Identification of potential biomarkers and pivotal biological pathways for prostate cancer using bioinformatics analysis methods

Zihao He 1, 2, 3, Xiaolu Duan 1, 2, 3, Guohua Zeng Corresp. 1, 2, 3

1 Department of Urology, Minimally Invasive Surgery Center, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China
2 Guangzhou Institute of Urology, Guangzhou, China
3 Guangdong Key Laboratory of Urology, Guangzhou, China

Corresponding Author: Guohua Zeng
Email address: gzgyzgh@vip.tom.com

Background: Prostate cancer (PCa) is a common urinary malignancy, whose molecular mechanism has not been fully elucidated. We aimed to screen for key genes and biological pathways related to PCa using bioinformatics method. Methods: Differentially expressed genes (DEGs) were filtered out from the GSE103512 dataset and subjected to the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. The protein-protein interactions (PPI) network was constructed, following by the identification of hub genes. The results of former studies were compared with ours. The relative expression levels of hub genes were examined in The Cancer Genome Atlas (TCGA) and Oncomine public databases. The UCSC Xena online tools were used to study whether the expression of hub genes was correlated with the survival of PCa patients from TCGA cohorts. Results: Totally, 252 (186 upregulated and 66 downregulated) DEGs were identified. GO analysis enriched mainly in ‘oxidation-reduction process’ and ‘positive regulation of transcription from RNA polymerase II promoter’; KEGG pathway analysis enriched mostly in ‘metabolic pathways’ and ‘protein digestion and absorption’. KLK3, CDH1, KLK2, FOXA1, and EPCAM were identified as hub genes from the PPI network. CDH1, FOXA1, and EPCAM were validated by other relevant GEO datasets. All hub genes were validated by both TCGA and Oncomine except KLK2. Two additional top DEGs (ABCC4 and SLPI) were found to be associated with the prognosis of PCa patients. Conclusions: This study excavated the key genes and pathways in PCa, which might be biomarkers for diagnosis, prognosis, and potential therapeutic targets.
Identification of Potential Biomarkers and Pivotal Biological Pathways for Prostate Cancer Using Bioinformatics Analysis

Methods

Zihao He¹,²,³, Xiaolu Duan¹,²,³, Guohua Zeng¹,²,³

1. Department of Urology, Minimally Invasive Surgery Center, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China
2. Guangzhou Institute of Urology, Guangzhou, China
3. Guangdong Key Laboratory of Urology, Guangzhou, China

Corresponding author: Guohua Zeng
Address: Kangda Road 1#, Haizhu District, Guangzhou, Guangdong, China, 510230
Email: gzgyzgh@vip.tom.com
Abstract

Background: Prostate cancer (PCa) is a common urinary malignancy, whose molecular mechanism has not been fully elucidated. We aimed to screen for key genes and biological pathways related to PCa using bioinformatics method.

Methods: Differentially expressed genes (DEGs) were filtered out from the GSE103512 dataset and subjected to the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. The protein-protein interactions (PPI) network was constructed, following by the identification of hub genes. The results of former studies were compared with ours. The relative expression levels of hub genes were examined in The Cancer Genome Atlas (TCGA) and Oncomine public databases. The UCSC Xena online tools were used to study whether the expression of hub genes was correlated with the survival of PCa patients from TCGA cohorts.

Results: Totally, 252 (186 upregulated and 66 downregulated) DEGs were identified. GO analysis enriched mainly in ‘oxidation-reduction process’ and ‘positive regulation of transcription from RNA polymerase II promoter’; KEGG pathway analysis enriched mostly in ‘metabolic pathways’ and ‘protein digestion and absorption’. KLK3, CDH1, KLK2, FOXA1, and EPCAM were identified as hub genes from the PPI network. CDH1, FOXA1, and EPCAM were validated by other relevant GEO datasets. All hub genes were validated by both TCGA and Oncomine except KLK2. Two additional top DEGs (ABCC4 and SLPI) were found to be associated with the prognosis of PCa patients.

Conclusions: This study excavated the key genes and pathways in PCa, which might be biomarkers for diagnosis, prognosis, and potential therapeutic targets.
Introduction

According to the Cancer Statistics of 2019, prostate cancer (PCa) has emerged as the second most frequently diagnosed malignancy and is estimated to be the second leading cause of cancer-related mortalities among American males (Siegel & Miller 2019). At present, prostate specific antigen (PSA) has been the most routinely used biomarker for the screening and monitoring of PCa (Gandaglia et al. 2019). However, elevated PSA does not necessarily indicate PCa and often leads to false positive results as well as overdiagnosis since it can also be seen in other benign lesions such as prostatitis or benign prostatic hyperplasia (BPH) (Inahara et al. 2006; Liu et al. 2019b). Therefore, excavating reliable and effective biomarkers and learning hub genes involved in the biological process of PCa is urgently needed.

With the development of high-throughput sequencing technology and bioinformatic analysis methods, the Gene Expression Omnibus (GEO) online public database has been widely utilized to screen out differentially expressed genes (DEGs), to study molecular signals and their relations, and to aid in constructing gene regulatory networks (Clough & Barrett 2016). Up to now, by either analysis of a single dataset or integrated analysis of multiple datasets in GEO, several studies have dug out genes that exert important influence on the occurrence and progression of PCa, such as CDH1 (Fang et al. 2017), CDCA8 (Zhao et al. 2017), RPS21 (Fan et al. 2018), PIK3R1 (He et al. 2018), EPCAM (Lu 2019), LMNB1 (Song et al. 2019b), IGF2 (Tan et al. 2019), and IKZF1 (Tong et al. 2019). However, the key genes detected by the above studies are largely different from each other and had little in common, and such discrepancy could be attributed to the fact that PCa is
considered as a heterogenous disease as a whole (Liu et al. 2019a; Thomas & Pachynski 2018).

As a consequence, further studies in this regard is still required for the exploration and validation of key genes.

In 2018, Brouwer-Visser et al. studied the immune microenvironment in four solid cancer types including PCa and uploaded the corresponding gene expression profile dataset GSE103512 (Brouwer-Visser et al. 2018). For the first time, the present study employed the GSE103512 dataset to screen out the DEGs between 60 PCa and 7 matched normal prostate (NP) tissue, and a total of 252 DEGs were detected. The Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the DEGs were performed, following by the construction of the protein-protein interaction (PPI) network which could uncover the underlying molecular mechanisms involved in the development of PCa. The cytoHubba and MCODE plugins of the Cytoscape software was applied to identify the hub genes and functional modules in the constructed PPI network, respectively. The Cancer Genome Atlas (TCGA) database, the Oncomine database, and the results reported by former studies were utilized for validation of our outcomes.

Materials and methods

Microarray data

The GSE103512 gene expression profile CEL files, which consisted of 60 PCa samples of different TNM staging and Gleason grading and 7 matched NP samples (Brouwer-Visser et al. 2018), were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The gene probe IDs in
the CEL files were converted to their corresponding gene symbols based on the annotation
information of the platform GPL13158 (Affymetrix HT HG-U133+ PM Array Plate), which
contains 54,715 probes.

Data processing and screening of DEGs

The CEL files of GSE103512 were read using the affy package of the R programming language
(Ver. 3.6.0). Data preconditioning (including background adjustment, normalization, and
summarization) were performed using the Robust Multi-array Average (RMA) method. The limma
package of R was adopted to identify the DEGs between PCa and NP samples, with a \(|\log_2 \text{ fold change}| \geq 1\) and a \(P < 0.05\) was deemed to be of statistically significance (Ritchie et al. 2015). The
overall upregulated and downregulated DEGs information were saved for further analyses.

Functional and enrichment analyses of DEGs

The present study adopted the Database for Annotation Visualization and Integrated Discovery
(DAVID; https://david.ncifcrf.gov/) to obtain GO annotation and KEGG pathway enrichment
information of the DEGs identified previously (Dennis et al. 2003). Results with \(P < 0.05\) with
count number \(\geq 2\) were regarded as statistically significant.

PPI network construction and modules analyses

PPI analysis serves as an entry point for better interpretation of relationships between different
proteins on a genome-wide scale, and might be helpful to provide novel insights into protein
function explanation (Stelzl et al. 2005). PPI relationships of DEGs was analyzed and the
corresponding PPI network was constructed by the Search Tool for the Retrieval of Interacting
Genes (STRING) database online tool (http://string-db.org/) (von Mering et al. 2003). All DEGs
were subsequently submitted to the Cytoscape software for network visualization and hub gene identification, which was realized by the cytoHubba plugin. In addition, the Molecular Complex Detection (MCODE) plugin of Cytoscape was applied to identify potential functional modules in the PPI network with default parameters (degree cutoff $\geq 2$, node score cutoff $\geq 2$, K-core $\geq 2$, and max depth = 100) (Bandettini et al. 2012).

Validation of hub genes

First, we performed a literature search regarding gene expression profiles of PCa in Pubmed and then extracted hub genes reported by the eligible studies, which was used to compare with those identified in the present study. Next, we examined the relative mRNA expression levels of the identified hub genes using the Gene Expression Profiling Interactive Analysis (GEPIA) and Oncomine online tools for further validation. The GEPIA (http://gepia.cancer-pku.cn/) provides comprehensive online services based on TCGA database, and we used the differential expression analysis and patient survival analysis functions in this case (Tang et al. 2017). The Oncomine online database (https://www.oncomine.org/) was applied to examine the mRNA expression of hub genes in both multiple cancers and their corresponding normal tissues (Rhodes et al. 2004).

Survival analysis of hub genes

Provided by the University of California Santa Cruz (UCSC), the UCSC Xena online tools (https://xenabrowser.net) enables researchers to study functional genomic datasets for correlations between genomic and/or phenotypic variables. In this case we used this tool to study whether the expression of hub genes was correlated with the survival of PCa patients from TCGA cohorts. Patients were categorized into a relatively high and low expression group respectively based on
the median expression value of genes, and their overall survivals were analyzed using the Kaplan-
Meier method with a log-tank test. A P < 0.05 was deemed as statistically significant.

Results

Data processing and screening of DEGs

The PCa microarray expression dataset GSE103512 contained information of mRNA expression
of 60 PC samples and 7 matched NP samples. A total of 19,918 official gene symbols were
discerned and the gene expression matrix was constructed. The DEGs were filtered using the
limma R package (criteria: adjusted P < 0.05 and $|\log_2$ fold change$| \geq 1$). A total of 252 DEGs were
identified between PCa and NP samples, including 186 upregulated genes and 66 downregulated
genres. The top 10 upregulated and downregulated DEGs based on fold changes are listed in Table
1. The volcano plot of all the genes detected and the cluster heatmap of the 252 DEGs were
collected in Fig. 1.

Functional and pathway enrichment analyses of DEGs

GO function annotation of the identified DEGs was obtained using the DAVID database and its
online analysis tool, which included the following 3 portions: biological process (BP), cell
component (CC), and molecular function (MF). The results were deemed as statistically significant
if P < 0.05, and the top 15 GO terms of the upregulated and downregulated DEGs are compiled in
Table 2. As shown in Fig. 2A, the upregulated genes were mainly enriched in oxidation-reduction
process (ontology: BP), extracellular exosome (ontology: CC), and calcium ion binding (ontology:
MF). As shown in Fig. 2B, the downregulated genes were mainly enriched in positive regulation
of transcription from RNA polymerase II promoter (ontology: BP), extracellular exosome (ontology: CC), and protein binding (ontology: MF).

The ensuing KEGG pathway analyses were also performed using the online analysis tool of DAVID database and the results were collected in Table 3. As shown in Fig. 2C, the upregulated DEGs enriched in 5 pathways: 1) metabolic pathways, 2) fatty acid metabolism, 3) arginine and proline metabolism, 4) PPAR signaling pathway, and 5) fatty acid biosynthesis. As shown in Fig. 2D, the downregulated DEGs enriched in 5 pathways: 1) protein digestion and absorption, 2) aldosterone-regulated sodium reabsorption, 3) carbohydrate digestion and absorption, 4) mineral absorption, and 5) endocrine and other factor-regulated calcium reabsorption. The KEGG pathways bubble plot of all DEGs was presented in Fig. 3.

PPI network construction and modules analyses

The STRING online database was applied to analyze the 252 identified DEGs and to construct a PPI network, which consisted of 181 nodes interacting with each other via 403 edges. The results were downloaded for further analysis by Cytoscape software. According to the descending order of degree value, the top 5 hub genes were subsequently screened, namely Kallikrein-related peptidase 3 (KLK3), cadherin 1 (CDH1), Kallikrein-related peptidase 2 (KLK2), forkhead box A1 (FOXA1), and epithelial cell adhesion molecule (EPCAM), as presented in Table 4.

Furthermore, seven functional cluster modules were filtered from the PPI network using the MCODE plugin of the Cytoscape software, and the top 2 modules (Fig. 4) were selected for further KEGG pathway enrichment analyses using the DAVID database. As shown in Table 5, the results indicated that the genes in Module 1 are mainly enriched in adherens junction and Hippo signaling.
pathway; while the genes in Module 2 are mainly enriched in fatty acid metabolism, fatty acid biosynthesis, PPAR signaling pathway, and fatty acid degradation.

Validation of hub genes expression in multiple databases

As shown in Table 6, the literature search in Pubmed/GEO regarding gene expression profiles of PCa versus NP tissues yielded 10 relevant studies, 5 (Chen et al. 2012; Endo et al. 2009; Fan et al. 2018; Fang et al. 2017; He et al. 2018) of which were based on a single GEO dataset whereas the other 5 studies (Lu 2019; Song et al. 2019b; Tan et al. 2019; Tong et al. 2019; Zhao et al. 2017) were integrated bioinformatic analyses based on multiple GEO datasets. The hub genes reported by the 10 eligible studies were extracted and compared with those identified in the present study.

As marked in bold fonts in Table 6, there were some overlapped findings between our studies and other studies: our identified hub genes FOXA1, CDH1, and EPCAM were seen in 2 (Endo et al. 2009; Tan et al. 2019), 1 (Fang et al. 2017), and 1 (Lu 2019) other studies, respectively.

Next, we further validated our findings using data from TCGA. GEPIA was applied to determine the expression differences of hub genes between PCa and NP tissues. As shown in Fig. 5A, mRNA levels of KLK3, CDH1, FOXA1, and EPCAM were significantly upregulated in PCa samples compared with NP samples; but there was no significant difference in the mRNA level of KLK2 between PCa and NP samples. Besides, we investigate the prognostic values of hub genes using TCGA data and conducted survival analyses with UCSC Xena online tools. As suggested in Fig. 5B, all 5 hub genes had no statistically significant impact on PCa patients’ overall survivals.

Furthermore, we put the top 10 significantly upregulated and downregulated DEGs into survival analysis and found 2 genes had significant impact: the relative higher expression of ABCC4

...
(upregulated) as well as relative lower expression of SLPI (downregulated) were associated with poorer overall survival.

Furthermore, an overview of hub genes expression in multiple types of cancers revealed that all 5 hub genes were remarkably overexpressed in PCa tissues compared with NP tissues according to the Oncomine database (Fig. 6).

Discussion
As one of the leading healthcare concerns worldwide, prostate cancer has become one of the most prevalent malignancies of adult males with an incidence of around 0.01% (Jemal et al. 2010). Although numerous progress has been made to uncover the molecular mechanisms of PCa development and progression, the outcomes remain obscure and the inconsistencies among different published studies are obvious (Barbieri et al. 2017). The main reason for such phenomenon is generally thought to be the complexity and heterogeneity nature of PCa as a whole (Liu et al. 2019a; Thomas & Pachynski 2018). Thus, it’s of vital significance to perform further research in this regard to validate and update the information acquired.

In current study, we first screened a total of 252 (186 upregulated and 66 downregulated) DEGs from the unused GEO dataset GSE103512. The GO/KEGG pathway enrichment analyses were implemented thereafter. Upregulated DEGs were mainly enriched in metabolic pathways, fatty acid metabolism, and PPAR signaling pathway; whereas downregulated DEGs were mostly associated with protein digestion and absorption. Metabolism alteration is a hallmark of cancer. It is well demonstrated that the cancerous cellular metabolisms were urged to adapt themselves to
the high proliferation rate and nutritional requirement so as to constantly support the increased cell division (Chandel 2014; Sancho et al. 2016). For PCa, one of the most noticeable metabolic changes is, as our results indicated, the fatty acid (FA) or lipid metabolism (Deep & Schlaepfer 2016; Yang et al. 2016). The lipid biosynthesis is essential for membrane formation and cell signaling; for instance, the metabolic intermediates of de novo lipogenesis can serve as second messengers and regulate PCa migration and invasion (Ferro et al. 2017). Moreover, the lipid metabolism of PCa is closely related to androgen by androgen receptor (AR) signaling. For examples, AR signaling can elevate the uptake of exogenous lipids by PCa cells (Liu et al. 2010) as well as stimulate adipose tissues to release FA (O'Reilly et al. 2014). Furthermore, the peroxisome-activated receptors (PPARs) family participate greatly in metabolic regulation. For example, PPARγ is a nuclear FA receptor which can interact with androgen receptor (AR) and regulate growth of PCa (Olokpa et al. 2017). The Fatty acid-binding protein 5 (FABP5)- PPARγ signaling pathway promotes the malignant progression of castration-resistant PCa cells (Jing et al. 2000; Kawaguchi et al. 2016). Recently, the FABP5 inhibitor dmrFABP5 were illustrated to inhibit the FABP5- PPARγ pathway to suppress the malignant progression of PCa cells (Al-Jameel et al. 2019). Therefore, targeting the above pathways might aid in developing effective strategy for the treatment of PCa.

Next, we constructed PPI networks with the DEGs and identified the following top 5 hub genes: KLK3, CDH1, KLK2, FOXA1, and EPCAM. All hub genes exhibited upregulated expressions in PCa compared to NP tissues. Both KLK3 and KLK2 belong to the Kallikrein-related peptidases (KLks) family. Expressed and secreted by glandular epithelia of many organs including
skin, mammary gland, prostate, colon, pancreas and brain, KLKs are a sort of serine proteases existed in the body fluids excreted by the above organs such as sweat, milk, and seminal fluid, involving in various physiological functions like electrolyte balance, extracellular matrix remodeling, and prohormone processing (Borgono et al. 2004; Shaw & Diamandis 2007). Aberrant expression of KLKs can be found in several kinds of malignancies including ovarian cancer (Loessner et al. 2018), breast cancer (Figueroa et al. 2018), gastrointestinal cancer (Kontos et al. 2013), and prostate cancer (Mavridis et al. 2014; McDonald & Parsons 2016). Further on, it was suggested that KLKs could lead to proliferation of the epithelium via protease-activated receptors (PARs), which might contribute to PCa development (Ramsay et al. 2008). Several studies have highlighted the association between KLK3 gene polymorphism and susceptibility to PCa (Chen & Xin 2017; Ding et al. 2018; Motamedi et al. 2019), and Zambon et al. successfully combined the KLK3 genetics analysis and free-to-total PSA ratio for PCa diagnosis (Zambon et al. 2012). Similarly, KLK2 was reported to be potential marker for diagnosis because its polymorphism can increase vulnerability to PCa (David et al. 2002; Nam et al. 2006). And more notably, the loss of miR-378 (a regulator of KLK2) in PCa was found to be correlated with aggressive phenotype and short-term relapse events (Avgeris et al. 2014). Both KLK2 and KLK3 were reported to elevate the level of insulin-like growth factor (IGF) by cleaving the IGF-binding proteins, which consequently affects cell survival, mitogenesis, and differentiation. Analogously, KLK3 can cleave the TGFβ-binding proteins and activate TGFβ, resulting in cell proliferation (Borgono et al. 2004).

The encoding product of CDH1 gene, namely E-cadherin, participates in cell-cell adhesion
and the maintenance of cell differentiation as well as normal structure of epithelium (Mayer et al. 1993; Takeichi 1995). Pathologically, inactivation or absence of CDH1 expression can decrease intercellular junctions and thus promote cancer invasion and metastasis, as seen in diffuse gastric cancer (Melo et al. 2017) and lobular breast cancer (McVeigh et al. 2014). Contrary to expectation, our results indicated that CDH1 expression turned out to be significantly upregulated in PCa compared to NP tissues. Consistent with our results, another study using the GSE26910 dataset also found overexpressed CDH1 in PCa samples (Fang et al. 2017). It was hypothesized that overexpression of CDH1 might have impact on oncogenesis since previous research have indicated the intercellular adhesion between cancer and normal cells were enhanced in the late stage of tumor formation (Albelda 1993). Plus, polymorphism of CDH1 was also found to elevate the risk of PCa (Imtiaz et al. 2019; Qiu et al. 2009). Future work is merited to explore the expression changes of CDH1 in different stages of PCa.

Forkhead Box A1 (FOXA1), also known as hepatocyte nuclear factor 3α (HNF3α), is a transcription factor that plays important roles in development as well as cancer formation (Bernardo & Keri 2012). FOXA1 has multiple impact on PCa, including 1) controls the morphogenesis and cell differentiation of prostate (Gao et al. 2005); 2) facilitates AR transactivation which is essential for PCa proliferation and survival (Zhao et al. 2014); 3) inhibits PCa progress towards neuroendocrine prostate cancer (NEPC), whose prognosis is worse (Kim et al. 2017); 4) downregulates TGF-β signaling and thus suppresses castration-resistant prostate cancer (CRPC) progression (Song et al. 2019a). Our findings were validated by another expression profiling study (Endo et al. 2009), and the complexity of functions of FOXA1 still calls for further
research to fully interpret its role in PCa development and progression.

EPCAM encodes epithelial cell adhesion molecule, also known as CD326, is usually high expressed in cancer tissues and participates in intercellular adhesion limitation, cell signaling, migration, proliferation and differentiation (Maetzel et al. 2009). Consistent with our results, EPCAM is shown to be overexpressed in localized and metastatic PCa (Massoner et al. 2014). Other integrated bioinformatic analysis using 4 GEO datasets also validated our findings (Lu 2019).

When it comes to validation of the identified hub genes, 3 of them (CDH1, FOXA1, and EPCAM) were validated by other studies using relevant GEO datasets; 4 genes (except KLK2) were validated by TCGA data; and all 5 genes were validated by the Oncomine data. This implied their potentials as effective and reliable biomarkers for diagnosis and as possible therapeutic targets. However, in survival analysis using TCGA cohorts, all hub genes turned out to have insignificant impact on PCa patients’ overall survival (OS). To add readability, we put the top 10 upregulated and downregulated DEGs into the survival analysis, outputting 2 significant genes, namely ABCC4 and SLPI. ABCC4 (ATP-cassette binding protein 4), which was significantly upregulated as our results revealed, its relatively higher expression was associated with poorer OS of PCa patients. This negative effect of ABCC4 expression towards PCa might result from that ABCC4 could decrease the efficacy of docetaxel in treating PCa cells (Oprea-Lager et al. 2013). Conversely, the relatively higher level of downregulated DEG SLPI (Stomatin-like protein 1) was related to better OS of PCa patients. Not much studies focused on SLPI. Intriguingly, it is revealed that overexpression of SLPI stabilizes F-box protein Fbw7-γ by inhibiting its degradation, which
is implicated in the degradation of oncogene c-Myc (Zhang et al. 2012). So SLPI might impact negatively on cell proliferation of PCa through the above mechanism. Therefore, these 2 genes might hopefully serve as prognostic indicators for PCa patients.

We also conducted a module analysis on the existing PPI network and selected the top 2 significant modules for the ensuing KEGG pathway enrichment analysis of the genes contained. The results demonstrated that the genes in Module 1 are mainly enriched in adherens junction and Hippo signaling pathway; while the genes in Module 2 are mainly enriched in fatty acid metabolism, fatty acid biosynthesis, PPAR signaling pathway, and fatty acid degradation. Apparently, the modules KEGG analyses were quite conformed to the overall KEGG results. The FA metabolism related pathways, PPAR signaling pathway, and cell adhesion-related hub genes and pathways were already discussed above. It was demonstrated that he activation of Hippo signaling pathway regulates cell growth and proliferation by inhibiting YAP and TAZ transcription co-activators, and its dysregulation impacts significantly on development of cancers including colon, liver, breast, lung, ovary, and prostate (Salem & Hansen 2019; Yu et al. 2015).

Conclusions

By bioinformatic analyses including GO/KEGG enrichment, PPI network, hub gene identification, and module analysis, the current study validated KLK3, CDHI, FOXA1, and EPCAM might potentially serve as effective and reliable molecular biomarkers for diagnosis of PCa; and ABCC4 and SLPI might be utilized as prognostic indicators of PCa. However, further basic and clinical researches are necessary for the verification of the clinical value of our findings.
Acknowledgments

Special thanks to Mr. Fucai Tang for his assistance in the use of R programming Language.

References

Al-Jameel W, Gou X, Jin X, Zhang J, Wei Q, Ai J, Li H, Al-Bayati A, Platt-Higgins A, Pettitt A, Rudland PS, and Ke Y. 2019. Inactivated FABP5 suppresses malignant progression of prostate cancer cells by inhibiting the activation of nuclear fatty acid receptor PPARgamma. *Genes Cancer* 10:80-96. 10.18632/genesandcancer.192

Albelda SM. 1993. Role of integrins and other cell adhesion molecules in tumor progression and metastasis. *Lab Invest* 68:4-17.

Avgeris M, Stravodimos K, and Scorilas A. 2014. Loss of miR-378 in prostate cancer, a common regulator of KLK2 and KLK4, correlates with aggressive disease phenotype and predicts the short-term relapse of the patients. *Biol Chem* 395:1095-1104. 10.1515/hsz-2014-0150

Bandettini WP, Kellman P, Mancini C, Booker OJ, Vasu S, Leung SW, Wilson JR, Shanbhag SM, Chen MY, and Arai AE. 2012. MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. *J Cardiovasc Magn Reson* 14:83. 10.1186/1532-429x-14-83

Barbieri CE, Chinnaiyan AM, Lerner SP, Swanton C, and Rubin MA. 2017. The Emergence of Precision Urologic Oncology: A Collaborative Review on Biomarker-driven Therapeutics. *Eur Urol* 71:237-246. 10.1016/j.eururo.2016.08.024

Barbero GM, and Keri RA. 2012. FOXA1: a transcription factor with parallel functions in development and cancer. *Biosci Rep* 32:113-130. 10.1042/bsr20110046

Borgono CA, Michael IP, and Diamandis EP. 2004. Human tissue kallikreins: physiologic roles and applications in cancer. *Mol Cancer Res* 2:257-280.

Brouwer-Visser J, Cheng WY, Bauer-Mehren A, Maisel D, Lechner K, Andersson E, Dudley JT, and Milletti F. 2018. Regulatory T-cell Genes Drive Altered Immune Microenvironment in Adult Solid Cancers and Allow for Immune Contextual Patient Subtyping. *Cancer Epidemiol Biomarkers Prev* 27:103-112. 10.1158/1055-9965.epi-17-0461

Chandel NS. 2014. Mitochondria as signaling organelles. *BMC Biol* 12:34. 10.1186/1741-7007-12-34

Chen C, and Xin Z. 2017. Single-nucleotide polymorphism rs1058205 of KLK3 is associated with the risk of prostate cancer: A case-control study of Han Chinese men in Northeast China. *Medicine (Baltimore)* 96:e6280. 10.1097/md.0000000000006280

Chen JH, He HC, Jiang FN, Militar J, Ran PY, Qin GQ, Cai C, Chen XB, Zhao J, Mo ZY, Chen YR, Zhu JG, Liu X, and Zhong WD. 2012. Analysis of the specific pathways and networks of prostate cancer for gene expression profiles in the Chinese population. *Med Oncol* 29:1972-1984. 10.1007/s12032-011-0088-5
Clough E, and Barrett T. 2016. The Gene Expression Omnibus Database. *Methods Mol Biol* 1418:93-110. 10.1007/978-1-4939-3578-9_5

David A, Mabjeesh N, Azar I, Biton S, Engel S, Bernstein J, Romano J, Avidor Y, Waks T, Eshhar Z, Langer SZ, Lifschitz-Mercer B, Matzkin H, Rotman G, Toporik A, Savitsky K, and Mintz L. 2002. Unusual alternative splicing within the human kallikrein genes KLK2 and KLK3 gives rise to novel prostate-specific proteins. *J Biol Chem* 277:18084-18090. 10.1074/jbc.M102285200

Deep G, and Schlaepfer IR. 2016. Aberrant Lipid Metabolism Promotes Prostate Cancer: Role in Cell Survival under Hypoxia and Extracellular Vesicles Biogenesis. *Int J Mol Sci* 17. 10.3390/ijms17071061

Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, and Lempicki RA. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4:P3.

Ding WH, Ren KW, Yue C, Zou JG, and Zuo L. 2018. Association between three genetic variants in kallikrein 3 and prostate cancer risk. 38. 10.1042/bsr20181151

Endo T, Uzawa K, Suzuki H, Tanzawa H, and Ichikawa T. 2009. Characteristic gene expression profiles of benign prostatic hyperplasty and prostate cancer. *Int J Oncol* 35:499-509. 10.3892/ijo_00000361

Fan S, Liang Z, Gao Z, Pan Z, Han S, Liu X, Zhao C, Yang W, Pan Z, and Feng W. 2018. Identification of the key genes and pathways in prostate cancer. *OncoLett* 16:6663-6669. 10.3892/ol.2018.9491

Fang E, Zhang X, Wang Q, and Wang D. 2017. Identification of prostate cancer hub genes and therapeutic agents using bioinformatics approach. *Cancer Biomark* 20:553-561. 10.3233/cbm-170826

Ferro M, Terracciano D, Buonerba C, Lucarelli G, Bottero D, Perdona S, Autorino R, Serino A, Cantiello F, Damiano R, Andras I, De Placido S, Di Lorenzo G, Battaglia M, Jereczek-Fossa BA, Mirone V, and De Cobelli O. 2017. The emerging role of obesity, diet and lipid metabolism in prostate cancer. *Future Oncol* 13:285-293. 10.2217/fon-2016-0217

Figueroa CD, Molina L, Bhoola KD, and Ehrenfeld P. 2018. Overview of tissue kallikrein and kallikrein-related peptidases in breast cancer. *Biol Chem* 399:937-957. 10.1515/hsz-2018-0111

Gandaglia G, Albers P, Abrahamsson PA, Briganti A, Catto JWF, Chapple CR, Montorsi F, Mottet N, Roobol MJ, Sonksen J, Wirth M, and van Poppel H. 2019. Structured Population-based Prostate-specific Antigen Screening for Prostate Cancer: The European Association of Urology Position in 2019. *Eur Urol*. 10.1016/j.eururo.2019.04.033

Gao N, Ishii K, Mirosevich J, Kuwajima S, Oppenheimer SR, Roberts RL, Jiang M, Yu X, Shappell SB, Caprioli RM, Stoffel M, Hayward SW, and Matusik RJ. 2005. Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. *Development* 132:3431-3443. 10.1242/dev.01917

He Z, Tang F, Lu Z, Huang Y, Lei H, Li Z, and Zeng G. 2018. Analysis of differentially expressed genes, clinical value and biological pathways in prostate cancer. *Am J Transl Res* 10:1444-1456.

Imtiyaz H, Afroz S, Hossain MA, Bellah SF, Rahman MM, Kadir MS, Sultana R, Mazid MA, and Rahman MM. 2019. Genetic polymorphisms in CDH1 and Exo1 genes elevate the prostate cancer risk in Bangladeshi population. 41:1010428319830837. 10.1177/1010428319830837

Inahara M, Suzuki H, Kojima S, Komiya A, Fukasawa S, Imamoto T, Naya Y, and Ichikawa T. 2006. Improved prostate cancer detection using systematic 14-core biopsy for large prostate glands with normal digital rectal examination findings. *Urology* 68:815-819. 10.1016/j.urology.2006.05.010

Jemal A, Siegel R, Xu J, and Ward E. 2010. Cancer statistics, 2010. *CA Cancer J Clin* 60:277-300. 10.3322/caac.20073
Jing C, Beesley C, Foster CS, Rudland PS, Fujii H, Ono T, Chen H, Smith PH, and Ke Y. 2000. Identification of the messenger RNA for human cutaneous fatty acid-binding protein as a metastasis inducer. Cancer Res 60:2390-2398.

Kawaguchi K, Kinameri A, Suzuki S, Senga S, Ke Y, and Fujii H. 2016. The cancer-promoting gene fatty acid-binding protein 5 (FABP5) is epigenetically regulated during human prostate carcinogenesis. Biochem J 473:449-461. 10.1042/bj20150926

Kim J, Jin H, Zhao JC, Yang YA, Li Y, Yang X, Dong X, and Yu J. 2017. FOXA1 inhibits prostate cancer neuroendocrine differentiation. Oncogene 36:4072-4080. 10.1038/onc.2017.50

Kontos CK, Mavridis K, Talieri M, and Scorilas A. 2013. Kallikrein-related peptidases (KLKs) in gastrointestinal cancer: mechanistic and clinical aspects. Thromb Haemost 110:450-457. 10.1160/th12-11-0791

Liu J, Nie R, Chiarappa JA, Wada K, Tsai J, Burudpandee C, Behl A, Ranganath R, and Antonarakis ES. 2019a. Clinical outcomes associated with pathogenic genomic instability mutations in prostate cancer: a retrospective analysis of US pharmacy and medical claims data. J Med Econ 1:10.1080/13696998.2019.1649267

Liu J, Wang Q, Li M, Zhou MY, Yu YF, and Zhan WW. 2019b. Establishment of two new predictive models for prostate cancer to determine whether to require prostate biopsy when the PSA level is in the diagnostic gray zone (4-10 ng ml(-1)). Asian J Androl. 10.4103/aja.aja_46_19

Liu Y, Zuckier LS, and Ghesani NV. 2010. Dominant uptake of fatty acid over glucose by prostate cells: a potential new diagnostic and therapeutic approach. Anticancer Res 30:369-374.

Loessner D, Goettig P, Preis S, Felber J, Bronner J, Clements JA, Dorn J, and Magdolen V. 2018. Kallikrein-related peptidases represent attractive therapeutic targets for ovarian cancer. Expert Opin Ther Targets 22:745-763. 10.1080/14728222.2018.1512587

Lu W. 2019. Identification of key genes in prostate cancer gene expression profile by bioinformatics. Cancer Cell Int 51:e13169. 10.1186/s12935-019-0753-x

Maetzell D, Denzel S, Mack B, Canis M, Went P, Benk M, Kiecz C, Papior P, Baeuerle PA, Munz M, and Gires O. 2009. Nuclear signalling by tumour-associated antigen EpCAM. Nat Cell Biol 11:162-171. 10.1038/ncb1824

Massoner P, Thomm T, Mack B, Untergasser G, Martowicz A, Bobowski K, Klocker H, Gires O, and Puhr M. 2014. EpCAM is overexpressed in local and metastatic prostate cancer, suppressed by chemotherapy and modulated by MET-associated miRNA-200c/205. Br J Cancer 111:955-964. 10.1038/bjc.2014.366

Mavridis K, Avgeris M, and Scorilas A. 2014. Targeting kallikrein-related peptidases in prostate cancer. Expert Opin Ther Targets 18:365-383. 10.1517/14728222.2014.880693

Mayer B, Johnson JP, Leitl F, Jauch KW, Heiss MM, Schildberg FW, Birchmeier W, and Funke I. 1993. E-cadherin expression in primary and metastatic gastric cancer: down-regulation correlates with cellular dedifferentiation and glandular disintegration. Cancer Res 53:1690-1695.
Functional Impact of CDH1 Missense Mutations in Hereditary Diffuse Gastric Cancer. *Int J Mol Sci* 18. 10.3390/ijms18122687

Motamedi RK, Sarhangi N, Afshari M, Sattari M, Jamaldini SH, Samzadeh M, Mohsen Ziaei SA, Pourmand GR, and Hasanzad M. 2019. Kallikrein-related peptidase 3 common genetic variant and the risk of prostate cancer. 10.1002/jcb.28743

Nam RK, Zhang WW, Klotz LH, Trachtenberg J, Jewett MA, Sweet J, Toi A, Teahan S, Venkateswaran V, Sugar L, Loblaw A, Siminovitch K, and Narod SA. 2006. Variants of the hK2 protein gene (KLK2) are associated with serum hK2 levels and predict the presence of prostate cancer at biopsy. *Clin Cancer Res* 12:6452-6458. 10.1158/1078-0432.ccr-06-1485

O'Reilly MW, House PJ, and Tomlinson JW. 2014. Understanding androgen action in adipose tissue. *J Steroid Biochem Mol Biol* 143:277-284. 10.1016/j.jsbmb.2014.04.008

Olokpa E, Moss PE, and Stewart LV. 2017. Crosstalk between the Androgen Receptor and PPAR Gamma Signaling Pathways in the Prostate. 2017:9456020. 10.1155/2017/9456020

Oprea-Lager DE, Bijnsdorp IV, RJ VANM, AJ VDE, Hoekstra OS, and Geldof AA. 2013. ABCC4 Decreases docetaxel and not cabazitaxel efficacy in prostate cancer cells in vitro. *Anticancer Res* 33:387-391.

Qiu LX, Li RT, Zhang JB, Zhong WZ, Bai JL, Liu BR, Zheng MH, and Qian XP. 2009. The E-cadherin (CDH1)--160 C/A polymorphism and prostate cancer risk: a meta-analysis. *Tumour Biol* 17:244-249. 10.1177/1010428319830837

Ramsay AJ, Reid JC, Adams MN, Samaratunga H, Dong Y, Clements JA, and Hooper JD. 2008. Prostatic trypsin-like kallikrein-related peptidases (KLKs) and other prostate-expressed tryptic proteinases as regulators of signalling via proteinase-activated receptors (PARs). *Biol Chem* 389:653-668. 10.1515/bc.2008.078

Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, and Chinnaiyan AM. 2004. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* 6:1-6. 10.1016/s1476-5586(04)80047-2

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47. 10.1093/nar/gkv007

Salem O, and Hansen CG. 2019. The Hippo Pathway in Prostate Cancer. *Cells* 8. 10.3390/cells8040370

Sancho P, Barneda D, and Heeschen C. 2016. Hallmarks of cancer stem cell metabolism. *Br J Cancer* 114:1305-1312. 10.1038/bjc.2016.152

Shaw JL, and Diamandis EP. 2007. Distribution of 15 human kallikreins in tissues and biological fluids. *Clin Chem* 53:1423-1432. 10.1373/clinchem.2007.088104

Siegel RL, and Miller KD. 2019. Cancer statistics, 2019. 69:7-34. 10.3322/caac.21551

Song B, Park SH, Zhao JC, Fong KW, Li S, Lee Y, Yang YA, Sridhar S, Lu X, Abdulkadir SA, Vessella RL, Morrissey C, Kuzel TM, Catalona W, Yang X, and Yu J. 2019a. Targeting FOXA1-mediated repression of TGF-beta signaling suppresses castration-resistant prostate cancer progression. *J Clin Invest* 129:569-582. 10.1172/jci122367

Song ZY, Chao F, Zhuo Z, Ma Z, Li W, and Chen G. 2019b. Identification of hub genes in prostate cancer using robust rank aggregation and weighted gene co-expression network analysis. *Aging (Albany NY)* 11:4736-4756. 10.18632/aging.102087

Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A,
Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehraech H, and Wanker EE. 2005. A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122:957-968. 10.1016/j.cell.2005.08.029

Takeichi M. 1995. Morphogenetic roles of classic cadherins. *Curr Opin Cell Biol* 7:619-627.

Tan J, Jin X, and Wang K. 2019. Integrated Bioinformatics Analysis of Potential Biomarkers for Prostate Cancer. *Pathol Oncol Res* 25:455-460. 10.1007/s12253-017-0346-8

Tang Z, Li C, Kang B, Gao G, Li C, and Zhang Z. 2017. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* 45:W98-w102. 10.1093/nar/gkx247

Thomas TS, and Pachynski RK. 2018. Treatment of Advanced Prostate Cancer. *Mo Med* 115:156-161.

Tong Y, Song Y, and Deng S. 2019. Combined analysis and validation for DNA methylation and gene expression profiles associated with prostate cancer. *Biomark Med* 19:50. 10.2217/bmm-2019-0141

10.1186/s12935-019-0753-x

von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, and Snel B. 2003. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 31:258-261. 10.1093/nar/gkg034

Yang M, Ayuningtyas A, Kenfield SA, Sesso HD, Campos H, Ma J, Stampfer MJ, and Chavarro JE. 2016. Blood fatty acid patterns are associated with prostate cancer risk in a prospective nested case-control study. *Cancer Causes Control* 27:1153-1161. 10.1007/s10552-016-0794-6

Yu FX, Zhao B, and Guan KL. 2015. Hippo Pathway in Organ Size Control, Tissue Homeostasis, and Cancer. *Cell* 163:811-828. 10.1016/j.cell.2015.10.044

Zambon CF, Prayer-Galetti T, Basso D, Padoan A, Rossi E, Secco S, Peloso M, Fogar P, Navaglia F, Moz S, Zattoni F, and Plebani M. 2012. Effectiveness of the combined evaluation of KLK3 genetics and free-to-total prostate specific antigen ratio for prostate cancer diagnosis. *J Urol* 188:1124-1130. 10.1016/j.juro.2012.06.030

Zhang W, MacDonald EM, and Koepp DM. 2012. The stomatin-like protein SLP-1 and Cdk2 interact with the F-Box protein Fbw7-gamma. *PLoS One* 7:e47736. 10.1371/journal.pone.0047736

Zhao R, Wang Y, Zhang M, Gu X, Wang W, Tan J, Wei X, and Jin N. 2017. Screening of potential therapy targets for prostate cancer using integrated analysis of two gene expression profiles. *Oncol Lett* 14:5361-5369. 10.3892/ol.2017.6879

Zhao Y, Tindall DJ, and Huang H. 2014. Modulation of androgen receptor by FOXA1 and FOXO1 factors in prostate cancer. *Int J Biol Sci* 10:614-619. 10.7150/ijbs.8389
Figure 1

The volcano plot and heatmap of the DEGs.

(A) Volcano plot: The abscissa represents $\log_2 FC$ and the ordinate represents $-\log_{10}(\text{adjusted P-value})$. The blue, purple, grey, red, and yellow dots represent significantly upregulated DEGs, significantly downregulated DEGs, insignificant DEGs, top 10 up/downregulated DEGs, and hub genes between PCa and NP tissues, respectively. (B) Cluster heatmap of the DEGs. DEG, differentially expressed gene; FC, fold change; PCa: prostate cancer; NP, normal prostate.
Figure 2

The GO/KEGG enrichment analyses of the DEGs.

(A) Top 15 enriched GO terms of the upregulated DEGs. (B) Top 15 enriched GO terms of the downregulated DEGs. (C) Top 5 enriched pathways of the upregulated DEGs. (D) Top 5 enriched pathways of the downregulated DEGs. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; BP, biological process; CC, cell component; MF, molecular function.
Figure 3

The KEGG pathways bubble plot of DEGs.
Figure 4

Top 2 modules extracted from the PPI network.

(A) Module 1. (B) Module 2. Circles represent genes, lines represent interactions between gene-encoded proteins and line colors represent evidence of interactions between proteins. PPI, protein-protein interaction.
Figure 5

Validation of hub genes using TCGA data.

Relative expression of hub genes between PCa and NP samples in the GEPIA database (the threshold was set as $|\log_2 \text{FC}| = 1$, $p = 0.05$, *$p < 0.05$): (A) KLK3; (B) CDH1; (C) KLK2; (D) FOXA1; (E) EPCAM. The relationship of hub genes and top DEGs with the overall survival of TCGA PCa cohorts analyzed by the UCSC Xena online tools: (F) KLK3; (G) CDH1; (H) KLK2; (I) FOXA1; (J) EPCAM; (K) ABCC4; (L) SLPI. TCGA, The Cancer Genome Atlas; GEPIA, the Gene Expression Profiling Interactive Analysis database; UCSC: University of California Santa Cruz.
Figure 6

An overview of mRNA levels of hub genes in various types of cancer based on Oncomine data.

The numbers in colored grids shows the counts of datasets with statistically significant mRNA overexpression (red) or low expression (blue) of genes. Grid color was determined by the best gene rank percentile for analysis within the grids. The threshold was set as gene rank percentile = All, p = 0.05, and FC = 2.
### Analysis Type by Cancer

| Cancer Type                  | Cancer vs. Normal | Cancer vs. Normal | Cancer vs. Normal | Cancer vs. Normal | Cancer vs. Normal |
|-----------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Bladder Cancer              | 2                 | 1                 | 1                 | 1                 | 1                 |
| Brain and CNS Cancer        | 6                 | 3                 | 2                 | 7                 | 1                 |
| Breast Cancer               | 5                 | 4                 | 2                 | 2                 | 23                |
| Cervical Cancer             | 1                 | 1                 | 2                 | 1                 | 4                 |
| Colorectal Cancer           | 4                 | 4                 | 4                 | 1                 | 1                 |
| Esophageal Cancer           | 1                 | 1                 | 2                 | 1                 | 10                |
| Gastric Cancer              | 1                 | 2                 | 1                 | 4                 | 5                 |
| Head and Neck Cancer        | 2                 | 3                 | 2                 | 1                 | 8                 |
| Kidney Cancer               | 2                 | 1                 | 10                | 8                 | 13                |
| Leukemia                    | 1                 | 12                | 1                 | 1                 | 1                 |
| Liver Cancer                | 1                 | 1                 | 1                 | 1                 | 1                 |
| Lung Cancer                 | 1                 | 3                 | 6                 | 3                 | 9                 |
| Lymphoma                    | 2                 | 9                 | 4                 | 1                 | 3                 |
| Melanoma                    | 2                 | 3                 | 2                 | 1                 | 1                 |
| Myeloma                     | 1                 | 2                 | 3                 | 1                 | 1                 |
| Other Cancer                | 2                 | 2                 | 13                | 1                 | 1                 |
| Ovarian Cancer              | 1                 | 7                 | 1                 | 1                 | 1                 |
| Pancreatic Cancer           | 4                 | 2                 | 1                 | 2                 | 1                 |
| Prostate Cancer             | 3                 | 1                 | 2                 | 3                 | 4                 |
| Sarcoma                     | 2                 | 7                 | 2                 | 1                 | 8                 |
| **Significant Unique Analyses** | 11            | 36              | 64                | 55              | 20               |
| **Total Unique Analyses**   | 339               | 376              | 321               | 325              | 343              |

Cell color is determined by the best gene rank percentile for the analyses within the cell.

**NOTE:** An analysis may be counted in more than one cancer type.
**Table 1** (on next page)

Top 10 upregulated and downregulated DEGs between PCa and NP tissues.

DEG, differentially expressed gene; PCa, prostate cancer; NP, normal prostate; FC, fold change; adj.P.Val, adjusted P value.
| Gene symbol | Log$_2$FC | adj.P.Val | State       |
|-------------|-----------|-----------|-------------|
| NPY         | -3.362711603 | 0.000789175 | upregulated |
| TARP        | -3.206862002 | 4.85E-09   | upregulated |
| OR51E2      | -2.828448822 | 7.28E-05   | upregulated |
| NKX3-1      | -2.528690546 | 3.58E-12   | upregulated |
| MSMB        | -2.517978536 | 5.11E-06   | upregulated |
| ACPP        | -2.474802744 | 2.29E-09   | upregulated |
| SLC45A3     | -2.46284327  | 2.66E-08   | upregulated |
| C15orf21    | -2.327851135 | 0.003615423| upregulated |
| ABCC4       | -2.261776258 | 1.46E-08   | upregulated |
| LOC100287445| -2.170028348 | 0.000544702| upregulated |
| FOLR1       | 3.005860886  | 8.83E-09   | downregulated |
| KRT13       | 2.958317838  | 2.03E-08   | downregulated |
| SEMG1       | 2.881097184  | 0.001633121| downregulated |
| SEMG2       | 2.444264593  | 0.00246384 | downregulated |
| OLFM4       | 2.410939695  | 0.000494142| downregulated |
| Gene  | Score          | P-value     | Status     |
|-------|----------------|-------------|------------|
| SLPI  | 2.391184301    | 3.98E-05    | downregulated |
| SCGB1A1 | 2.324648206  | 2.40E-05    | downregulated |
| OSR1  | 2.028908547    | 3.00E-15    | downregulated |
| HSPB6 | 1.791949206    | 1.19E-06    | downregulated |
| CFD   | 1.71042558     | 1.42E-09    | downregulated |
Table 2 (on next page)

Top 15 enriched gene ontology terms of the upregulated and downregulated DEGs.

BP, biological process; CC, cell component; MF, molecular function.
| Category | Term                                                | Count | P.Value          | State      |
|----------|-----------------------------------------------------|-------|-----------------|------------|
| BP       | Oxidation-reduction process                         | 17    | 1.10E-04        | upregulated|
| BP       | Proteolysis                                         | 11    | 0.01703296      | upregulated|
| BP       | Long-chain fatty-acyl-CoA biosynthetic process      | 6     | 3.93E-05        | upregulated|
| BP       | Response to organic cyclic compound                 | 4     | 0.01023923      | upregulated|
| BP       | Response to endoplasmic reticulum stress            | 4     | 0.031611372     | upregulated|
| CC       | Extracellular exosome                               | 74    | 2.91E-19        | upregulated|
| CC       | Integral component of membrane                       | 65    | 0.001627446     | upregulated|
| CC       | Plasma membrane                                     | 56    | 7.29E-04        | upregulated|
| CC       | Membrane                                            | 34    | 0.00188457      | upregulated|
| CC       | Extracellular space                                 | 27    | 1.58E-04        | upregulated|
| MF       | Calcium ion binding                                 | 13    | 0.032748281     | upregulated|
| MF       | Oxidoreductase activity                             | 9     | 5.26E-04        | upregulated|
| MF       | Actin binding                                       | 9     | 0.004195153     | upregulated|
| MF       | Serine-type peptidase activity                      | 7     | 2.40E-05        | upregulated|
| MF       | Serine-type endopeptidase activity                  | 7     | 0.030442507     | upregulated|
| Rank | Domain | Description                                           | ID | P-value       | Status       |
|------|--------|-------------------------------------------------------|----|---------------|--------------|
| 1    | BP     | Positive regulation of transcription from RNA polymerase II promoter | 8  | 0.047270726   | downregulated|
| 2    | BP     | Negative regulation of transcription from RNA polymerase II promoter | 7  | 0.034624047   | downregulated|
| 3    | BP     | Negative regulation of cell growth                    | 4  | 0.008047189   | downregulated|
| 4    | BP     | Extracellular matrix organization                      | 4  | 0.028910713   | downregulated|
| 5    | BP     | Collagen fibril organization                           | 3  | 0.007739592   | downregulated|
| 6    | CC     | Extracellular exosome                                  | 23 | 3.41E-05      | downregulated|
| 7    | CC     | Extracellular space                                   | 20 | 2.27E-08      | downregulated|
| 8    | CC     | Extracellular region                                  | 15 | 4.96E-04      | downregulated|
| 9    | CC     | Extracellular matrix                                  | 6  | 0.002673098   | downregulated|
|      | CC     | Blood microparticle                                   | 4  | 0.013152355   | downregulated|
| 10   | MF     | Protein binding                                       | 34 | 0.021626321   | downregulated|
| 11   | MF     | Drug binding                                          | 4  | 0.001392245   | downregulated|
| 12   | MF     | Protein dimerization activity                         | 4  | 0.009398722   | downregulated|
| 13   | MF     | ATPase activity                                       | 4  | 0.016027571   | downregulated|
| 14   | MF     | Sodium:potassium-exchanging ATPase activity           | 2  | 0.028658029   | downregulated|
Table 3 (on next page)

Top 5 enriched pathways of the upregulated and downregulated DEGs.
| Pathway ID | Name                           | Count | P.Value     | Genes                                                                 | State        |
|-----------|--------------------------------|-------|-------------|-----------------------------------------------------------------------|--------------|
| hsa01100  | Metabolic pathways             | 27    | 0.001161421 | SAT1, NAMPT, SORD, GCNT2, GALNT7, GLUD1, PPT1, GGT1, ACSL1, ALDH1A3, ARG2, FASN, AMD1, ACSL3, DHCR24, UAP1, ACLY, ADI1, RDH11, FOLH1, ALOX15B, PLA2G7, PLA2G2A, ABAT, MBOAT2, SMS, DCXR | upregulated  |
| hsa01212  | Fatty acid metabolism          | 5     | 0.002342524 | ACSL1, SCD, FASN, PPT1, ACSL3                                        | upregulated  |
| hsa00330  | Arginine and proline metabolism| 4     | 0.020729538 | SAT1, ARG2, AMD1, SMS                                                 | upregulated  |
| hsa03320  | PPAR signaling pathway         | 4     | 0.043990087 | ACSL1, SCD, DBI, ACSL3                                               | upregulated  |
| hsa00061  | Fatty acid biosynthesis        | 3     | 0.009820063 | ACSL1, FASN, ACSL3                                                  | upregulated  |
| hsa04974  | Protein digestion and absorption| 3     | 0.017282616 | ATP1B1, ATP1A2, COL5A3                                               | downregulated|
| hsa04960  | Aldosterone-regulated sodium reabsorption| 2   | 0.087044909 | ATP1B1, ATP1A2                                                     | downregulated|
| hsa04973  | Carbohydrate digestion and absorption| 2   | 0.093437569 | ATP1B1, ATP1A2                                                      | downregulated|
| hsa04978  | Mineral absorption             | 2     | 0.097676003 | ATP1B1, ATP1A2                                                      | downregulated|
| hsa04961  | Endocrine and other factor-    | 2     | 0.099788246 | ATP1B1, ATP1A2                                                      | downregulated|
regulated calcium reabsorption
Table 4 (on next page)

Top 5 hub genes in the protein-protein interaction (PPI) network.
| Gene symbol | Gene description                        | Degree | Log$_{2}$FC | P Value    |
|-------------|-----------------------------------------|--------|-------------|------------|
| KLK3        | Kallikrein-related peptidase 3          | 26     | 2.019308668 | 2.04E-11   |
| CDH1        | Cadherin 1                              | 22     | 1.21816524  | 5.18E-09   |
| KLK2        | Kallikrein-related peptidase 2          | 21     | 1.95470456  | 5.64E-11   |
| FOXA1       | Forkhead Box A1                         | 20     | 1.922009952 | 1.66E-11   |
| EPCAM       | Epithelial cell adhesion molecule       | 19     | 1.076626501 | 0.000999371|
Pathway enrichment analysis of the genes in the top 2 Modules.
| Module    | Pathway                      | Count | P-value       | Genes            |
|-----------|------------------------------|-------|---------------|------------------|
| Module 1  | Adherens junction            | 2     | 1.0E-2        | CDH1, SNAI2      |
|           | Hippo signaling pathway      | 2     | 2.0E-2        | CDH1, SNAI2      |
| Module 2  | Fatty acid metabolism        | 4     | 1.97E-05      | ACSL1, SCD, FASN, ACSL3 |
|           | Fatty acid biosynthesis      | 3     | 9.79E-05      | ACSL1, FASN, ACSL3 |
|           | PPAR signaling pathway       | 3     | 0.002925067   | ACSL1, SCD, ACSL3 |
|           | Fatty acid degradation       | 2     | 0.047850686   | ACSL1, ACSL3     |
Table 6 (on next page)

Comparisons between the present study and other studies using the GEO datasets.

Bold fonts represent overlapped hub genes among studies.
| Study            | Dataset                   | Sample size | Hub genes                                      |
|------------------|---------------------------|-------------|-----------------------------------------------|
| 2009 Endo et al. | -                         | 3, 1        | TSPAN1, BMPR1B, **FOXA1**, STEAP1, RCAN3, S100P, LYZ, SCGB3A1, IL8, DPT |
| 2012 Chen et al. | -                         | 4, 4        | PPFIA2, PTPRT, PTPRR, PRR16, CHRM2, KRT23, CYP3A4, CYP3A7, DPYS, DUOXA1 |
| 2017 Fang et al. | GSE26910                  | 6, 6        | **CDH1**, BMP2, NKX3-1, PPARG, PRKAR2B         |
| 2017 Zhao et al. | GSE38241, GSE3933         | 47, 37      | CDCA8, CDCA5, UBE2C, TK1                       |
| 2018 Fan et al.  | GSE55945                  | 18, 8       | RPS21, FOXO1, BIRC5, POLR2H, RPL22L1, NPM1     |
| 2018 He et al.   | GSE46602                  | 36, 14      | PIK3R1, BIRC5, ITGB4, RRM2, TOP2A, ANXA1, LPAR1, ITGB8 |
| 2019 Lu et al.   | GSE32448, GSE45016, GSE46602, GSE104749 | 90, 59      | **EPCAM**, TWIST1, CD38, VEGFA                |
| 2019 Song et al. | GSE6919, GSE6956, GSE32448, GSE32571, GSE35988, GSE46602, GSE68555, GSE69223, GSE70768, GSE88808 | 569, 402   | LMNB1, TK1, RACGAP1, ZWINT                      |
| 2019 Tan et al.  | GSE38240, GSE26910        | 14, 10      | IGF2, GATA5, F10, CFI, AGTR1, **FOXA1**, BZRAP1-AS1, KRT8 |
| 2019 Tong et al. | GSE26910, GSE30174, GSE46602, GSE55945, GSE69223 | 140, 53    | IKZF1, PPM1A, FBP1, SMCHD1, ALPL, CASP5, PYHIN1, DAPK1, CASP8 |
| The present study | GSE103512                | 60, 7       | **KLK3**, **CDH1**, **KLK2**, **FOXA1**, **EPCAM** |