



A bioinformatic analysis of WFDC2 (HE4) expression in high grade serous ovarian cancer reveals tumor-specific changes in metabolic and extracellular matrix gene expression

Nicole E. James^{1,2} · Megan Gura³ · Morgan Woodman¹ · Richard N. Freiman³ · Jennifer R. Ribeiro^{1,2}

Received: 27 October 2021 / Accepted: 22 January 2022 / Published online: 15 May 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Human epididymis protein-4 (HE4/WFDC2) has been well-studied as an ovarian cancer clinical biomarker. To improve our understanding of its functional role in high grade serous ovarian cancer, we determined transcriptomic differences between ovarian tumors with high- versus low-WFDC2 mRNA levels in The Cancer Genome Atlas dataset. High-WFDC2 transcript levels were significantly associated with reduced survival in stage III/IV serous ovarian cancer patients. Differential expression and correlation analyses revealed secretory leukocyte peptidase inhibitor (SLPI/WFDC4) as the gene most positively correlated with WFDC2, while A kinase anchor protein-12 was most negatively correlated. WFDC2 and SLPI were strongly correlated across many cancers. Gene ontology analysis revealed enrichment of oxidative phosphorylation in differentially expressed genes associated with high-WFDC2 levels, while extracellular matrix organization was enriched among genes associated with low-WFDC2 levels. Immune cell subsets found to be positively correlated with WFDC2 levels were B cells and plasmacytoid dendritic cells, while neutrophils and endothelial cells were negatively correlated with WFDC2. Results were compared with DepMap cell culture gene expression data. Gene ontology analysis of k-means clustering revealed that genes associated with low-WFDC2 were also enriched in extracellular matrix and adhesion categories, while high-WFDC2 genes were enriched in epithelial cell proliferation and peptidase activity. These results support previous findings regarding the effect of HE4/WFDC2 on ovarian cancer pathogenesis in cell lines and mouse models, while adding another layer of complexity to its potential functions in ovarian tumor tissue. Further experimental explorations of these findings in the context of the tumor microenvironment are merited.

Keywords HE4 · WFDC2 · SLPI · Ovarian cancer · The Cancer Genome Atlas · Tumor microenvironment

Abbreviations

CCLE Cancer Cell Line Encyclopedia
CNA Copy number alteration

DEG Differentially expressed gene
ECM Extracellular matrix
FPKM Fragments Per Kilobase of transcript per Million mapped reads
GEO Gene Expression Omnibus
GO Gene ontology
HGSOC High grade serous ovarian cancer
HR Hazard ratio
OS Overall survival
OXPHOS Oxidative phosphorylation
PBMC Peripheral blood mononuclear cells
PFS Progression-free survival
TCGA The cancer genome atlas
TPM Transcripts per million

Nicole E. James and Megan Gura are equal contributors to this study.

✉ Jennifer R. Ribeiro
jrribeiro@wihri.org

¹ Department of Obstetrics and Gynecology, Program in Women's Oncology, Women and Infants Hospital of Rhode Island, 200 Chestnut St, room 208, Providence, RI 02905, USA

² Warren-Alpert Medical School of Brown University, Providence, RI 02903, USA

³ Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, Providence, RI 02912, USA

Introduction

Human epididymis protein-4 (HE4), encoded by the gene whey acidic protein (WAP) four-disulfide core domain 2 (*WFDC2*), has been extensively studied as biomarker in ovarian cancer [1, 2]. Serum HE4, when used in combination with serum CA125 in the Risk of Ovarian Malignancy Algorithm (ROMA), demonstrates significantly greater detection sensitivity at a set specificity of 75% than the Risk of Malignancy Index (RMI), which uses pelvic imaging, menopausal status, and serum CA125 values. Importantly, serum HE4 levels are less commonly elevated in benign gynecologic conditions than CA125 [2]. Moreover, *in vitro* and *in vivo* analyses have shown that HE4 promotes ovarian cancer cell growth, extracellular matrix production, metastatic properties, angiogenesis, and immune evasion through regulation of signaling pathways, including focal adhesion kinase (FAK), extracellular signal-regulated kinase (ERK), and signal transducer and activator of transcription 3 (STAT3) (reviewed in [3]). Our previous studies have uncovered a role for HE4 in promoting chemoresistance and extracellular matrix gene expression in ovarian cancer cells, as well as pro-angiogenic gene expression in immune cells [4–6]. Nonetheless, a complete understanding of how this enigmatic and multi-functional protein functions in the context of the human ovarian tumor microenvironment to promote chemoresistance and worse survival outcomes remains undetermined. While numerous studies have explored its use as a diagnostic and prognostic serum biomarker, no study to date has examined the transcriptomic signature associated with high- and low-*WFDC2* mRNA levels in high grade serous ovarian cancer (HGSOC) patient tumors.

The complexity of the ovarian cancer microenvironment highlights the importance of studying a particular protein in the context of this microenvironment in order to fully understand its significance in human ovarian tumors. Ovarian cancer is unique in that tumorigenic cells originating from the fallopian tubes “seed” onto the ovary, where they find an attractive niche to promote their growth. One factor that may aid in the production of this ovarian niche is the development of fibrosis, which occurs with advancing age [7]. Cancer associated fibroblasts (CAFs) promote tumorigenesis through production of cytokines, growth factors, and extracellular matrix components, leading to worsening fibrosis and tumor growth [8]. Subsequent to the growth of cancer cells in the ovary, these cells are susceptible to spread within the peritoneal cavity, homing to the highly vascular, fatty tissue of the omentum that uses immunologic and metabolic mechanisms to promote tumor growth [9]. In order to understand the role of HE4/*WFDC2* in ovarian cancer, it is important to consider its known functions and

how these might interplay with this complex ovarian tumor microenvironment.

HE4 is a small secreted glycoprotein and protease inhibitor that was originally identified in the epididymis and suggested to be important in sperm maturation. It was later found to be expressed in several tissue types, namely trachea, salivary gland, kidney, lung, and nasal epithelium, and may be involved in host defense of airways [10–12]. In addition to playing a role in ovarian cancer, it is a key fibroblast mediator of kidney fibrosis, inhibiting the ability of matrix metalloproteinases to degrade collagen I, and is also associated with fibrotic diseases of the lung and liver [13–17]. Its role in promoting fibrosis is a critical component to examine given the degree of desmoplastic, fibrotic matrix that characterizes EOC.

To begin investigating the function of HE4/*WFDC2* in the context of the ovarian tumor microenvironment, we performed gene expression analysis of 378 HGSOC tumors with RNA-sequencing (RNA-seq) data available through The Cancer Genome Atlas (TCGA), and compared differential gene expression in samples with high- versus low-*WFDC2* mRNA levels. We compared these findings to DepMap gene expression correlation data for serous ovarian cancer cell lines. We identified strong co-regulation of *WFDC2* and two other WAP-family members, secretory leukocyte peptidase inhibitor (*SLPI/WFDC4*) and peptidase inhibitor-3 (*PI3/WFDC14/elafin*), suggesting an important role for this family of protease inhibitors in promoting ovarian tumor pathogenesis. Moreover, our results reveal a potential novel role for HE4 in modulating oxidative phosphorylation metabolic gene expression in HGSOC. Our data support previously identified HE4 functions in mediating extracellular matrix, angiogenesis, and immune response in ovarian cancer cells, while also highlighting the critical importance of examining HE4 effects in the context of the tumor microenvironment.

Results

HE4 (*WFDC2*) expression correlates with clinical survival outcomes

As a first exploration of *WFDC2* expression in the ovarian cancer TCGA dataset (TCGA-OV), we compared *WFDC2* mRNA expression with HE4 protein levels using the cBioPortal, revealing a strong correlation between mRNA and protein (Spearman $r=0.74$, $p=2.56e-19$) (Fig. 1A). These results suggest that *WFDC2* mRNA levels are a relatively accurate representation of its protein expression in human ovarian tumors. Secondly, we examined mRNA expression in relationship to copy number alterations (CNAs). Five percent of 594 samples possess putative copy number

amplifications, and CNAs were correlated with mRNA expression, with lower mRNA in the samples with shallow deletion, and higher mRNA in the samples with amplification. Most samples were diploid or possessed copy number gains in *WFDC2* (Fig. 1B).

Next, as it has been reported that lower tumor mutation count is related to decreased progression-free survival (PFS) and overall survival (OS) in ovarian cancer [18], we sought to explore the relationship between *WFDC2* and tumor mutation count. There was a small but significant decrease in *WFDC2* mRNA expression in tumors with high (43–158) versus low (8–43) raw mutation count, which was also observed in a slight, but non-significant inverse correlation between *WFDC2* mRNA expression and mutation count (Spearman $r = -0.14$, $p = 0.058$) (Fig. 1C).

To explore the relationship of *WFDC2* mRNA with clinical outcomes, we performed Kaplan–Meier analysis for *WFDC2* in TCGA-OV and Gene Expression Omnibus (GEO) Series. There was no significant difference in PFS in *WFDC2*-high versus -low groups when examining all serous samples, and only a non-significant trend toward worse OS in patients with high-*WFDC2* (HR = 1.17, $p = 0.083$) (data not shown). Next, we narrowed down our analysis to stage III and IV, grade 2 and 3 ($n = 975$). Using the upper quartile cutoff, we again found no significant difference between *WFDC2*-high and -low groups with regards to PFS; however, *WFDC2*-high patients had significantly worse OS (HR = 1.22, $p = 0.046$) (Fig. 1D). When this group was narrowed further to include only optimally debulked patients ($n = 495$), the *WFDC2*-high group showed decreased PFS (HR = 1.31, $p = 0.037$) and OS (HR = 1.38, $p = 0.027$) (Fig. 1E). Collectively, these results agree with previously published studies showing HE4 serum levels are prognostic for ovarian cancer patients [19], and also suggests that patients with suboptimal debulking may experience poor outcomes regardless of HE4 levels.

Differential gene expression reveals a positive correlation of *WFDC2* and *SLPI* across many cancers

We next performed differential gene expression analysis in TCGA-OV dataset, which was split into high- and low-*WFDC2* groups according to median *WFDC2* fragments per kilobase of transcript per million mapped reads (FPKM) levels (Table S1). Principal component analysis (PCA) revealed no strong grouping of the samples, which is unsurprising given the large population with a high degree of biological variability inherently present (Fig. 2A). We identified 512 significant differentially expressed genes (DEGs) ($p\text{-adj.} < 0.05$, \log_2 fold-change ≥ 10.51 , protein-coding), with 399 DEGs corresponding to high-*WFDC2* expression (“high-*WFDC2* DEGs”) and 113 DEGs corresponding to

low-*WFDC2* expression (“low-*WFDC2* DEGs”) (Table S2). A volcano plot was generated to show the top five DEGs, which reassuringly included *WFDC2* (Fig. 2B). Notably, another WAP-domain containing protein, secretory leukocyte peptidase inhibitor (*SLPI/WFDC4*) was also among the top five DEGs that were associated with high-*WFDC2* levels. We then performed correlation analyses of all DEGs with \log_2 fold-change of ≥ 0.5 in either direction. The cBioPortal co-expression feature was used to generate Spearman r values using RNA Seq V2 RSEM data. As expected, the direction of correlations matched well with the differential gene expression analysis (data not shown). All DEGs that were significantly correlated ($p < 0.01$) with Spearman $r \geq 0.3$ were compared by heatmap analysis to fold-change values, displaying a high degree of similarity between the differential gene expression analysis and correlation analysis results (Fig. 2C).

As previously noted, *SLPI* emerged as a top high-*WFDC2* DEG that was also the most strongly positively correlated gene (Spearman $r = 0.59$, $p = 6.9e-30$). A kinase anchor protein-12 (*AKAP12*) emerged as the low-*WFDC2* DEG that was most strongly negatively correlated with *WFDC2* (Spearman $r = -0.36$, $p = 1.09e-10$) (Fig. 2D–G). Since *SLPI* was very strongly correlated with *WFDC2* and is also a WAP-domain containing protease inhibitor, we suspected that these two proteins may show a tendency toward co-regulation. We performed a pan-cancer correlation analysis of *WFDC2* and *SLPI*, which revealed their strong correlation across many cancers, in particular pancreatic adenocarcinoma (PADD), thymoma (THYM), and uterine carcinosarcoma (UCS) (Fig. 2H). While *SLPI* has been reported to play a role in ovarian cancer pathogenesis and is associated with worse outcomes in ovarian cancer patients [20–22], there is no clear understanding of the potential overlapping or divergent roles *HE4/WFDC2* and *SLPI* may have in regulating tumorigenic properties.

In addition to *SLPI*, the gene peptidase inhibitor-3 (*PI3*), encoding the protein elafin, was also among the top *WFDC2*-high DEGs that was strongly correlated with *WFDC2* ($r = 0.386$, $p = 2.39e-12$). Elafin is also a WFDC protein that is overexpressed in ovarian cancer and related to poor outcomes [23, 24]; however, even less is known about elafin’s role in ovarian cancer than *SLPI*’s. Nonetheless, the co-regulation of these three WFDC proteins in ovarian cancer suggests an important role for this protein family in this disease.

Gene ontology analysis implicates metabolism and extracellular matrix correlations with *WFDC2* mRNA expression

Gene ontology (GO) analysis was performed on the DEGs of high- and low-*WFDC2* expressing tumors. The complete

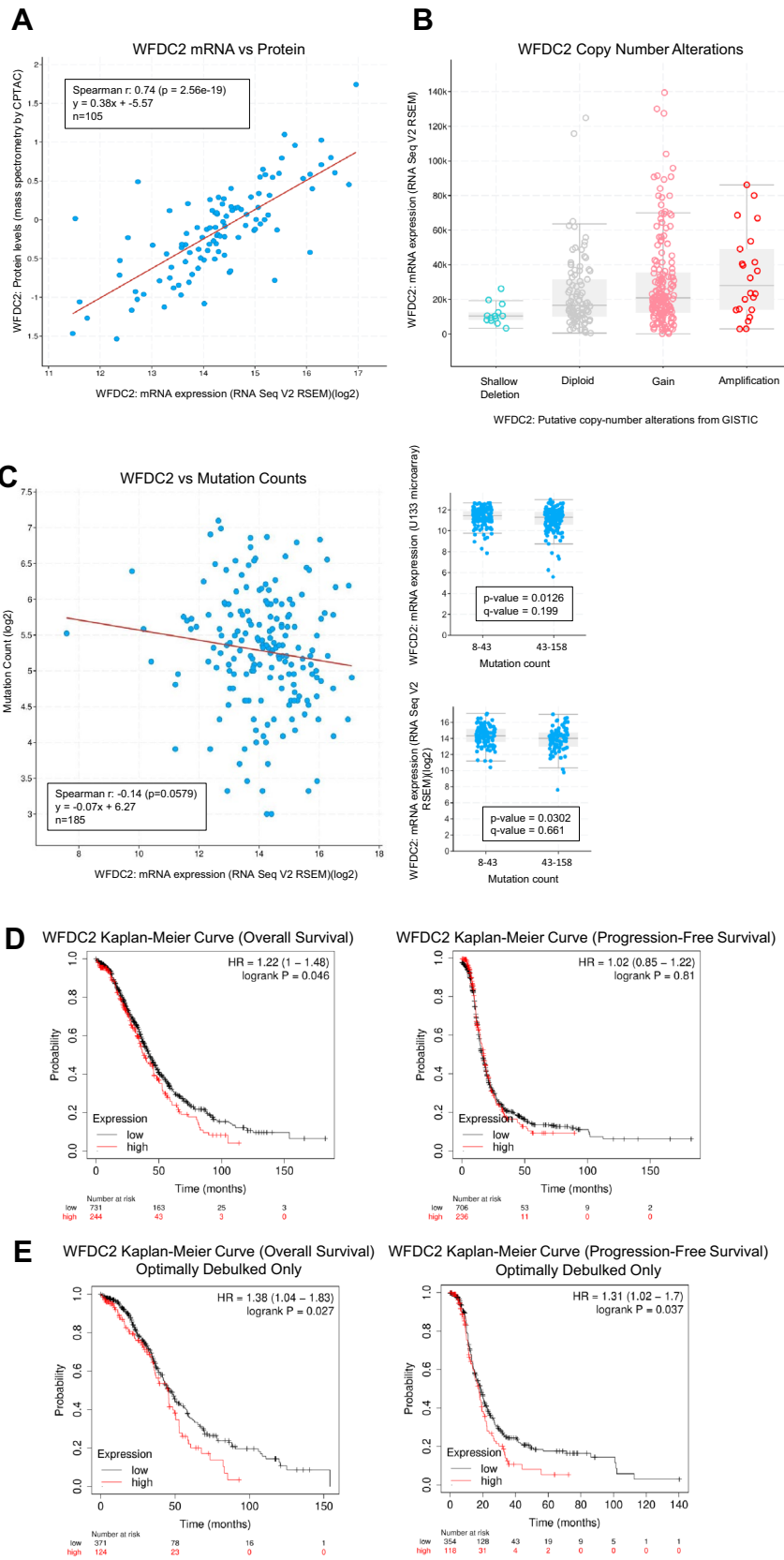


Fig. 1 HE4 (*WFDC2*) expression correlates with clinical survival outcomes **A** *WFDC2* mRNA and HE4 protein levels were correlated by Spearman rank correlation analysis using all samples of the Firehose Legacy cohort with mRNA (RNA Seq V2 RSEM) and protein (mass spectrometry by CPTAC) levels available ($n=105$), showing strong correlation between mRNA and protein. **B** Putative copy number alterations (CNAs) were plotted against *WFDC2* mRNA (RNA Seq V2 RSEM) levels in the Firehose Legacy cohort. A majority of samples were diploid ($n=95$) or had CNA gains ($n=169$). Median mRNA levels generally corresponded to CNAs. **C** *WFDC2* mRNA levels (RNA Seq V2 RSEM) were correlated to mutation counts using Spearman rank correlation (left panel). Firehose Legacy TCGA samples were split into low (8–43) and high (43–158) mutation count groups and median *WFDC2* mRNA levels (RNA Seq V2 RSEM and U133 microarray) determined for each group (right panels). There was a small, non-significant inverse correlation between *WFDC2* and mutation count, as well as a small, but significant decrease in *WFDC2* mRNA levels in patients with fewer mutations. **D** Kaplan–Meier curves for overall survival and progression-free survival were determined for *WFDC2* using all cohorts (GEO Series and TCGA) available for ovarian cancer at <http://KMplot.com>. Analysis was restricted to serous Stage III and IV, grade 2 and 3. Patients with higher *WFDC2* levels had worse overall survival. **E** Kaplan–Meier curves for overall survival and progression-free survival were determined for *WFDC2* using all cohorts (GEO Series and TCGA) available for ovarian cancer at <http://KMplot.com>. Analysis was restricted to serous Stage III and IV, grade 2 and 3, optimally debulked only. Patients with higher *WFDC2* levels had worse overall survival and progression-free survival

lists of enriched categories are shown in Tables S3 and S4. Strikingly, categories related to oxidative phosphorylation (OXPHOS)/mitochondrial metabolism were highly enriched in high-*WFDC2* DEGs. Categories termed “antimicrobial humoral response” and “neutrophil activation” were also significantly enriched, which is interesting in light of a previously reported role for HE4 in regulating innate immunity of the respiratory tract [10] (Fig. 3A, Table S3). Of note, in the “neutrophil activation” category, CXCL8 (Interleukin-8; IL8), was among the identified DEGs, which is in agreement with our previous study reporting on the regulation of CXCL8/IL8 by HE4 in immune cells [6].

GO assessment of low-*WFDC2* DEGs revealed enrichment for categories related to extracellular matrix, vascular development, epithelial cell proliferation, and ERK signaling (Fig. 3B, Table S4). These results were particularly surprising given the consistent reports of the stimulatory role of *WFDC2* in metastasis, angiogenesis, proliferation, and ERK signaling [3, 6], raising the possibility of negative feedback mechanisms producing this unexpected result. In support of this hypothesis, we noted that expression of tenascin-C (*TNC*), which we have previously found to be upregulated by HE4 overexpression or treatment in OVCAR8 ovarian cancer cells [5], was a low-*WFDC2* DEG. Alternatively, the enrichment of these categories may occur through selective alterations in genes that negatively regulate these processes. In support of this hypothesis, we noted that early growth response-1 (*EGR1*) was a low-*WFDC2* DEG. We have

previously reported that cisplatin-induced *EGR1* expression was suppressed in *WFDC2*-overexpressing cells [4], supporting a role for HE4 in suppressing the apoptosis-promoting effects of *EGR1* in response to chemotherapy. Moreover, *AKAP12* was listed in the “ERK1 and ERK2 cascade” category, and is described as a tumor suppressor known to suppress the ERK signaling pathway [25]. Collectively, HE4 may regulate these various pathways through a combination of feedback mechanisms or negative regulatory approaches, which in some cases may be more accurately represented in the complex tumor microenvironment than in ovarian cancer cell lines.

Survival outcomes related to top correlated DEGs

We then performed Kaplan–Meier analyses for OS for all of the top correlated DEGs. To keep the comparison between *WFDC2* and these genes comparable, we used the same parameters we found to produce the most prognostic results for *WFDC2* in Fig. 1 (stage III, IV; grade 2, 3 disease, optimally debulked, top quartile cutoff). All genes with hazard ratios (HR) ≤ 0.75 or ≥ 1.5 (log-rank $p < 0.01$) are shown in Fig. 4A–N. Several genes had higher HRs than *WFDC2*, including *SLPI* (HR = 1.63 [1.23–2.16], $p = 0.00056$) and several NADH:ubiquinone oxidoreductase (NDU) family genes. Using average expression of all these NDU genes also demonstrated a significant HR of 1.8 [1.36–2.4] ($p = 3.6e-5$) (Fig. 4M). The most prognostic gene was reactive oxygen species modulator-1 (*ROMO1*), with an HR of 2.7 [1.6–4.55], $p = 0.00011$ (Fig. 4D). These data suggest that the coordinate expression of specific sets of genes related to *WFDC2* expression may indicate patient clinical outcomes in EOC.

WFDC2 levels influence immune cell infiltration

Using TIMER 2.0, we examined all immune deconvolution methods to determine immune cell populations significantly correlating with *WFDC2* expression in TCGA. B cells (TIMER) and plasmacytoid dendritic cells (XCELL) were positively correlated with *WFDC2* levels (Spearman $r = 0.243$, $p = 1.02e-04$ and Spearman $r = 0.221$, $p = 4.41e-04$, respectively).

Conversely, neutrophils (MCPCOUNTER) (Spearman $r = -0.278$, $p = 8.45e-06$), and endothelial cells (MCP-COUNTER, EPIC) (Spearman $r = -0.309$, $p = 6.43e-07$ and Spearman $r = -0.206$, $p = 1.07e-03$, respectively) were significantly negatively correlated with *WFDC2* levels, however the correlations were overall weak (Fig. 5A–F). The putative reduction of endothelial cells in *WFDC2*-high tumors could reflect hypoxia-induced increases in *WFDC2* levels, as has been reported to occur during renal fibrosis and in gastric cancer [26, 27].

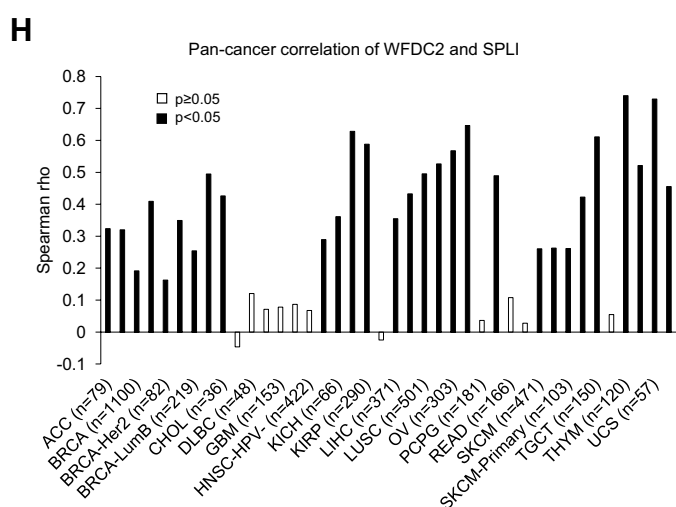
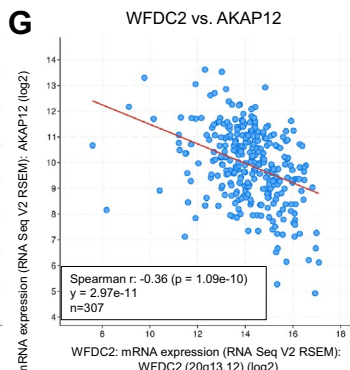
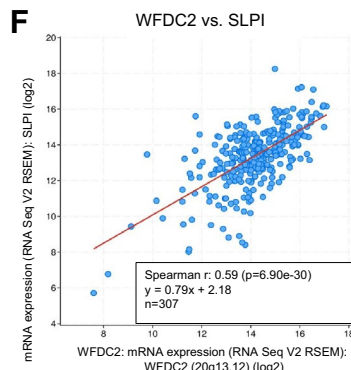
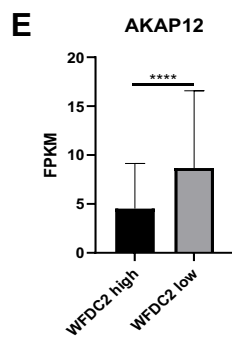
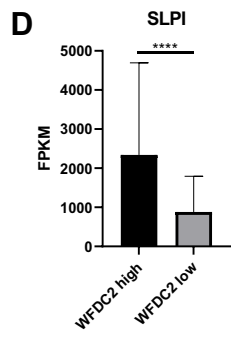
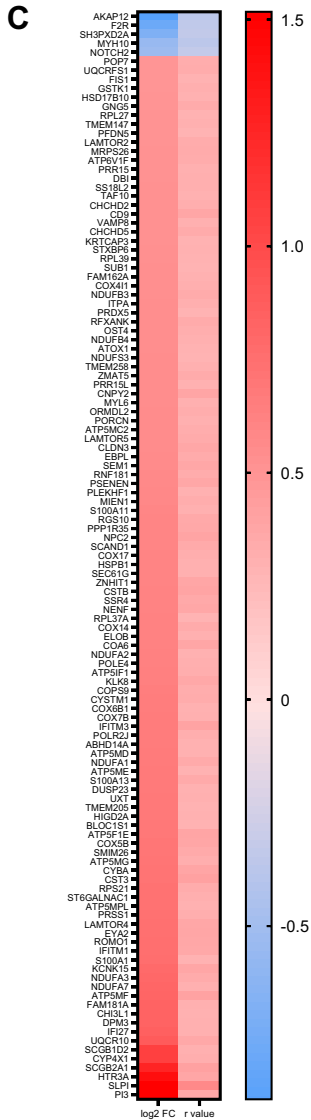
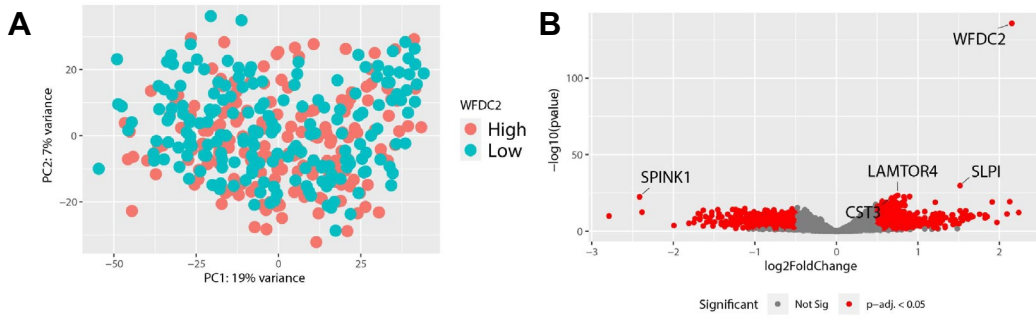


Fig. 2 Differential gene expression reveals a positive correlation of *WFDC2* and *SLPI* across many cancers. **A** Principal component analysis (PCA) of all TCGA-OV samples. **B** Volcano plot analysis showing top five differentially expressed genes (DEGs) between high- and low-*WFDC2* levels. Protein-coding genes significantly changed ($p\text{-adj.} < 0.05$) with \log_2 fold-change ≥ 0.5 in either direction are shown as red dots. **C** All DEGs were correlated with *WFDC2* in cBioPortal. \log_2 fold-change (\log_2 FC) and Spearman r -values are represented in a heat map side-by-side comparison for all genes that significantly correlated with *WFDC2* ≥ 0.3 in either direction. Fold-change data versus correlation data show a high degree of similarity. **D, E** *SLPI* was determined to be the high-*WFDC2* DEG that most strongly correlated with *WFDC2* in the Firehose Legacy cohort, while *AKAP12* was the low-*WFDC2* DEG most negatively correlated with *WFDC2*. Average FPKM values for *SLPI* and *AKAP12* were plotted for *WFDC2*-high versus *WFDC2*-low samples. $****p < 0.0001$ **F, G** Spearman rank correlations are shown for *SLPI* and *AKAP12*. **H** Pan-cancer Spearman rank correlation analysis of *WFDC2* and *SLPI*

Comparison of TCGA data with DepMap ovarian cancer cell line data

We next went on to compare DepMap Cancer Cell Line Encyclopedia (CCLE) cell line expression data at <https://depmap.org/portal/> (Table S5). First, we analyzed *SLPI* and *PI3* (elafin) correlation data for 24 HGSOC cell lines. We found these genes to be strongly and significantly correlated with *WFDC2*, confirming their relationship with *WFDC2* (Fig. 6A, B). Next, we stratified the cell lines according to median *WFDC2* transcripts per million (TPM) and then analyzed the gene expression of the top five low- and top five high-*WFDC2* expressing cell lines using iDEP.94 k-means clustering [28]. PCA revealed a good clustering of low- versus high-*WFDC2* cell lines (Fig. 6C). Next, we performed k-means clustering using the top 500 most variable genes (Fig. 6D, Table S6). Gene ontology analysis of the resulting clusters revealed that the genes associated with low-*WFDC2* were enriched for categories including extracellular matrix, cell-substrate adhesion, and angiogenesis, which matched the enriched categories we found in our TCGA analysis (Fig. 6E, Table S7). When examining the genes associated with high-*WFDC2*, we found the enriched categories were involved in epidermis development, epithelial cell proliferation, and negative regulation of peptidase activity (Fig. 6F, Table S8). These results are not entirely unexpected, since HE4 is a protease inhibitor with a known role in promoting proliferation [3]; however, these categories differ from enriched categories in our TCGA analysis. It is possible these differences are related to the lack of tumor microenvironment interactions and hypoxia in the cell lines that are present in ovarian tumors. Overall, these results confirm the role of HE4 in regulating extracellular matrix functions, but suggest it may also have additional effects in an in vivo context versus in vitro.

Discussion

In this analysis, we identified *SLPI* as the gene most prominently correlated with *WFDC2*. In differential gene expression analysis, it was also among the top five genes with significantly higher levels in HGSOC samples with high-*WFDC2* levels. *SLPI* and HE4 are both WFDC proteins with reported anti-protease functions [11]. While HE4 has been extensively studied as a biomarker and for its tumor promoting functions, *SLPI* has received overall less attention. Our results confirmed that *SLPI* is a prognostic factor in ovarian cancer, with its transcript levels showing better prognostication than *WFDC2*. We also determined that *SLPI* is broadly correlated with *WFDC2* across cancer types, suggesting that these two genes are frequently co-regulated. One unanswered question regarding the functions of *SLPI* and HE4 is whether they possess redundant anti-protease functions or if their functions are unique. Supporting their potential unique roles is one study showing their co-expression in similar tissues of the respiratory tract, but divergent subcellular localizations [10]. The question of these two proteins' overlapping versus divergent functions will be important to answer if either protein were to be targeted therapeutically.

One surprising result from gene ontology analysis of high-*WFDC2* DEGs was the enrichment of OXPHOS/mitochondrial metabolism categories. No previous studies in ovarian cancer cell lines or mouse models have identified a potential relationship between HE4 and metabolism. While previous dogma stated that tumors undergo a glycolytic switch whereby they move from relying on OXPHOS to glycolysis, more recent studies show that OXPHOS is an important metabolic pathway in several tumor types [29]. In ovarian cancer, OXPHOS is important in mediating resistance to chemotherapy [30–32]. Interestingly, one study also reported that ovarian cancer stem cells prioritize OXPHOS for their metabolism [33], suggesting high OXPHOS metabolism to be a potential defining characteristic of a subset of chemoresistant cancer stem cells. Single-cell sequencing of ovarian tumors identified epithelial and stromal cell subsets in ovarian tumors, with the epithelial subset being characterized by OXPHOS genes [34]. These studies point to the importance of OXPHOS as a key energy production pathway in ovarian cancer and our results raise the possibility of a connection between HE4/*WFDC2* and hypoxia, fibrosis, and OXPHOS metabolism. However, the interplay between these various features remains to be elucidated.

Other enriched categories in high-*WFDC2* DEGs were related to immune response, including “antimicrobial humoral immunity”, and “neutrophil activation”. An early study on HE4 reported its potential role in innate immunity and epithelial host defense in the respiratory tract [10].

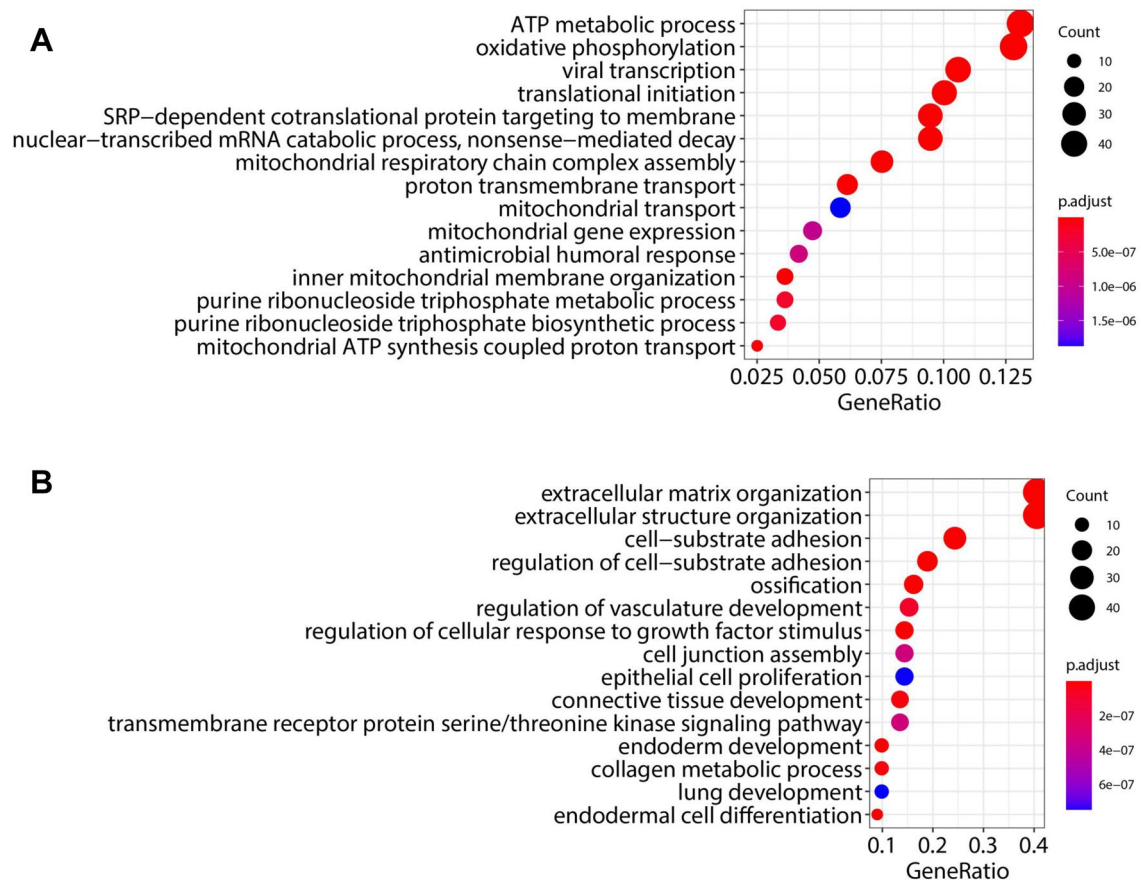


Fig. 3 Gene ontology analysis implicates metabolism and extracellular matrix correlations with *WFDC2* mRNA expression. **A** Gene ontology analysis was performed for all DEGs associated with high-*WFDC2*, revealing enrichment in categories related to metabolism/oxidative phosphorylation. Number of genes in each category (“Count”) are indicated by circle size, while adjusted p-value (“p.adjust”) is indicated by color. **B** Gene ontology analysis was performed for all differentially expressed genes associated with low-*WFDC2*, revealing enrichment in categories related to extracellular matrix, vascular development, and proliferation. Number of genes in each category (“Count”) are indicated by circle size, while adjusted p-value (“p.adjust”) is indicated by color.

adjust”) is indicated by color. **B** Gene ontology analysis was performed for all differentially expressed genes associated with low-*WFDC2*, revealing enrichment in categories related to extracellular matrix, vascular development, and proliferation. Number of genes in each category (“Count”) are indicated by circle size, while adjusted p-value (“p.adjust”) is indicated by color.

Recently, a correlation between innate immune response and serum HE4 was confirmed in COVID-19 patients [35]. We have also studied its role in regulating gene expression in immune cells, finding that it regulates several important immune genes and promotes secretion of Interleukin-8 (*IL8/CXCL8*) from peripheral blood mononuclear cells (PBMCs), and contributes to suppression of PBMC cytotoxicity against cancer cells [6, 36, 37]. Given that *IL8/CXCL8*, which was also a high-*WFDC2* DEG in this current analysis, is a potent neutrophil chemotactic factor [38], it made sense that categories related to neutrophil activation and degranulation were enriched. However, somewhat contradictorily, we observed a weak association with reduced neutrophil infiltration according to the TIMER analysis. Further studies are required to query this TIMER result and elucidate the reasoning behind these contradictory findings.

Several interesting GO categories were enriched in the low-*WFDC2* DEG group, with a heavy focus on categories

related to extracellular matrix (ECM) organization. As previously mentioned, this enrichment of ECM categories in the low-*WFDC2* group could result from negative feedback mechanisms, or through downregulation of genes that negatively regulate these processes. Notably, many collagen genes were downregulated, which could signal negative feedback mechanisms or an imbalance of various collagen types, leading to a fibrotic phenotype. The relative amounts and spatial location of various collagens may play a role in the development of fibrosis [39]. Furthermore, an interesting connection between OXPHOS categories and ECM categories exists, with studies reporting that metabolic functions and ECM regulation are intricately linked. Modulation of OXPHOS in osteosarcoma cells resulted in significant changes in ECM genes [40], and triple-negative breast cancer demonstrated increased OXPHOS with decreasing collagen density [41]. Metformin, a drug used to treat diabetes and a known inhibitor of fibrosis [42, 43] and OXPHOS [44], suppresses ovarian cancer growth and metastasis [45],

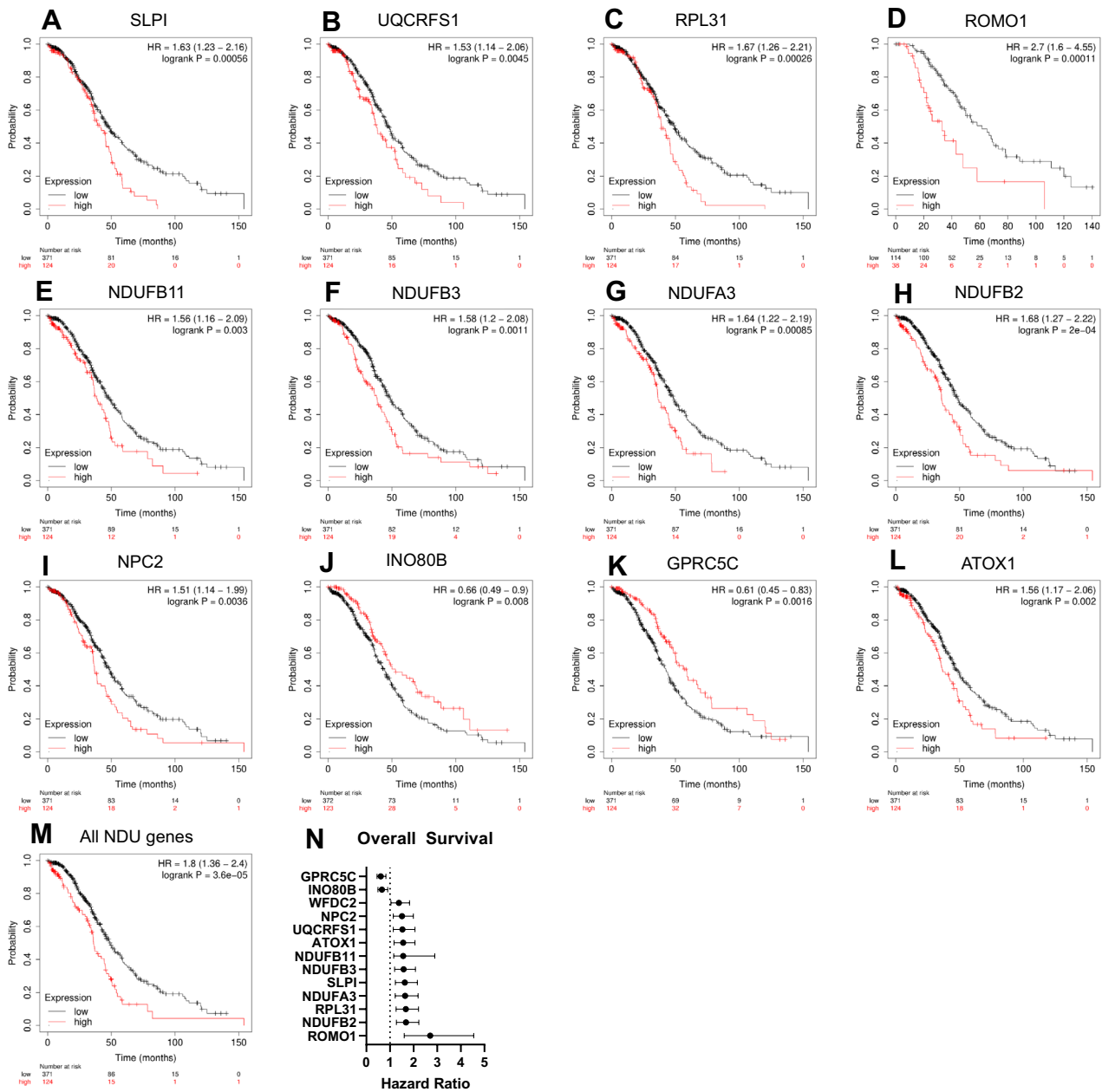


Fig. 4 Survival outcomes related to top correlated DEGs. Kaplan–Meier curves were generated at <http://KMplot.com> for all available datasets (TCGA and GEO Series). Top DEGs that were most associated with overall survival (OS) are shown in **A–L** (hazard ratio

(HR) > 1.5; $p < 0.01$). **M** All NADH:Ubiquinone (NDU) genes were combined and analyzed by Kaplan–Meier. **N** Summary of hazard ratios

further supporting the connection between these tumor features. Further studies are needed to determine if HE4-mediated changes in ECM regulate metabolic pathways in HGSOC.

“Regulation of response to growth factor stimulus” and the “ERK1 and ERK2 cascade” were two other enriched categories in the low-*WFDC2* DEG group. It has been well-documented that HE4 promotes activation of the ERK pathway and other signaling pathways in ovarian cancer

(reviewed in [3]). Therefore, a possible explanation for this GO finding can be found by examining the specific genes found in the ERK enrichment group, which include *AKAP12* and elastin microfibril interfacer-1 (*EMILIN1*)—both negative regulators of the ERK pathway [25, 46]. Finally, the “vascular development” category was also enriched in the low-*WFDC2* DEG group, which highly overlapped with ECM genes. *EGR1*, which mediates apoptosis, was also listed in this category, which supports our previous study

finding HE4-mediated suppression of *EGR1* expression following cisplatin treatment [4]. It is unclear whether the presence of this category actually signals a reduction in vascular formation or angiogenesis, which would disagree with our previous study showing an angiogenesis promoting effect of HE4 via regulation of STAT3 and IL8 in vitro [6]. However, the TIMER immune cell data does suggest reduced endothelial cell infiltration, although these computational findings will need to be confirmed experimentally.

The comparison of these TCGA analysis results with correlation data from high- and low-*WFDC2* expressing HGSOC cell lines revealed some interesting findings. We observed an agreement in GO analysis of low-*WFDC2* associated genes in the categories of extracellular matrix, cell-substrate adhesion, and angiogenesis, confirming an important role for HE4 in regulating these functions. However, the enriched categories among high-*WFDC2* associated genes were notably different between TCGA and HGSOC cell lines, with TCGA data pointing to metabolic functions

and the DepMap data pointing to peptidase regulation and epithelial cell proliferation. These data do not preclude the regulation of peptidase activity and proliferation in an in vivo context, but suggest there are tumor microenvironmental effects that lead to a strong association of HE4 with OXPHOS metabolism. Despite these differences, we again noted a strong correlation of *WFDC2* with *SLPI* and *PI3* (elafin) in the cell lines, confirming the association of these genes in ovarian cancer. In summary, we have identified genes differentially expressed between HGSOC with high-versus low-*WFDC2* expression in human ovarian tumors and cell lines, revealing novel insights but also raising new questions regarding the pathogenic processes that may be regulated by the HE4 protein in the context of the tumor microenvironment in HGSOC. Future studies will further elaborate on the role of HE4 in extracellular matrix regulation and angiogenesis that we have already reported in vitro [5, 6], as well as studies on the metabolic effects of HE4 in an in vivo context.

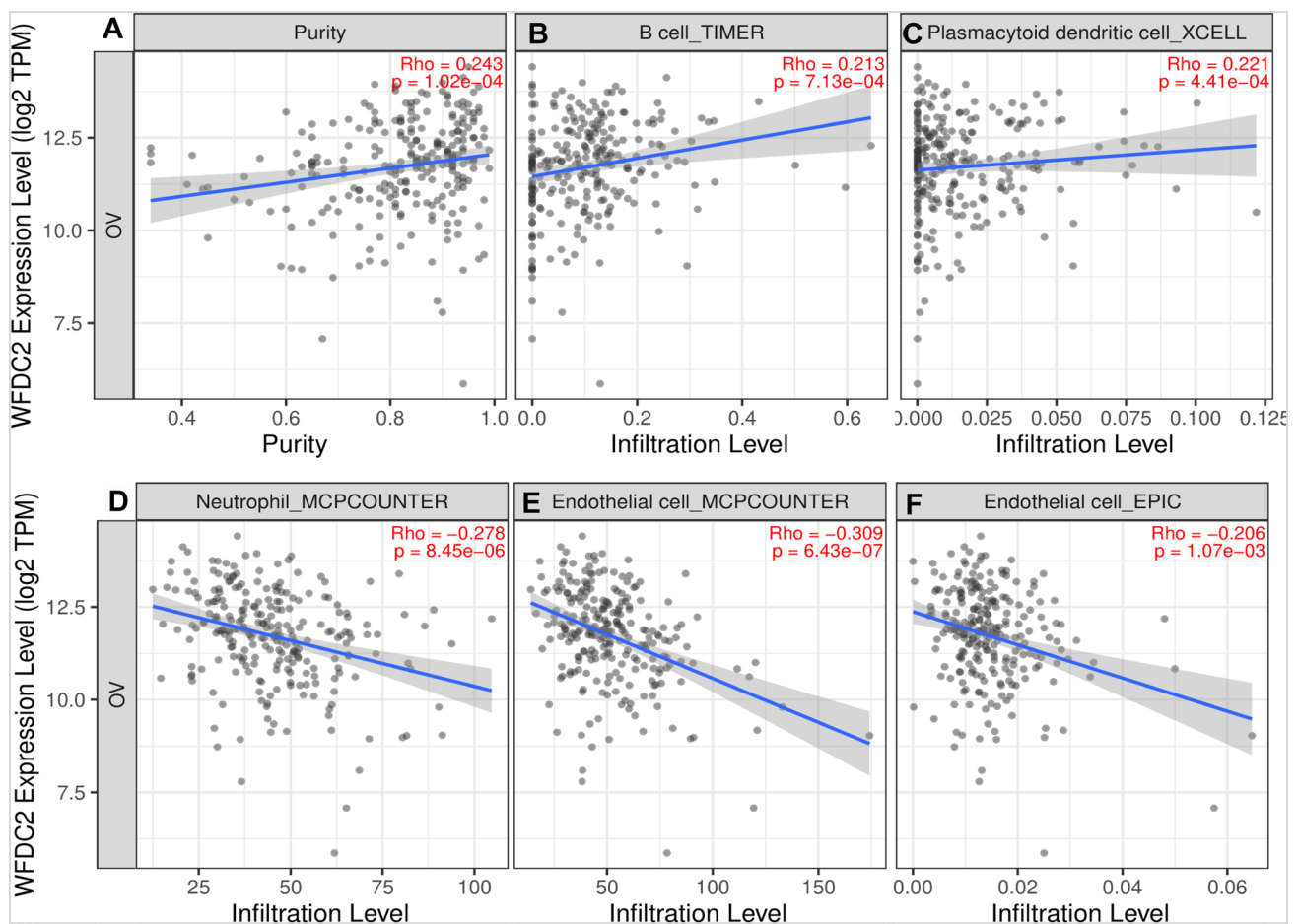


Fig. 5 *WFDC2* levels influence immune cell infiltration. TIMER 2.0 was used to determine the relationship between immune cell infiltration and *WFDC2* transcripts per million (TPM) in TCGA-OV dataset.

Purity correction was performed for all analyses (A). Significantly correlated immune subsets using the indicated algorithm (TIMER, XCELL, MCPCOUNTER, or EPIC) are shown in B–F

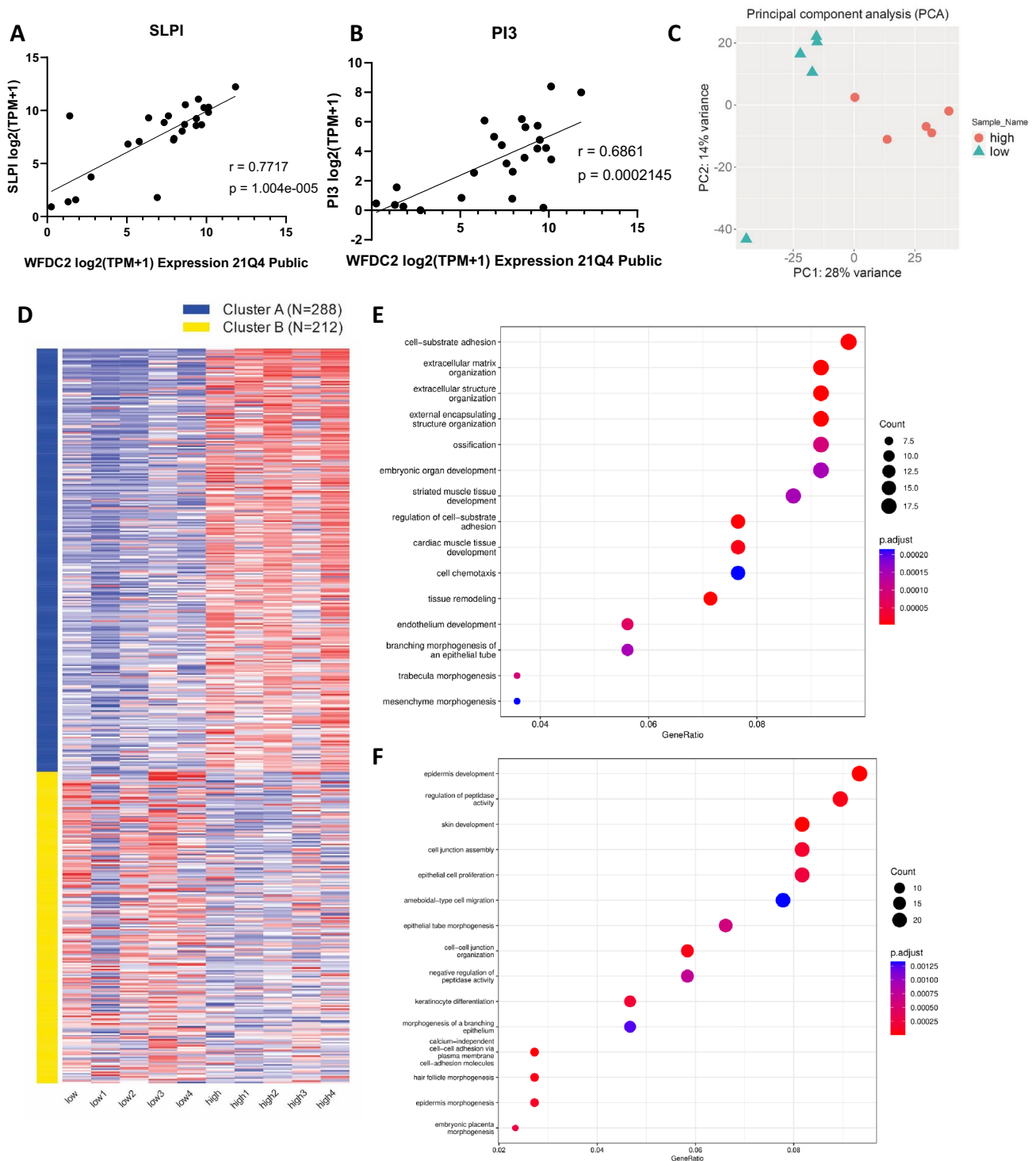


Fig. 6 Comparison of TCGA data with DepMap ovarian cancer cell line data. **A, B** Pearson correlation of *SLPI* and *PI3* (elafin) with *WFDC2* TPMs in DepMap HGSOc cell lines. **C** Principal component analysis (PCA) of the top five low- and high-*WFDC2* expressing HGSOc cell lines. **D** Two k-means clustering analysis of the top five low- and high-*WFDC2* expressing cell lines, using the 500 most variable genes. **E** Gene ontology analysis was performed for Cluster B genes (genes associated with low-*WFDC2* in DepMap HGSOc cell lines), revealing enrichment in categories related to extracellular

matrix and adhesion. Number of genes in each category (“Count”) are indicated by circle size, while adjusted p-value (“p.adjust”) is indicated by color. **F** Gene ontology analysis was performed for k-means cluster A genes (genes associated with high-*WFDC2* in DepMap HGSOc cell lines), revealing enrichment in categories related to epidermis development, proliferation, and peptidase activity. Number of genes in each category (“Count”) are indicated by circle size, while adjusted p-value (“p.adjust”) is indicated by color

Methods

cBioPortal

For all cBioPortal [47, 48] analyses, TCGA ovarian cancer Firehose Legacy dataset was explored. All gene correlations were performed using the “co-expression” feature. Protein and mRNA correlation and mRNA levels according to copy number alterations were performed using the “plots” feature. Mutation count analyses were performed using the “plots” feature and the “mutations count” feature selecting “median”. RNA Seq V2 RSEM or U133 microarray data were used where indicated.

DepMap Portal

Cell line expression data (CCLE Expression Public 21Q4) available in the DepMap portal [49] (<https://depmap.org/portal/>) were downloaded for k-means clustering analysis or analyzed using the “Data Explorer” feature in “Tools”. Data were analyzed for 24 HGSOC cell lines and correlation data were downloaded for *SLPI* and *PI3* with *WFDC2*.

Differential gene expression analysis

The Cancer Genome Atlas (TCGA) ovarian cancer dataset with complete RNA-sequencing results in Fragments Per Kilobase of transcript per Million mapped reads (FPKM) ($n=378$) was obtained using GenomicDataCommons (version 1.12.0) and RStudio (R version 4.0.0) [50, 51]. The R scripts used for plotting and to identify the high- and low-*WFDC2* DEGs in this publication are available to the public: https://github.com/mg859337/WFDC2_TCGA_Analysis. The median FPKM of *WFDC2* was calculated using the FPKM table from TCGA. FPKM values were used to split the samples into *WFDC2*-high and -low groups. Based on this, a metadata file for this dataset was manually created in Excel and saved as a csv. The gene count table from TCGA was used to create a PCA plot by variance-stabilizing transformation (vst) of the dds created using DESeq2 (v1.28.1) and plotted using ggplot2 (v3.3.3) in RStudio (Rv4.0.2) (deseq ref, ggplot ref). DESeq2 differential expression analysis was run with the design of “~*WFDC2*”. Differentially expressed genes (DEGs) were defined as protein-coding genes with log₂ fold-change of ≥ 0.5 or ≤ -0.5 that had a p-adjust value < 0.05 . The volcano plot was created using ggplot2, dplyr (v1.0.2), and ggrepel (v0.8.2) with DEGs plotted in red.

iDEP.94 k-means clustering analysis

HGSOC cell line expression data (CCLE Expression Public 21Q4) available in the DepMap portal [49] (<https://depmap.org/portal/>) was downloaded. The cell lines were stratified according to median transcripts per million (TPM), and then the top five low-*WFDC2* and top five high-*WFDC2* expressing cell lines were determined. Data for these cell lines were uploaded into the iDEP.94 portal [28], and the data were pre-processed, log-transformed, and very low expressing genes were filtered (genes with 0 TPMs in half the samples). A PCA plot was generated, and k-means clustering analysis was performed for two clusters with the 500 most variable genes.

Gene ontology analysis

The R scripts used for gene ontology analysis are available to the public: https://github.com/mg859337/WFDC2_TCGA_Analysis. The DEGs identified using DESeq2 were divided into “high-*WFDC2* DEGs” and “low-*WFDC2* DEGs” based on their log₂ fold-change. The gene lists for the DepMap gene ontology analysis were generated from k-means clustering data. The gene lists were saved into text files and used as input for clusterProfiler (v3.16.1). To use clusterProfiler, loading the packages org.Hs.eg.db (v3.11.4), DOSE (V3.14.0), and ggplot2 (v3.3.3) into RStudio (Rv4.0.2) was also necessary [52–55]. From the gene ontology results of clusterProfiler, dotplots were constructed for each of the DEG lists ordered by “GeneRatio”.

Kaplan–Meier curves

The ovarian cancer plotter at <http://KMplot.com> [56] was used to determine progression-free survival (PFS) and overall survival (OS) in TCGA and Gene Expression Omnibus (GEO) Series cohorts limited to stage III and IV, grade 2 and 3 samples. $N=472$ (PFS) and $n=495$ (OS) for all genes limited to optimally debulked samples except *ROMO1*, for which only 384–387 samples were available for analysis. $n=942$ (PFS) and $n=975$ (OS) for analysis of *WFDC2* in all serous stage III and IV, grade 2 and 3 samples. Top quartile of expression was used to delineate low and high expressing groups for all analyses.

TIMER immune cell subset analysis

TIMER 2.0 for immune cell subsets with purity adjustment was performed for *WFDC2* in TCGA-OV dataset ($n=303$), using the website <http://timer.comp-genomics.org/>. *WFDC2* and *SLPI* pancancer correlation analysis was also performed using TIMER 2.0 “Gene_Corr” feature [57].

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12032-022-01665-4>.

Acknowledgements The results shown here are in whole or part based upon data generated by TCGA Research Network: <https://www.cancer.gov/tcga>. We would like to thank Dr. Joselynn Wallace in the COBRE Center for CBHD at Brown University for assistance in obtaining TCGA data. This research was supported by the Rhode Island Foundation and Swim Across America.

Author contribution JRR designed the study, obtained and organized TCGA data, analyzed data, generated figures, performed the k-means clustering, gene ontology analysis, and figure generation of the DepMap data, and wrote the manuscript. MAG performed the differential gene expression analysis of TCGA data, performed gene ontology analysis, and generated associated figures. NEJ assisted in study design, performed data analysis, and assisted in writing the manuscript. RNF provided conceptual feedback and oversight of the study. MW assisted in writing and editing of the manuscript. All authors edited and approved the final submitted version.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Hellström I, et al. The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. *Cancer Res*. 2003;63:3695–700.
- Moore RG, et al. A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol Oncol*. 2009;112:40–6.
- James NE, et al. Beyond the biomarker: understanding the diverse roles of human epididymis protein 4 in the pathogenesis of epithelial ovarian cancer. *Front Oncol*. 2018;8:124.
- Ribeiro JR, et al. HE4 promotes collateral resistance to cisplatin and paclitaxel in ovarian cancer cells. *J Ovarian Res*. 2016;9:28.
- Ribeiro JR, et al. Human epididymis protein 4 promotes events associated with metastatic ovarian cancer via upregulation of the extracellular matrix. *Front Oncol*. 2017;7:332.
- James NE, et al. The biomarker HE4 (WFDC2) promotes a pro-angiogenic and immunosuppressive tumor microenvironment via regulation of STAT3 target genes. *Sci Rep*. 2020;10:8558.
- McCloskey CW, et al. Metformin abrogates age-associated ovarian fibrosis. *Clin Cancer Res*. 2020;26:632–42.
- Dasari S, et al. Cancer associated fibroblasts: naughty neighbors that drive ovarian cancer progression. *Cancers (Basel)*. 2018;10:406.
- Ma X. The omentum, a niche for premetastatic ovarian cancer. *J Exp Med*. 2020;217:e20192312.
- Bingle L, et al. WFDC2 (HE4): a potential role in the innate immunity of the oral cavity and respiratory tract and the development of adenocarcinomas of the lung. *Respir Res*. 2006;7:61.
- Bingle L, et al. The putative ovarian tumour marker gene HE4 (WFDC2), is expressed in normal tissues and undergoes complex alternative splicing to yield multiple protein isoforms. *Oncogene*. 2002;21:2768–73.
- Chhikara N, et al. Human epididymis protein-4 (HE-4): a novel cross-class protease inhibitor. *PLoS One*. 2012;7:e47672.
- Nagy B, et al. (2016) Serum HE4 is a suitable inflammatory biomarker in cystic fibrosis: YF8. *Clin Chem Lab Med*
- Nagy B, et al. Human epididymis protein 4 (HE4) levels inversely correlate with lung function improvement (delta FEV1) in cystic fibrosis patients receiving ivacaftor treatment. *J Cyst Fibros*. 2019;18:271–7.
- LeBleu VS, et al. Identification of human epididymis protein-4 as a fibroblast-derived mediator of fibrosis. *Nat Med*. 2013;19:227–31.
- Hou Y, et al. Clinical significance of serum human epididymis protein 4 in liver fibrosis: an experimental study. *Medicine (Baltimore)*. 2020;99:e23428.
- Wan J, et al. Elevated serum concentrations of HE4 as a novel biomarker of disease severity and renal fibrosis in kidney disease. *Oncotarget*. 2016;7:67748–59.
- Liu J, et al. Multi-omics analysis of tumor mutational burden combined with prognostic assessment in epithelial ovarian cancer based on TCGA database. *Int J Med Sci*. 2020;17:3200–13.
- Chen W-T, et al. HE4 as a serum biomarker for ROMA prediction and prognosis of epithelial ovarian cancer. *Asian Pac J Cancer Prev*. 2014;15(1):101–105.
- Hoskins E, et al. Paracrine SLPI secretion upregulates MMP-9 transcription and secretion in ovarian cancer cells. *Gynecol Oncol*. 2011;122:656–62.
- Devoogdt N, et al. Overexpression of protease inhibitor-dead secretory leukocyte protease inhibitor causes more aggressive ovarian cancer in vitro and in vivo. *Cancer Sci*. 2009;100:434–40.
- Simpkins FA, et al. The alarm anti-protease, secretory leukocyte protease inhibitor, is a proliferation and survival factor for ovarian cancer cells. *Carcinogenesis*. 2008;29:466–72.
- Labidi-Galy SI, et al. Elafin drives poor outcome in high-grade serous ovarian cancers and basal-like breast tumors. *Oncogene*. 2015;34:373–83.
- Clauss A, et al. Overexpression of elafin in ovarian carcinoma is driven by genomic gains and activation of the nuclear factor kappaB pathway and is associated with poor overall survival. *Neoplasia*. 2010;12:161–72.
- Su B, et al. SSeCKS/Gravin/AKAP12 inhibits cancer cell invasiveness and chemotaxis by suppressing a protein kinase C- Ral/MEK/ERK pathway. *J Biol Chem*. 2010;285:4578–86.
- Zhang L, et al. Hypoxia-induced HE4 in tubular epithelial cells promotes extracellular matrix accumulation and renal fibrosis via NF- κ B. *FASEB J*. 2020;34:2554–67.
- Peng C, et al. Hypoxia-induced upregulation of HE4 is responsible for resistance to radiation therapy of gastric cancer. *Mol Ther Oncolytics*. 2019;12:49–55.
- Ge SX, et al. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics*. 2018;19:534.
- Nayak AP, et al. Oxidative phosphorylation: a target for novel therapeutic strategies against ovarian cancer. *Cancers (Basel)*. 2018;10:337.
- Matassa DS, et al. Oxidative metabolism drives inflammation-induced platinum resistance in human ovarian cancer. *Cell Death Differ*. 2016;23:1542–54.
- Shen L, et al. PGC1 α regulates mitochondrial oxidative phosphorylation involved in cisplatin resistance in ovarian cells via nucleo-mitochondrial transcriptional feedback. *Exp Cell Res*. 2020. <https://doi.org/10.1016/j.yexcr.2020.112369>.
- Dar S, et al. Bioenergetic adaptations in chemoresistant ovarian cancer cells. *Sci Rep*. 2017;7:8760.
- Pastò A, et al. Cancer stem cells from epithelial ovarian cancer patients privilege oxidative phosphorylation, and resist glucose deprivation. *Oncotarget*. 2014;5:4305–19.
- Winterhoff BJ, et al. Single cell sequencing reveals heterogeneity within ovarian cancer epithelium and cancer associated stromal cells. *Gynecol Oncol*. 2017;144:598–606.

35. Schirinzì A, et al. New insights in laboratory testing for COVID-19 patients: looking for the role and predictive value of human epididymis secretory protein 4 (HE4) and the innate immunity of the oral cavity and respiratory tract. *Microorganisms*. 2020;8:1718.
36. James NE, et al. Human epididymis secretory protein 4 (HE4) compromises cytotoxic mononuclear cells via inducing dual specificity phosphatase 6. *Front Pharmacol*. 2019;10:216.
37. James NE, et al. HE4 suppresses the expression of osteopontin in mononuclear cells and compromises their cytotoxicity against ovarian cancer cells. *Clin Exp Immunol*. 2018;193:327–40.
38. Bickel M. The role of interleukin-8 in inflammation and mechanisms of regulation. *J Periodontol*. 1993;64:456–60.
39. Karsdal MA, et al. The good and the bad collagens of fibrosis - their role in signaling and organ function. *Adv Drug Deliv Rev*. 2017;121:43–56.
40. van Waveren C, et al. Oxidative phosphorylation dysfunction modulates expression of extracellular matrix–remodeling genes and invasion. *Carcinogenesis*. 2006;27:409–18.
41. Mah EJ, et al. Collagen density modulates triple-negative breast cancer cell metabolism through adhesion-mediated contractility. *Sci Rep*. 2018;8:17094.
42. Landry DA, et al. The significance of ovarian fibrosis. *Oncotarget*. 2020;11:4366–70.
43. Sato N, et al. Metformin attenuates lung fibrosis development via NOX4 suppression. *Respir Res*. 2016;17:107.
44. Bridges HR, et al. Effects of metformin and other biguanides on oxidative phosphorylation in mitochondria. *Biochem J*. 2014;462:475–87.
45. Rattan R, et al. Metformin suppresses ovarian cancer growth and metastasis with enhancement of cisplatin cytotoxicity in vivo. *Neoplasia*. 2011;13:483–91.
46. Modica TME, et al. The extracellular matrix protein EMILIN1 silences the RAS-ERK pathway via $\alpha 4\beta 1$ integrin and decreases tumor cell growth. *Oncotarget*. 2017;8:27034–46.
47. Cerami E, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2:401–4.
48. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al. Integrative analysis of complex cancer genomics and clinical. *Sci Signal*. 2013. <https://doi.org/10.1126/scisignal.2004088>.
49. Ghandi M, et al. Next-generation characterization of the cancer cell line encyclopedia. *Nature*. 2019;569:503–8.
50. Team, R.C. (2013) R: a language and environment for statistical computing.
51. Morgan, M. and Davis, S. R: a language and environment for statistical computing, *NIH/NCI Genomic Data Commons Access*. [Online]. Available: <https://bioconductor.org/packages/GenomicDataCommons>, <http://github.com/Bioconductor/GenomicDataCommons>. Accessed 15 Jun 2020.
52. Yu G, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16:284–7.
53. Yu G, et al. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics*. 2015;31:608–9.
54. Datta A, et al. On nearest-neighbor Gaussian process models for massive spatial data. *Wiley Interdiscip Rev Comput Stat*. 2016;8:162–71.
55. Carlson, M. (2019) *org.Hs.eg.db: Genome wide annotation for Human*, R package version 3.8.2.
56. Györfy B, et al. Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. *Endocr Relat Cancer*. 2012;19:197–208.
57. Li T, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res*. 2020;48:W509–14.

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