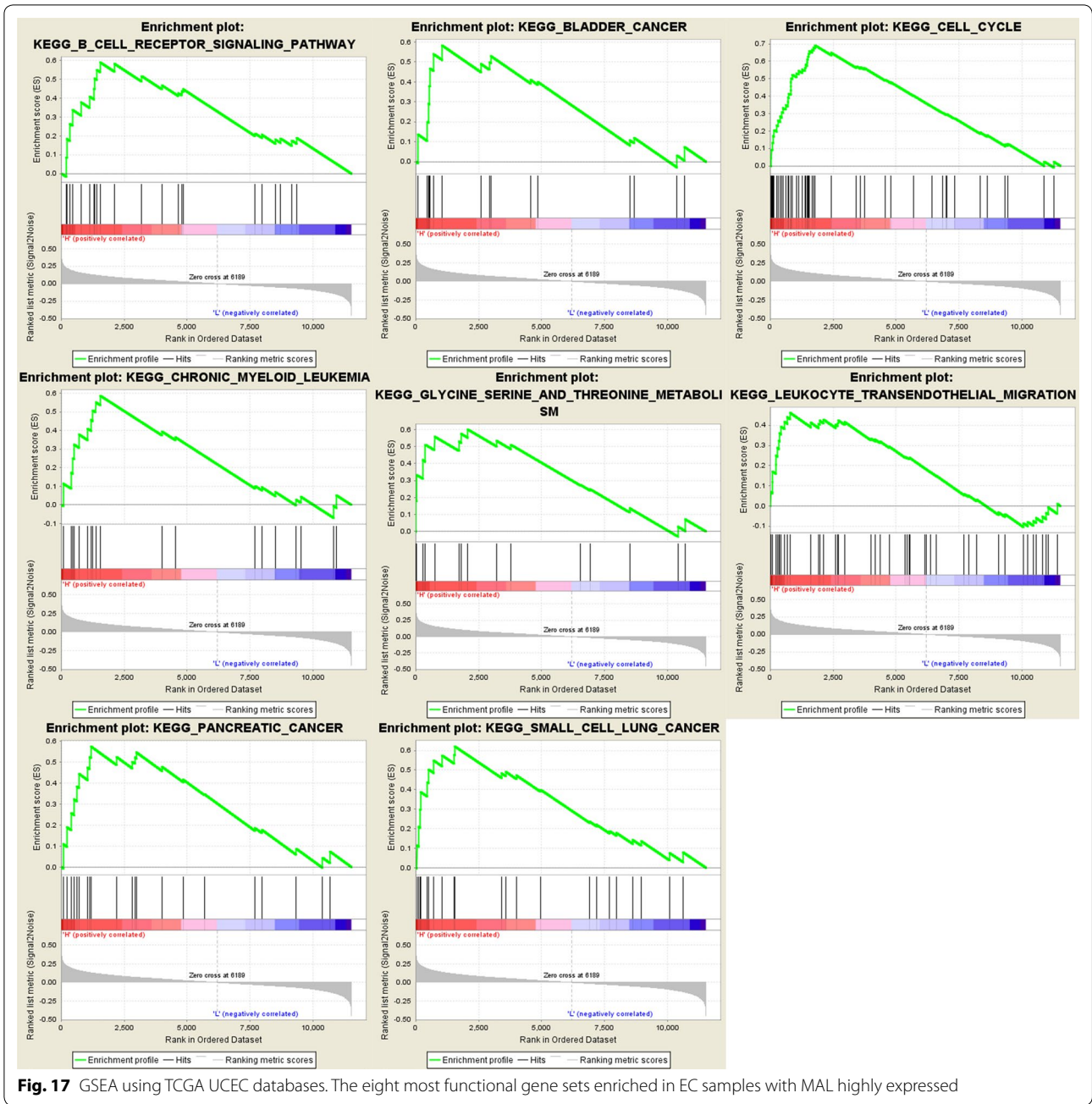


Fig. 16 GSEA using TCGA UCEC databases. The eight most functional gene sets enriched in EC samples with ASPM highly expressed



Conclusion

In conclusion, BCHE, MAL and ASPM may be potential prognostic markers in EC. Thioguanosine, resveratrol and trichostatin A can be used as antagonists against EC.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12935-020-1140-3>.

Additional file 1: Figure S1. Heatmap of top 200 genes in GSE17025.

Additional file 2: Figure S2. Heatmap of all genes in TCGA.

Additional file 3: Figure S3. The volcano plot of all DEGs in GSE17025. Red dots represent up-regulated genes, green dots represent down-regulated genes, and black dots represent genes without differential expression.

Additional file 4: Figure S4. The volcano plot of all DEGs in TCGA. Red dots represent up-regulated genes, green dots represent down-regulated genes, and black dots represent genes without differential expression.

Additional file 5: Figure S5. Results of CMap analysis.

Additional file 6: Figure S6. Univariate Cox proportional hazards regression analysis showed the top 10 EC-relative genes.

Additional file 7: Figure S7. Multivariate Cox proportional hazards regression analysis further screened out 6 hub genes.

Additional file 8: Figure S8. Expression of the six genes in low- and high-risk groups based on TCGA dataset. Red represents high-risk groups, blue represents low-risk groups.

Additional file 9: Figure S9. The heatmap of the six-gene expression levels between high- and low-risk groups in clinical information based on the TCGA dataset.

Abbreviation

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; ROC: Receiver operating characteristic; GSEA: Gene set enrichment analysis; RMA: Robust multi-array average; HR: Hazard ratio; CI: Confidence interval.

Acknowledgements

Not applicable.

Authors' contributions

JL was responsible for experimental design and data collection, SL was responsible for article writing and figures editing, MF and SN were responsible for clinical data collection and performing clinical validation. HW, SW, JQ and JZ were responsible for proofreading. WC was responsible for the review. All authors read and approved the final manuscript.

Funding

This work was supported by the National Nature Science Foundation of China (81472442, 81872119) and the Jiangsu province medical innovation team (CXTDA2017008).

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the Nanjing Medical University and samples were obtained with written informed consent from all patients, and all informed consent were conducted in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors report no conflicts of interest in this work.

Received: 15 June 2019 Accepted: 10 February 2020

Published online: 19 February 2020

References

- Fang F, Munck J, Tang J, Taverna P, Wang Y, Miller DF, Pilrose J, Choy G, Azab M, Pawelczak KS, et al. The novel, small-molecule DNA methylation inhibitor SGI-110 as an ovarian cancer chemosensitizer. *Clin Cancer Res*. 2014;20(24):6504–16.
- Braun MM, Overbeek-Wager EA, Grumbo RJ. Diagnosis and management of endometrial cancer. *Am Fam Physician*. 2016;93(6):468–74.
- Suri V, Arora A. Management of endometrial cancer: a review. *Rev Recent Clin Trials*. 2015;10(4):309–16.
- McAlpine JN, Temkin SM, Mackay HJ. Endometrial cancer: not your grandmother's cancer. *Cancer*. 2016;122(18):2787–98.
- Bendifallah S, Ballester M, Darai E. Endometrial cancer: predictive models and clinical impact. *Bull Cancer*. 2017;104(12):1022–31.
- Lee YC, Lheureux S, Oza AM. Treatment strategies for endometrial cancer: current practice and perspective. *Curr Opin Obstet Gynecol*. 2017;29(1):47–58.
- Tsimberidou AM. Targeted therapy in cancer. *Cancer Chemother Pharmacol*. 2015;76(6):1113–32.
- Tomczak K, Czerwinska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Poznan, Poland)*. 2015;19(1a):A68–77.
- Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. *PLoS Comput Biol*. 2008;4(8):e1000117.
- Xia L, Su X, Shen J, Meng Q, Yan J, Zhang C, Chen Y, Wang H, Xu M. ANLN functions as a key candidate gene in cervical cancer as determined by integrated bioinformatic analysis. *Cancer Manag Res*. 2018;10:663–70.
- Yuan L, Zeng G, Chen L, Wang G, Wang X, Cao X, Lu M, Liu X, Qian G, Xiao Y, et al. Identification of key genes and pathways in human clear cell renal cell carcinoma (ccRCC) by co-expression analysis. *Int J Biol Sci*. 2018;14(3):266–79.
- Lin Y, Liang R, Qiu Y, Lv Y, Zhang J, Qin G, Yuan C, Liu Z, Li Y, Zou D, et al. Expression and gene regulation network of RBM8A in hepatocellular carcinoma based on data mining. *Aging*. 2019;11(2):423–47.
- Clarke C, Madden SF, Doolan P, Aherne ST, Joyce H, O'Driscoll L, Gallagher WM, Hennessy BT, Moriarty M, Crown J, et al. Correlating transcriptional networks to breast cancer survival: a large-scale co-expression analysis. *Carcinogenesis*. 2013;34(10):2300–8.
- Day RS, McDade KK, Chandran UR, Lisovich A, Conrads TP, Hood BL, Kolli VS, Kirchner D, Litz T, Maxwell GL. Identifier mapping performance for integrating transcriptomics and proteomics experimental results. *BMC Bioinform*. 2011;12:213.
- Harbig J, Sprinkle R, Enkemann SA. A sequence-based identification of the genes detected by probesets on the Affymetrix U133 plus 2.0 array. *Nucleic Acids Res*. 2005;33(3):e31.
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics (Oxford, England)*. 2003;4(2):249–64.
- Gautier L, Cope L, Bolstad BM, Irizarry RA. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics (Oxford, England)*. 2004;20(3):307–15.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*. 2010;26(1):139–40.
- Galili T, O'Callaghan A, Sidi J, Sievert C. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics (Oxford, England)*. 2018;34(9):1600–2.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7.

22. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43(Database issue):D447–52.
23. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504.
24. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, Varambally S. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia (New York, NY).* 2017;19(8):649–58.
25. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):126.
26. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics (Oxford, England).* 2007;23(23):3251–3.
27. Liu J, Chen S, Wang W, Ning BF, Chen F, Shen W, Ding J, Chen W, Xie WF, Zhang X. Cancer-associated fibroblasts promote hepatocellular carcinoma metastasis through chemokine-activated hedgehog and TGF-beta pathways. *Cancer Lett.* 2016;379(1):49–59.
28. Pan B, Shen J, Cao J, Zhou Y, Shang L, Jin S, Cao S, Che D, Liu F, Yu Y. Interleukin-17 promotes angiogenesis by stimulating VEGF production of cancer cells via the STAT3/GiV signaling pathway in non-small-cell lung cancer. *Sci Rep.* 2015;5:16053.
29. Li H, Zhang YX, Liu Y, Wang Q. Effect of IL-17 monoclonal antibody Secukinumab combined with IL-35 blockade of Notch signaling pathway on the invasive capability of hepatoma cells. *Genet Mol Res GMR.* 2016. <https://doi.org/10.4238/gmr.15028174>.
30. Huda MN, Erdene-Ochir E, Pan CH. Assay for phosphorylation and microtubule binding along with localization of tau protein in colorectal cancer cells. *J Vis Exp JoVE.* 2017;128:e55932.
31. Zhang Q, Su R, Shan C, Gao C, Wu P. Non-SMC Condensin I Complex, Subunit G (NCAPG) is a novel mitotic gene required for hepatocellular cancer cell proliferation and migration. *Oncol Res.* 2018;26(2):269–76.
32. Song B, Du J, Song DF, Ren JC, Feng Y. Dysregulation of NCAPG, KNL1, miR-148a-3p, miR-193b-3p, and miR-1179 may contribute to the progression of gastric cancer. *Biol Res.* 2018;51(1):44.
33. Goto Y, Kurozumi A, Arai T, Nohata N, Kojima S, Okato A, Kato M, Yamazaki K, Ishida Y, Naya Y, et al. Impact of novel miR-145-3p regulatory networks on survival in patients with castration-resistant prostate cancer. *Br J Cancer.* 2017;117(3):409–20.
34. Vethe NT, Bremer S, Bergan S. IMP dehydrogenase basal activity in MOLT-4 human leukaemia cells is altered by mycophenolic acid and 6-thioguanosine. *Scand J Clin Lab Invest.* 2008;68(4):277–85.
35. Yang T, Zhang J, Zhou J, Zhu M, Wang L, Yan L. Resveratrol inhibits Interleukin-6 induced invasion of human gastric cancer cells. *Biomed Pharmacother Biomed Pharmacother.* 2018;99:766–73.
36. Zhong LX, Nie JH, Liu J, Lin LZ. Correlation of ARHI upregulation with growth suppression and STAT3 inactivation in resveratrol-treated ovarian cancer cells. *Cancer Biomarkers Sect A Dis Markers.* 2018;21(4):787–95.
37. Bartolacci C, Andreani C, Amici A, Marchini C. Walking a tightrope: a perspective of resveratrol effects on breast cancer. *Curr Protein Pept Sci.* 2018;19(3):311–22.
38. Kamal R, Chadha VD, Dhawan DK. Physiological uptake and retention of radiolabeled resveratrol loaded gold nanoparticles ((99 m) Tc-Res-AuNP) in colon cancer tissue. *Nanomed Nanotechnol Biol Med.* 2018;14(3):1059–71.
39. Ranganathan P, Rangnekar VM. Exploiting the TSA connections to overcome apoptosis-resistance. *Cancer Biol Ther.* 2005;4(4):391–2.
40. Hajji N, Wallenborg K, Vlachos P, Nyman U, Hermanson O, Joseph B. Combinatorial action of the HDAC inhibitor trichostatin A and etoposide induces caspase-mediated AIF-dependent apoptotic cell death in non-small cell lung carcinoma cells. *Oncogene.* 2008;27(22):3134–44.
41. Gu Y, Chow MJ, Kapoor A, Mei W, Jiang Y, Yan J, De Melo J, Seliman M, Yang H, Cutz JC, et al. Biphasic alteration of butyrylcholinesterase (BChE) during prostate cancer development. *Transl Oncol.* 2018;11(4):1012–22.
42. Bernardi CC, Ribeiro Ede S, Cavalli JJ, Chautard-Freire-Maia EA, Souza RL. Amplification and deletion of the AChE and BChE cholinesterase genes in sporadic breast cancer. *Cancer Genet Cytogenet.* 2010;197(2):158–65.
43. Aviello G, Corr SC, Johnston DG, O'Neill LA, Fallon PG. MyD88 adaptor-like (Mal) regulates intestinal homeostasis and colitis-associated colorectal cancer in mice. *Am J Physiol Gastrointest Liver Physiol.* 2014;306(9):G769–78.
44. Choi B, Han TS, Min J, Hur K, Lee SM, Lee HJ, Kim YJ, Yang HK. MAL and TMEM220 are novel DNA methylation markers in human gastric cancer. *Biomarkers.* 2017;22(1):35–44.
45. van Baars R, van der Marel J, Snijders PJ, Rodriguez-Manfredi A, ter Harmsel B, van den Munckhof HA, Ordi J, del Pino M, van de Sandt MM, Wentzensen N, et al. CADM1 and MAL methylation status in cervical scrapes is representative of the most severe underlying lesion in women with multiple cervical biopsies. *Int J Cancer.* 2016;138(2):463–71.
46. Zanotti L, Romani C, Tassone L, Todeschini P, Tassi RA, Bandiera E, Damia G, Ricci F, Ardighieri L, Calza S, et al. MAL gene overexpression as a marker of high-grade serous ovarian carcinoma stem-like cells that predicts chemoresistance and poor prognosis. *BMC Cancer.* 2017;17(1):366.
47. Lin SY, Pan HW, Liu SH, Jeng YM, Hu FC, Peng SY, Lai PL, Hsu HC. ASPM is a novel marker for vascular invasion, early recurrence, and poor prognosis of hepatocellular carcinoma. *Clin Cancer Res.* 2008;14(15):4814–20.
48. Xie JJ, Zhuo YJ, Zheng Y, Mo RJ, Liu ZZ, Li BW, Cai ZD, Zhu XJ, Liang YX, He HC, et al. High expression of ASPM correlates with tumor progression and predicts poor outcome in patients with prostate cancer. *Int Urol Nephrol.* 2017;49(5):817–23.
49. Bruning-Richardson A, Bond J, Alsiary R, Richardson J, Cairns DA, McCormack L, Hutson R, Burns P, Wilkinson N, Hall GD, et al. ASPM and microcephalin expression in epithelial ovarian cancer correlates with tumour grade and survival. *Br J Cancer.* 2011;104(10):1602–10.
50. Alsiary R, Bruning-Richardson A, Bond J, Morrison EE, Wilkinson N, Bell SM. Deregulation of microcephalin and ASPM expression are correlated with epithelial ovarian cancer progression. *PLoS ONE.* 2014;9(5):e97059.
51. Wang WY, Hsu CC, Wang TY, Li CR, Hou YC, Chu JM, Lee CT, Liu MS, Su JJ, Jian KY, et al. A gene expression signature of epithelial tubulogenesis and a role for ASPM in pancreatic tumor progression. *Gastroenterology.* 2013;145(5):1110–20.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.