Proteomic Characterization of Primary and Metastatic Prostate Cancer Reveals Reduced Proteinase Activity in Aggressive Tumors

Qing Kay Li (qli23@jhmi.edu)
The Johns Hopkins Hospital
Jing Chen
Arrowhead Pharmaceuticals Inc
Yingwei Hu
Johns Hopkins Hospital: Johns Hopkins Medicine
Naseruddin Hoti
Johns Hopkins Hospital: Johns Hopkins Medicine
Stefani N Thomas
University of Minnesota Medical Center Fairview: University of Minnesota Health
Li Chen
National Institutes of Health
Sujayita Roy
Johns Hopkins Hospital: Johns Hopkins Medicine
Alan Meeker
Johns Hopkins Hospital: Johns Hopkins Medicine
Punit Shah
Johns Hopkins Hospital: Johns Hopkins Medicine
Lijun Chen
Johns Hopkins Hospital: Johns Hopkins Medicine
G. Steven Bova
Tampereen Yliopisto
Bai Zhang
Johns Hopkins Hospital: Johns Hopkins Medicine
Hui Zhang
Johns Hopkins Hospital: Johns Hopkins Medicine

Research

Keywords: proteomics, primary and metastatic prostate cancer, aggressive prostate cancer, proteinase DPP4, SWATH-MS

Posted Date: May 19th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-522029/v1

Abstract

Background

Prostate cancer (PCa) is a heterogeneous group of tumors with variable clinical courses. Clinically, it is critical to separate and treat aggressive PCa (AG) from non-aggressive PCa (NAG). Although recent genomic studies have identified a spectrum of molecular abnormalities associated with PCa, it is still challenge to separate AG from NAG. To better understand the functional consequence of PCa progression and the unique features of AG from NAG, we studied proteomic signatures of primary AG, NAG and metastatic PCa.

Methodes

39 PCa and 10 benign prostate controls in discovery cohort and 57 PCa in independent-collected validation cohort were analyzed using data-independent acquisition (DIA) SWATH–MS platform. Proteins with the highest variances (top 500 proteins) were annotated for the pathway enrichment analysis. Functional analysis of differentially expressed proteins in NAG and AG was performed. Data was further validated using validation cohort, as well as by comparison with TCGA mRNA expression and immnunochemistry on PCa tissue microarray (TMA).

Results

4,415 proteins were identified in the tumor and benign control tissues, including 158 up-regulated and 116 down-regulated proteins in AG tumors. A functional analysis of tumor-associated proteins revealed the reduced expression of several proteinases, including dipeptidyl peptidase 4 (DPP4), carboxypeptidase E (CPE) and prostate specific antigen (KLK3) in AG and metastatic PCa. A targeted analysis using SWATH data further identified that the reduced expression of DPP4 was associated with the accumulation of DPP4 substrates in AG tumors, including the reduced ratio of DPP4 cleaved peptide to intact substrate peptide. Findings were further validated using independent-collected cohort, by comparison with TCGA mRNA data, and the immunochemical stains of our tumor microarray (TMA).

Conclusions

Our study is the first large-scale proteomics analysis of PCa tissue using DIA SWATH-MS platform. It not only provides an interrogative proteomic signature of PCa subtypes, but also indicates critical roles of certain proteinases, especially DPP4, in PCa progression. The spectrum map and protein profile generated in the study can be used to investigate potential biological mechanisms involved in the PCa progression as well as for the development of a clinical assay to distinguish aggressive from indolent PCa.

Background

Prostate cancer (PCa) is the most common cancer and the second leading cause of cancer death in men in the United States, with estimated 192,000 new cases and 33,000 deaths in 2020[1]. The clinical

behavior of PCa is highly variable. Majority of PCa presents as a localized and/or slow-growing disease, which can be safely observed by active surveillance without invasive treatments. However, A subset of tumors have an aggressive behavior leading to progression, metastasis and death [2]. Several recent studies have also demonstrated that the incidence rate for localized disease continues to decline, whereas, the incidence rate for advanced-stage disease continues to arise in men aged > 50 years [2–4] following the guideline of US Preventive Services Task Force (USPSTF) [5, 6]. To separate the high-risk aggressive PCa (AG) from low-risk non-aggressive indolent tumors (NAG), multiple risk stratification systems have been developed, including the combination of both clinical and pathological parameters (such as Gleason score/ISUP grade, PSA levels, clinical and pathological staging); however, these tools are still suboptimal for adequately predict disease progression and the outcome[7, 8].

Recently, further risk stratification using molecular features have been developed. The genomic analysis and the Cancer Genome Atlas (TCGA) demonstrate a substantial heterogeneity of PCa, including the spectrum of molecular abnormalities; and these findings are correlated with variable clinical courses of PCa [9–14]. In the TCGA study [13], the comprehensive genomic analysis of 333 primary PCa revealed that a majority of tumors (74%) fell into one of seven subtypes defined by specific gene fusions (*ERG*, *ETV1/4, FLI1*) or mutations (*SPOP*, *FOXA1*, *IDH1*). Furthermore, epigenetic profiles also found that PCa with *IDH1*-mutation had unique methylated phenotype [11–14]. In *SPOP* and *FOXA1* mutant, the androgen receptor (AR) activity varied widely, having the highest levels of AR-induced transcripts. It was also reported that 25% of PCa had alterations in the PI3K or MAPK signaling pathways, and approximately 20% of PCa had abnormalities in DNA repair genes [9–14]. These studies reveal not only genomic heterogeneity among primary PCa, but also identifies potentially actionable molecular aberrations.

The Sequential Window Acquisition of all THeoretical fragment ion spectra (SWATH), also called data-independent acquisition (DIA) of mass spectrometry (MS), has emerged as an unbiased and alternative technology for proteomic analysis of biological samples to overcome certain limitations of conventional data-dependent acquisition (DDA)-based analysis, such as the stochastic nature of precursor ion selection and low sampling efficiency [15–18]. Our previous studies and others indicate that SWATH can create a comprehensive fragmentation map of all detectable precursors for accurate quantification of the given sample by dividing peptide precursor ions into several consecutive windows during fragmentation. Other benefits of the SWATH platform include requiring less quantity of clinical samples, providing sufficient proteome coverage with quantitative consistency, and analytic accuracy [15, 19, 20]. SWATH-MS is a fast, simple and reproducible method for large-scale quantitative proteomic analysis.

To further understand the molecular features of the tumor progression in PCa, we performed a proteomic analysis of primary (including both indolent NAG and AG subtypes) and metastatic PCa using SWATH-MS platform. The purposes of our study are to generate a comprehensive proteomic map of AG, NAG and metastatic PCa by using DIA approach and to identify unique AG-associated proteinases, which may be used for the development of clinical assay to separate AG from NAG PCa.

Methods And Materials

Materials

BCA protein assay kit, Urea, and tris (2-carboxyethyl) phosphine (TCEP) were from Thermo Fisher Scientific (Waltham, MA); sequencing-grade trypsin was from Promega (Madison, WI); C18 columns and Strong Cation Exchange (SCX) columns were from Glygen (Columbia, MD); all other chemicals were from Sigma-Aldrich (St. Louis, MO).

Clinical samples collection and preparation

Tumor samples were collected from radical prostatectomy or transurethral resection with patients' informed consent and in a manner to protect patients' identity. A total of 106 samples were included in the study, including 48 cases of primary and 48 cases of metastatic PCa, and 10 benign prostate controls. Among samples, the discovery cohort included 10 NAG, 9 AG and 20 metastatic PCa; the independent-collected validation cohort included 19 NAG, 10 AG and 28 metastatic PCa. All specimens were snap-frozen, embedded in optimal cutting temperature (OCT) compound and stored at -80°C until use. OCT-embedded frozen tumor tissues were sectioned and enriched using a cryostat microdissection as previously described [16, 21]. Proteins were extracted using 8M urea and digested with trypsin [22, 23]. Digested peptides were thoroughly cleaned with C18 and SCX columns, vacuum dried and resuspended in 0.2% formic acid.

The hematoxylin and eosin (H&E) stained tumor sections were reviewed by pathologists to ensure the representation of tumor area. The study was approval by the Institutional Review Board of Johns Hopkins Medical Institutions.

Data dependent (DDA) LC-MS/MS on 5600^{+} Triple TOF

To build the spectral library, an equal amount of peptides from each tissue group in the discovery cohort were pooled, and then, the pooled samples were separated to 24 fractions with high pH reversed-phase chromatography as previously described [24]. Each fraction was analyzed on a SCIEX 5600^{+} Triple TOF mass spectrometer (SCIEX, Framingham, MA) with an Eksigent ekspertTM nanoLC 400 system in DDA mode. Peptides were loaded onto a 200μm×0.5mm cHiPLC trap column followed by a 75μm×15cm nano cHiPLC column, and separated using a 90-min gradient from 5–35%, buffer A 0.1% (v/v) formic acid in water, buffer B 0.1% (v/v) formic acid in ACN at a flow rate of 300 μL/min. The MS1 spectra was collected in the range 400–1,800 m/z for 250 ms. The 30 most intense precursors with a charge state of 2–5 which exceeded 50 counts per second were selected. The MS2 spectra was collected in the range 100–1,800 m/z for 50 ms. Precursor ions were dynamically excluded from reselection for 6 s. The duty cycle time was ~1.8s.

Protein identification and quantification

MS/MS spectra of 24 fractionated raw data from the 5600^{+} TripleTOF were searched using ProteinPilot 4.5 against the UniProtKB/Swiss-Prot complete human proteome database containing 20,274 entries (version of December, 2013). The database search included static modifications of 57.021 Da for cysteine, dynamic modification of 15.995 Da for oxidation, and dynamic modification of 42.011 Da for acetylation (N-terminus only). Ion libraries were generated using PeakView from ProteinPilot search files for proteins identified at 1% FDR. MS/MS intensities of selected transitions of non-shared peptides were extracted and quantified using PeakView 2.0 with the SWATH all MS/MS 2.0 microapp. Protein identification was counted based on all the peptides passing 1% FDR requirement in individual sample. Protein quantification was the sum of abundances of all corresponding peptides passing 1% FDR requirement in at least one sample.

SWATH-MS measurement

SWATH-MS datasets from 106 tissue samples were acquired using a 5600^{+} Triple TOF mass spectrometer. The chromatographic system and settings were the same as those for DDA LC-MS/MS as described above. In SWATH-MS mode, the instrument was optimized to quadrupole settings for the selection of 25-m/z wide precursor ion selection windows. Using an isolation width of 26 m/z (containing 1 m/z for the window overlap), 34 overlapping windows was constructed covering the precursor mass range of 400–1,250 m/z. SWATH MS2 spectra were collected from 100 to 1,800 m/z. The collision energy (CE) was optimized for each window according to the calculation for a charge 2+ ion centered upon the window with a spread of 15 eV. An accumulation time (dwell time) of 100 ms was used for all fragment-ion scans in high-sensitivity mode. For each SWATH-MS cycle, a survey scan in high-resolution mode was also acquired for 50 ms, resulting in a duty cycle of ~3.5 s.

Bioinformatics analysis of SWATH proteomics data

All samples were normalized to the reference (AG1) based on the assumption that all samples have a median relative expression of 1. The absolute protein abundances were transformed to log2 ratio values. The protein biological coefficient of variation (CV) was calculated based on the CV of the log2 averaged abundances of all samples. The protein technical CVs was calculated based upon replicates of samples in 5 tissue groups (C, NAG, AG, Nmet and M); the maximum value from these groups was used. Proteins with a biological CV less than the technical CV were not used for the further analysis. 3865 out of 4415 proteins were included for the further analysis.

Unsupervised hierarchical clustering was performed using clustermap function of Seaborn Python package (metric='euclidean', and method='ward'). Proteins with the highest variances (top 500 proteins) were included in the clustering and annotated for the pathway enrichment analysis. Functional analysis such as Over-Representation Analysis (ORA) of differentially expressed proteins between sample clusters was performed using WebGestalt (http://www.webgestalt.org/). Protein features were evaluated against a background dataset of 3865 proteins quantified by the SWATH-MS analysis. Principal component

analysis (PCA) was performed by OmicsOne [25] to further investigate the sample distances between the cancer subtypes and control samples.

Differential analysis based on student t-tests were performed on all qualified proteins passed CV filtration (including tumors verse control as well as AG verse NAG). For tumors and control comparison, the t-test p-values were adjusted by Benjamin-Hochberg (BH) correction. The significantly up- and down-regulated proteins were labelled if adjusted p values < 0.01 and absolute fold changes ≥ 2 during the t-tests. For AG and NAG comparison, proteins with ≥1.5 fold changes and the p value < 0.05 were considered to be subtype-associated proteins. TCGA RNA-seq data was downloaded from Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/) for the verification study.

Permutation test was performed to select subtype-associated stable proteins discriminating between AG and NAG primary tumors. In the permutation test, data of protein expression from AG and NAG groups was generated by random sampling (Permutated). The procedure was performed 500 times. Proteins were ranked with p values with the smallest p value as rank 1, based upon student's t-test on log-transformed data. The average rank was plotted for permutated and non-permutated analysis against the number of proteins.

The log2 fold changes of proteins were analyzed by Gene Set Enrichment Analysis (GSEA) function of WebGestalt to find the enriched Gene Ontology (GO) terms of biological process and molecular function databases under 5% FDR.

Immunohistochemistry of PCa tumor tissue microarray

The PCa tissue microarray (TMA) was constructed using surgical resected tumors (n=60 cases), including Gleason score 3 (i.e. 3+3), 4 (i.e. 3+4, 4+3, or 4+4) to 5 (i.e. 5+4 or 4+5) tumors. All tissues were fixed in 10% buffered formalin and embedded in paraffin. The 6 mm core was used for the TMA, including 215 cores of PCa and 111 cores of adjacent tumor-matched benign tissue. The use of human tumor tissue was approved by the Johns Hopkins Institutional Review Board.

Immunohistochemical (IHC) study of DPP4 was performed on TMA using a Ventana Discovery Ultra autostainer (Roche Diagnostics). Briefly, following dewaxing and rehydration on board, epitope retrieval was performed using Ventana Ultra CC1 buffer (catalog# 6414575001, Roche Diagnostics) at 96^{o}C for 48 minutes. Primary antibody, anti-DD4/CD26 (D6D8K, catalog #67138, Cell signaling Inc) at 1:250 dilution was applied at 36^{o}C for 60 minutes. Primary antibodies were detected using an anti-rabbit HQ detection system (catalog# 7017936001 and 7017812001, Roche Diagnostics) followed by Chromomap DAB IHC detection kit (catalog # 5266645001, Roche Diagnostics), counterstaining with Mayer's hematoxylin, dehydration and mounting. The intensity of IHC staining pattern was semi-quantitatively by two researchers QKL (the American Board of Pathology certified pathologist) and NH, using a 4-tier system as: 0 (0%, no staining), 1 (<10%, weak and focally staining), 2 (10-50%, medium and focally staining), or 3 (>50%, strong and diffusely staining) in tumor cells.

Results

Experimental design

Our cohort consisted of 106 fresh frozen prostate tissues, including 48 primary and 48 metastatic PCa and 10 benign prostate tissue controls (C). In primary PCa, 29 cases were indolent NAG with the Gleason score of 6 (follow-up data up to 16 years); and 19 cases were AG with the Gleason score >7 (and died of PCa). In metastatic PCa, 38 cases were treated with androgen deprivation therapy (castration resistant metastases, M) and 10 cases were without androgen deprivation therapy (castration naïve prostate metastases, Nmet), Autopsies were also performed in deceased cases as part of the Project to Eliminate Lethal Prostate Cancer (PELICAN) and the Johns Hopkins Autopsy Study of Lethal Prostate Cancer (JHASPC). All benign prostate controls were collected from healthy transplant donors who died of other conditions rather than PCa.

Proteomic profiles were generated using the SWATH-MS analytic platform. The workflow was demonstrated in the **Figure 1**. We first characterized the protein signatures of PCa in discovery cohort, including 5 different groups (10 C, 10 NAG, 9 AG, 10 Nmet and 10 M). The SWATH data files from individual sample were searched against a customized spectral library constructed from a combined sample pool and analyzed using data dependent LC-MS/MS on 5600^{+} Triple TOF and the Human Proteome Map. We further validated our findings (verification phase) by using a targeted re-examination of SWATH-MS maps using independently collected cohort, including 19 NAG, 10 AG and 28 M.

Global proteomic profile of non-aggressive, aggressive and metastatic PCa

The proteomic profile of PCa and 10 benign controls were performed for the comprehensive proteome characterization. Using the spectra from DDA runs, total of 4,415 proteins were identified and quantified from SWATH maps, with an average protein identification number of 1275±246 in the individual run (**Supplementary Table 1**). We also evaluated data reproducibility through the analysis. In the study, 48 samples were analyzed in duplicate. The reproducibility of the protein measurements was assessed using these replicated runs. A correlation score of 0.89 was achieved among replicates, indicating a satisfactory reproducibility of the SWATH-MS workflow in the quantification of the proteome data.

We investigated the protein differential expression patterns in all samples. The non-supervised hierarchical clustering was performed on top 500 most variant proteins. The heat map was constructed using cancer-associated proteins that the expression values were transformed into Z score at the protein level (**Figure 2A**). Four major protein clusters were identified among PCa subtypes, including NAG, AG, metastases and benign control C. Furthermore, several sample clusters were also identified among individual samples. To further correlate the major protein clusters with individual samples, biological processes (BP) annotation in each major protein cluster was generated based on Over-Representation Analysis (ORA) via WebGestalt. Among the four major protein clusters, we found that the GO terms of muscle contraction (cluster 2 in orange color), transcriptional misregulation in cancer (cluster 3 in blue color), and immune response (cluster 4 in red color) were enriched (**Figure 2A**, **Supplementary Table 2**).

There is no significant GO term enriched in cluster 1. These biological annotations were correlated with individual sample clustering. For example, protein cluster 2 demonstrated differential pattern among primary and metastatic PCa.

The PCA analysis of all samples further illustrated the differences of protein expression in different sample groups (**Figure 2B**). In the analysis, the tumor subtypes demonstrated distinct patterns than benign control. All tumor groups, including the primary (NAG and AG) and metastatic (Nmet and M) subtypes, except one case, could also be clearly separated based on their protein expression patterns.

Further comparison of tumor (NAG, AG, Nmet and M) with controls, we identified 204 up-regulated and 237 down-regulated proteins with significantly differential expression (adjusted p-value <0.01 and absolute fold change ≥ 2) (**Figure 2C, Supplementary Table 2**). Based upon the GSEA analysis, we found differential protein localizations among up- and down-regulated proteins. The top three BP terms enriched in the up-regulated proteins were endoplasmic reticulum, RNA catabolic process, and RNA localization, whereas, the top three BP terms enriched in the down-regulated proteins were muscle system process, extracellular structure organization, and homotypic cell-cell adhesion (**Figure 2D, Supplementary Table 2**).

Functional analysis revealed downregulation of proteinases

We further analyzed the signature of differentially expressed proteins between NAG and AG groups. We identified 158 up-regulated and 116 down-regulated proteins with subtype-associated differential expression (p-value <0.05 and fold change ≥ 1.5 (**Figure 3A**, **Supplementary Table 3**). All the proteins with subtype-associated differential expression were further verified by permutation test as stable proteins discriminating between AG and NAG primary tumors (**Figure 3B, Supplementary Table 3**).

We performed the molecular function (MF) annotation using WebGestalt to categorize biological functions of AG tumor-associated proteins. The serine hydrolases activity was associated with a reduced expression pattern (**Figure 3C, Supplementary Table 3**). The most differentially expressed proteinases were dipeptidyl peptidase 4 (DPP4), prostate specific antigen (KLK3), and chymase (CMA1) (**Figure 3D, 3E** and **3F**). Similar expression patterns were also found in proteinases, carboxypeptidase E (CPE), integrin beta-1 (ITGB1) and aminopeptidase N (ANPEP) (**Figure 3G, 3H,** and **3I**). Our data demonstrated that these proteinases were markedly downregulated in primary AG and the majority of metastatic PCa.

Further verification of proteinases signature using independent cohort and transcriptomic data

We further verified the altered proteinases expression of serine hydrolase in the discovery samples set using an independently collected cohort, containing 29 primary PCa and 28 metastatic tumors (**Figure 4**). We found a similar expression patterns of DPP4, KLK3 and CMA1 in the verification sample set (**Figure 4A-C**). Proteinases CPE, ITGB1 and ANPEP also revealed similar expression patterns (**Figure 4D-F**).

To assess whether the observed decreased expression of proteinases is regulated at the transcriptional level, we interrogated and compared our findings with the TCGA prostate adenocarcinoma RNA-seq data. The RNA-seq data (mRNA z-score) and clinical information from TCGA study was downloaded from

Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/). In the analysis, we only included TCGA PCa cases without history of any other malignancy and no neoadjuvant treatment. Among cases, 114 cases were identified as AG (Gleason score ≥7, nor biochemical recurrence, nor involvement of regional lymph nodes, nor clinical identified metastases); and 43 cases were identified as NAG (Gleason score of 6, nor biochemical recurrence, nor lymph nodes involvement or clinical identified metastases) (**Supplementary Table 4**). The mRNA expressions of DPP4, CPE and KLK3 were markedly down-regulated in AG PCa in the TCGA data set (**Figure 5A-C)**. The comparison of our proteomics data with TCGA data demonstrated that our findings of down-regulated expression of proteinases were in the agreement with the mRNA expression pattern.

Loss of DPP4 correlate with the accumulation of substrates in aggressive PCa

We further analyzed proteolytic product abundance in NAG and AG by re-examination of the SWATH data for substrate levels of DPP4. Two forms of DPP4 related products were identified, including the intact peptide neuropeptide Y (NPY(1-36) containing the sequence of YPSKPDNPGEDAPAEDMAR, and the cleaved peptide of NPY (NPY (3-36) containing the sequence of SKPDNPGEDAPAEDMAR, where 2 amino acids are truncated from the N-terminus of intact NPY by DPP4 enzymatic activity [26, 27]. Both these two peptides contain c-terminal sequence of HYINLITRQR.

With SWATH-MS data from all 106 cases of PCa and benign controls, we quantified 3 peptides including the intact NPY (1-36), cleaved peptide of NPY (3-36) and c-terminal peptide of NPY. A positive correlation between DPP4 protein expression and NPY cleaved peptide (SKPDNPGEDAPAEDMAR/*YP*SKPDNPGEDAPAEDMAR ratio) was found with a Pearson correlation coefficient of 0.45 (**Figure 5D**), indicating that peptidase activity of DPP4 was correlated with the accumulation of its substrate NPY. A lower ratio of NPY cleaved peptide to intact substrate peptide (SKPDNPGEDAPAEDMAR/*YP*SKPDNPGEDAPAEDMAR ratio) was observed in the AG tumors (**Figure 5E**), further confirming a reduced DPP4 activity in aggressive tumors. The reduced cleaved peptide (NPY 3-36) and accumulation of intact substrate peptide (NPY 1-36) were also identified (**Figure 5F**).

Verification of reduced DPP4 expression as a signature of aggressiveness by PCa tissue microarray and immunochemistry

To further evaluate the expression of DPP4 in prostate tissues, IHC stain of PCa TMA was performed (**Figure 6)**. A total of 215 tumor cores and 111 tumor-matched benign tissue cores were constructed in the TMA. In tumor cores, 87 (40.5%) had a Gleason score 3, 52 (24.2%) had a Gleason score 4 and 76 (35.3%) had a Gleason score 5 (the highest Gleason score component in the cases). The staining patterns of DPP4 in tumors were analyzed using a semi-quantitative scoring system (**Figure 6A**). The decreased expression of DPP4 was identified in AG tumors (**Figure 6B**). A stronger staining pattern was found in Gleason score 3 tumors; in contrast, a weakly staining pattern was found in in Gleason score 4 and Gleason 5 tumors. The intensities of DPP4 staining in tumor-matched controls, Gleason 3, Gleason 4 and Gleason 5 tumors were 2.51±0.75, 2.69±0.57, 1.96±0.92 and 1.47±0.92. The DPP4 expression was significantly decreased in tumors with the Gleason ≥4 ($p < 0.05$) (**Figure 6C**).

Discussion

PCa is a heterogeneous group of tumors. The majority of PCa is diagnosed as an indolent tumor with localized disease [3–7], in which the patient can be cured by either conventional surgical therapy and/or safely observed in an active surveillance program without interventional treatments. Approximately 10% of PCa presents as lethal PCa. Recently, great efforts have been made to construct a comprehensive genomic landscape of PCa using large-scale genomic data to better understand the biology of PCa and potential therapeutic strategies. For example, the TCGA Network comprehensively characterized primary PCa and identified novel molecular features of tumors by using seven genomic platforms. However, the separation of aggressive PCa from indolent tumors remains challenging. To gain further knowledge into the proteomic heterogeneity of aggressive PCa and to investigate the molecular taxonomy of the tumor for future diagnostic, prognostic, and therapeutic stratification, we profiled both primary and metastatic PCa, and comprehensively integrated those data to assess the robustness of previously defined aggressive PCa subtypes. More importantly, we also investigated the potential proteomic alterations associated with the tumor progression and potential prognostic indications.

One of the challenges of proteomic analysis of relatively large number of clinical samples is to choose an effective way to provide sufficient proteome coverage with quantitative consistency and analytic accuracy. In this study, we used SWATH-MS platform to identify proteomic profile in PCa. Our results demonstrated that SWATH-MS could be effectively applied for characterizing global proteomes in large-scale clinical cohorts. SWATH is an unbiased methodology that allows peptide precursor ions divided into several consecutive windows during fragmentation resulting in a comprehensive fragmentation map of all detectable precursor ions for accurate quantification of the given sample. A recent study involving 11 institutions worldwide, demonstrated that SWATH-MS is a rapid, simple and reproducible method for large-scale proteomic quantitative analysis [20]. Other benefits of using SWATH-MS based technology include requiring less quantity of clinical samples compared to conventional DDA-based MS approaches and providing sufficient proteome coverage with quantitative consistency and analytic accuracy [15, 19].

The most unique feature of our study is the integrative analysis of a spectrum of PCa, including from primary NAG to AG, and to metastatic tumors. Our study included 48 primary and 48 metastatic PCa as well as 10 normal/benign prostate tissues using the SWATH–MS. In the study, total of 4,415 proteins were identified in PCa, including 158 up-regulated and 116 down-regulated proteins in AG subtype, comparing to NAG samples. A functional analysis demonstrated that the expression of certain proteinases such as dipeptidyl peptidase 4 (DPP4), carboxypeptidase E (CPE) and prostate specific antigen (KLK3), are significantly altered in both primary aggressive PCa and metastatic tumors compared to primary NAG PCa and normal prostate tissues. The functional role of DPP4 is further revealed by the targeted re-examination of SWATH-MS maps, including the accumulation of the DPP4 substrate, neuropeptide Y (NPY), and the reduction of NPY cleaved peptides in tumors.

DPP4 plays critical roles in regulating cellular signaling pathways and biological functions such as cell proliferation and migration [16, 28–30]. Previously published data also suggest that the DPP4 expression

is frequently lost in cancer [31, 32]. Through studying primary PCa and metastatic PCa, we provide the first evidence that decreased DPP4 expression and activity is associated with PCa aggressiveness. DPP4 is an epithelial membrane-bound serine protease, which can target numerous growth factors/cytokines signaling pathways; it also has oncogenic or tumor suppressor properties [31, 32]. Studies have shown that patients treated with DPP4 inhibitor, a commonly used therapy for type 2 diabetes, may accelerate PCa progression following androgen deprivation therapy [33]. A reduced serum DPP4 level was also found in PCa patients with metastatic disease [34]. Our finding of decreased DPP4 levels in aggressive and metastatic PCa is consistent with these previous studies, indicating a selective oncogenic activity to downregulate DPP4 in aggressive tumors.

The strong selective pressure to keep DPP4 levels low would then result in the accumulation of bio-active substrates. To further verify our findings, we examined the levels of enzymatic products of DPP4, including the intact substrate peptide NPY (1–36) and cleaved peptide NPY (3–36). Intact NPY (1–36) is a 36 amino acid neuropeptide that has been shown to be involved in several hormone-regulated cancers including breast, ovarian and prostate cancer [35–37], where it modulates tumor cell proliferation through the activation of the Y1-receptor signaling pathway [37–39]. Interestingly, cleaved NPY (3–36) loses its ability to activate the Y1-receptor, and becomes a selective activator of Y2-receptor [40]. These studies suggested that NPY (1–36) and NPY (3–36) may initiate differential intracellular signaling pathways. The decreased DPP4 activity in aggressive prostate tumors may lead to increased tumor cell proliferation through the accumulation of the substrate peptides NPY (1–36).

Taken together, our findings suggested that loss of DPP4 expression and activity may promote prostate cancer aggressiveness through the regulatory effect of NPY (1–36). The consequence of the decrease expression of DPP4 and subsequent increase in bio-active substrate levels promote tumor cell proliferation and disease progression. It can be served as a signature of AG tumors.

Conclusion

We analyzed PCa tumor tissues, including primary NAG, AG, and metastatic tumors, using DIA SWATH–MS platform. We identified a comprehensive proteomic map containing 4,415 proteins. We also characterized AG-associated proteins, including 158 up-regulated and 116 down-regulated proteins. A functional analysis of tumor-associated proteins revealed the reduced expression of several proteinases, including dipeptidyl peptidase 4 (DPP4), carboxypeptidase E (CPE) and prostate specific antigen (KLK3), particularly in AG and metastatic PCa. The functional role of DPP4 is further revealed by the targeted re-examination of SWATH-MS maps; and findings of accumulation of the DPP4 substrate, neuropeptide Y (NPY), and the reduction of NPY cleaved peptides in AG tumors. The decreased level of DPP4 in AG further validated using independent-collected cohort, as well as by comparison with TCGA mRNA data and the immunochemical stains using our tumor microarray (TMA). Our findings demonstrate that DPP4 play a critical role in PCa progression. It may be served as a clinical biomarker. Additional studies are necessary to determine the clinical significance of these proteinases and their potential diagnostic and therapeutic value.

Abbreviations

PCa: prostate cancer; AG: aggressive prostate cancer; NAG: non-aggressive prostate cancer; C: control; M: metastasis with androgen deprivation therapy; Nmet: metastasis without androgen deprivation therapy; USPSTF: US preventive services task force; ISUP: international society of urological pathology MS: mass spectrometry; TMA: tumor microarray; DIA: data-independent acquisition; SWATH: the sequential window acquisition of all theoretical fragment ion spectra; TCGA: the Cancer Genome Atlas; DDA: data-dependent acquisition; TCEP: tris (2-carboxyethyl) phosphine; SCX: strong cation exchange column; OCT: optimal cutting temperature; H&E: hematoxylin and eosin; ACN: Acetonitrile; FDR: false discovery rate; CE: collision energy; CV: coefficient of variations; PCA: principal component analysis; BH: Benjamin-Hochberg; SD: standard deviation; HPA: Human Proteome Atlas; QC: quality control; GO: Gene Ontology; GSEA: gene set enrichment analysis; FDA: Food and Drug Administration. IHC: immunochemistry; PELICAN: project to eliminate lethal prostate cancer; JHASPC: the Johns Hopkins autopsy study of lethal prostate cancer; ORA: over-representation analysis; DPP4: dipeptidyl peptidase 4; KLK3: prostate specific antigen; CMA1: chymase 1; CPE: carboxypeptidase E; ITGB1: integrin beta-1; ANPEP: aminopeptidase N; NPY: neuropeptide Y.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

This manuscript has been read and approved by all the authors to publish and is not submitted or under consideration for publication elsewhere.

Availability of data and material

The datasets generated and/or analyzed during the current study are available in the supplementary files (supplementary Tables S1 to S4 (.xlsx)).

Competing interests

All authors declare that they have no competing interests.

Funding

This work is supported in part by the National Institutes of Health under grants and contracts of National Cancer Institute, the Early Detection Research Network (EDRN, U01CA152813) and Clinical Proteomics Tumor Analysis Consortium (U24CA160036 and U24CA210985).

Author Contributions

QKL, JC and HZ contributed to organize data, experimental design, and draft the manuscript; JC, LJC, PS contributed to perform experiments; LC, NH, YWH, SNT, PS, BZ contributed to data analysis; GSB contributed to sample collection; NH, AM, SR contributed to IHC staining.

Acknowledgments

Not applicable.

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," *CA: A Cancer Journal for Clinicians,* vol. 70, no. 1, pp. 7-30, 2020, doi: 10.3322/caac.21590.

[2] A. Jemal, J. Ma, R. Siegel, S. Fedewa, O. Brawley, and E. M. Ward, "Prostate Cancer Incidence Rates 2 Years After the US Preventive Services Task Force Recommendations Against Screening," *JAMA Oncology,* vol. 2, no. 12, pp. 1657-1660, 2016, doi: 10.1001/jamaoncol.2016.2667.

[3] J. C. Hu *et al.*, "Increase in Prostate Cancer Distant Metastases at Diagnosis in the United States," (in eng), *JAMA oncology,* vol. 3, no. 5, pp. 705-707, 2017, doi: 10.1001/jamaoncol.2016.5465.

[4] A. Jemal, M. B. Culp, J. Ma, F. Islami, and S. A. Fedewa, "Prostate Cancer Incidence 5 Years After US Preventive Services Task Force Recommendations Against Screening," *JNCI: Journal of the National Cancer Institute,* vol. 113, no. 1, pp. 64-71, 2021, doi: 10.1093/jnci/djaa068.

[5] U. S. P. S. T. Force, "Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement," *JAMA,* vol. 319, no. 18, pp. 1901-1913, 2018, doi: 10.1001/jama.2018.3710.

[6] S. Negoita *et al.*, "Annual Report to the Nation on the Status of Cancer, part II: Recent changes in prostate cancer trends and disease characteristics," *Cancer,* 10.1002/cncr.31549 vol. 124, no. 13, pp. 2801-2814, 2018, doi: 10.1002/cncr.31549.

[7] A. Houston Keisha, J. King, J. Li, and A. Jemal, "Trends in Prostate Cancer Incidence Rates and Prevalence of Prostate Specific Antigen Screening by Socioeconomic Status and Regions in the United States, 2004 to 2013," *Journal of Urology,* vol. 199, no. 3, pp. 676-682, 2018, doi: 10.1016/j.juro.2017.09.103.

[8] Q. K. Li *et al.*, "Serum fucosylated prostate-specific antigen (PSA) improves the differentiation of aggressive from non-aggressive prostate cancers," *Theranostics,* vol. 5, no. 3, pp. 267-276, 2015, doi: 10.7150/thno.10349.

[9] M. S. Lawrence *et al.*, "Discovery and saturation analysis of cancer genes across 21 tumour types," *Nature,* vol. 505, no. 7484, pp. 495-501, 2014, doi: 10.1038/nature12912.

[10] F. Islami, R. L. Siegel, and A. Jemal, "The changing landscape of cancer in the USA — opportunities for advancing prevention and treatment," *Nature Reviews Clinical Oncology,* vol. 17, no. 10, pp. 631-649, 2020, doi: 10.1038/s41571-020-0378-y.

[11] C. S. Grasso *et al.*, "The mutational landscape of lethal castration-resistant prostate cancer," *Nature,* vol. 487, no. 7406, pp. 239-243, 2012, doi: 10.1038/nature11125.

[12] B. S. Taylor *et al.*, "Integrative genomic profiling of human prostate cancer," *Cancer Cell,* vol. 18, no. 1, pp. 11-22, 2010, doi: 10.1016/j.ccr.2010.05.026.

[13] N. Cancer Genome Atlas Research, "The Molecular Taxonomy of Primary Prostate Cancer," *Cell,* vol. 163, no. 4, pp. 1011-1025, 2015, doi: 10.1016/j.cell.2015.10.025.

[14] M. F. Berger *et al.*, "The genomic complexity of primary human prostate cancer," *Nature,* vol. 470, no. 7333, pp. 214-220, 2011, doi: 10.1038/nature09744.

[15] L. C. Gillet *et al.*, "Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis," *Mol Cell Proteomics,* vol. 11, no. 6, pp. O111.016717-O111.016717, 2012, doi: 10.1074/mcp.O111.016717.

[16] Y. Liu *et al.*, "Glycoproteomic analysis of prostate cancer tissues by SWATH mass spectrometry discovers N-acylethanolamine acid amidase and protein tyrosine kinase 7 as signatures for tumor aggressiveness," *Mol Cell Proteomics,* vol. 13, no. 7, pp. 1753-1768, 2014, doi: 10.1074/mcp.M114.038273.

[17] S. N. Thomas, B. Friedrich, M. Schnaubelt, D. W. Chan, H. Zhang, and R. Aebersold, "Orthogonal Proteomic Platforms and Their Implications for the Stable Classification of High-Grade Serous Ovarian Cancer Subtypes," *iScience,* vol. 23, no. 6, p. 101079, 2020, doi: 10.1016/j.isci.2020.101079.

[18] K.-C. Cho *et al.*, "Deep Proteomics Using Two Dimensional Data Independent Acquisition Mass Spectrometry," *Analytical Chemistry,* vol. 92, no. 6, pp. 4217-4225, 2020, doi: 10.1021/acs.analchem.9b04418.

[19] C. Ludwig, L. Gillet, G. Rosenberger, S. Amon, B. C. Collins, and R. Aebersold, "Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial," *Mol Syst Biol,* vol. 14, no. 8, pp. e8126-e8126, 2018, doi: 10.15252/msb.20178126.

[20] B. C. Collins *et al.*, "Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry," *Nature Communications,* vol. 8, no. 1, p. 291, 2017, doi: 10.1038/s41467-017-00249-5.

[21] J. Chen, S. Toghi Eshghi, G. S. Bova, Q. K. Li, X. Li, and H. Zhang, "Epithelium percentage estimation facilitates epithelial quantitative protein measurement in tissue specimens," *Clinical Proteomics,* vol. 10, no. 1, p. 18, 2013, doi: 10.1186/1559-0275-10-18.

[22] J. Chen, J. Xi, Y. Tian, G. S. Bova, and H. Zhang, "Identification, prioritization, and evaluation of glycoproteins for aggressive prostate cancer using quantitative glycoproteomics and antibody-based assays on tissue specimens," *Proteomics,* vol. 13, no. 15, pp. 2268-2277, 2013, doi: 10.1002/pmic.201200541.

[23] Y. Tian, Z. Yao, R. B. S. Roden, and H. Zhang, "Identification of glycoproteins associated with different histological subtypes of ovarian tumors using quantitative glycoproteomics," (in eng), *Proteomics,* vol. 11, no. 24, pp. 4677-4687, 2011, doi: 10.1002/pmic.201000811.

[24] S. Yang *et al.*, "Glycoproteins identified from heart failure and treatment models," *Proteomics,* vol. 15, no. 2-3, pp. 567-579, 2015, doi: 10.1002/pmic.201400151.

[25] Y. Hu, M. Ao, and H. Zhang, "OmicsOne: Associate Omics Data with Phenotypes in One-Click," *bioRxiv,* p. 756544, 2019, doi: 10.1101/756544.

[26] R. Mentlein, "Dipeptidyl-peptidase IV (CD26)-role in the inactivation of regulatory peptides," *Regulatory Peptides,* vol. 85, no. 1, pp. 9-24, 1999, doi: 10.1016/S0167-0115(99)00089-0.

[27] N. Frerker *et al.*, "Neuropeptide Y (NPY) cleaving enzymes: Structural and functional homologues of dipeptidyl peptidase 4," *Peptides,* vol. 28, no. 2, pp. 257-268, 2007, doi: 10.1016/j.peptides.2006.09.027.

[28] R. Sueyoshi, K. M. Woods Ignatoski, M. Okawada, B. Hartmann, J. Holst, and D. H. Teitelbaum, "Stimulation of intestinal growth and function with DPP4 inhibition in a mouse short bowel syndrome model," *American Journal of Physiology-Gastrointestinal and Liver Physiology,* vol. 307, no. 4, pp. G410-G419, 2014, doi: 10.1152/ajpgi.00363.2013.

[29] N. Ervinna *et al.*, "Anagliptin, a DPP-4 Inhibitor, Suppresses Proliferation of Vascular Smooth Muscles and Monocyte Inflammatory Reaction and Attenuates Atherosclerosis in Male apo E-Deficient Mice," *Endocrinology,* vol. 154, no. 3, pp. 1260-1270, 2013, doi: 10.1210/en.2012-1855.

[30] U. V. Wesley, M. McGroarty, and A. Homoyouni, "Dipeptidyl Peptidase Inhibits Malignant Phenotype of Prostate Cancer Cells by Blocking Basic Fibroblast Growth Factor Signaling Pathway," *Cancer Research,* vol. 65, no. 4, p. 1325, 2005, doi: 10.1158/0008-5472.CAN-04-1852.

[31] P. A. Havre, M. Abe, Y. Urasaki, K. Ohnuma, C. Morimoto, and N. H. Dang, "The role of CD26/dipeptidyl peptidase IV in cancer," *Front Biosci,* vol. 13, pp. 1634-45, Jan 1 2008, doi: 10.2741/2787.

[32] B. Pro and N. H. Dang, "CD26/dipeptidyl peptidase IV and its role in cancer," (in eng), *Histol Histopathol,* vol. 19, no. 4, pp. 1345-51, Oct 2004, doi: 10.14670/hh-19.1345.

[33] J. W. Russo *et al.*, "Downregulation of Dipeptidyl Peptidase 4 Accelerates Progression to Castration-Resistant Prostate Cancer," *Cancer research,* vol. 78, no. 22, pp. 6354-6362, 2018, doi: 10.1158/0008-5472.CAN-18-0687.

[34] K. Ueda *et al.*, "Plasma Low-Molecular-Weight Proteome Profiling Identified Neuropeptide-Y as a Prostate Cancer Biomarker Polypeptide," *Journal of Proteome Research,* vol. 12, no. 10, pp. 4497-4506, 2013, doi: 10.1021/pr400547s.

[35] P. J. Medeiros, B. K. Al-Khazraji, N. M. Novielli, L. M. Postovit, A. F. Chambers, and D. N. Jackson, "Neuropeptide Y stimulates proliferation and migration in the 4T1 breast cancer cell line," *International Journal of Cancer,* 10.1002/ijc.26350 vol. 131, no. 2, pp. 276-286, 2012, doi: 10.1002/ijc.26350.

[36] P. J. Medeiros and D. N. Jackson, "Neuropeptide Y Y5-receptor activation on breast cancer cells acts as a paracrine system that stimulates VEGF expression and secretion to promote angiogenesis," *Peptides,* vol. 48, pp. 106-113, 2013, doi: 10.1016/j.peptides.2013.07.029.

[37] M. Ruscica, E. Dozio, M. Motta, and P. Magni, "Relevance of the neuropeptide Y system in the biology of cancer progression," *Curr Top Med Chem,* vol. 7, no. 17, pp. 1682-91, 2007, doi: 10.2174/156802607782341019.

[38] M. Ruscica *et al.*, "Activation of the Y1 Receptor by Neuropeptide Y Regulates the Growth of Prostate Cancer Cells," *Endocrinology,* vol. 147, no. 3, pp. 1466-1473, 2006, doi: 10.1210/en.2005-0925.

[39] P. Magni and M. Motta, "Expression of neuropeptide Y receptors in human prostate cancer cells," *Annals of Oncology,* vol. 12, pp. S27-S29, 2001, doi: 10.1093/annonc/12.suppl_2.S27.

[40] M. Körner and J. C. Reubi, "NPY receptors in human cancer: A review of current knowledge," *Peptides,* vol. 28, no. 2, pp. 419-425, 2007, doi: 10.1016/j.peptides.2006.08.037.

Figures

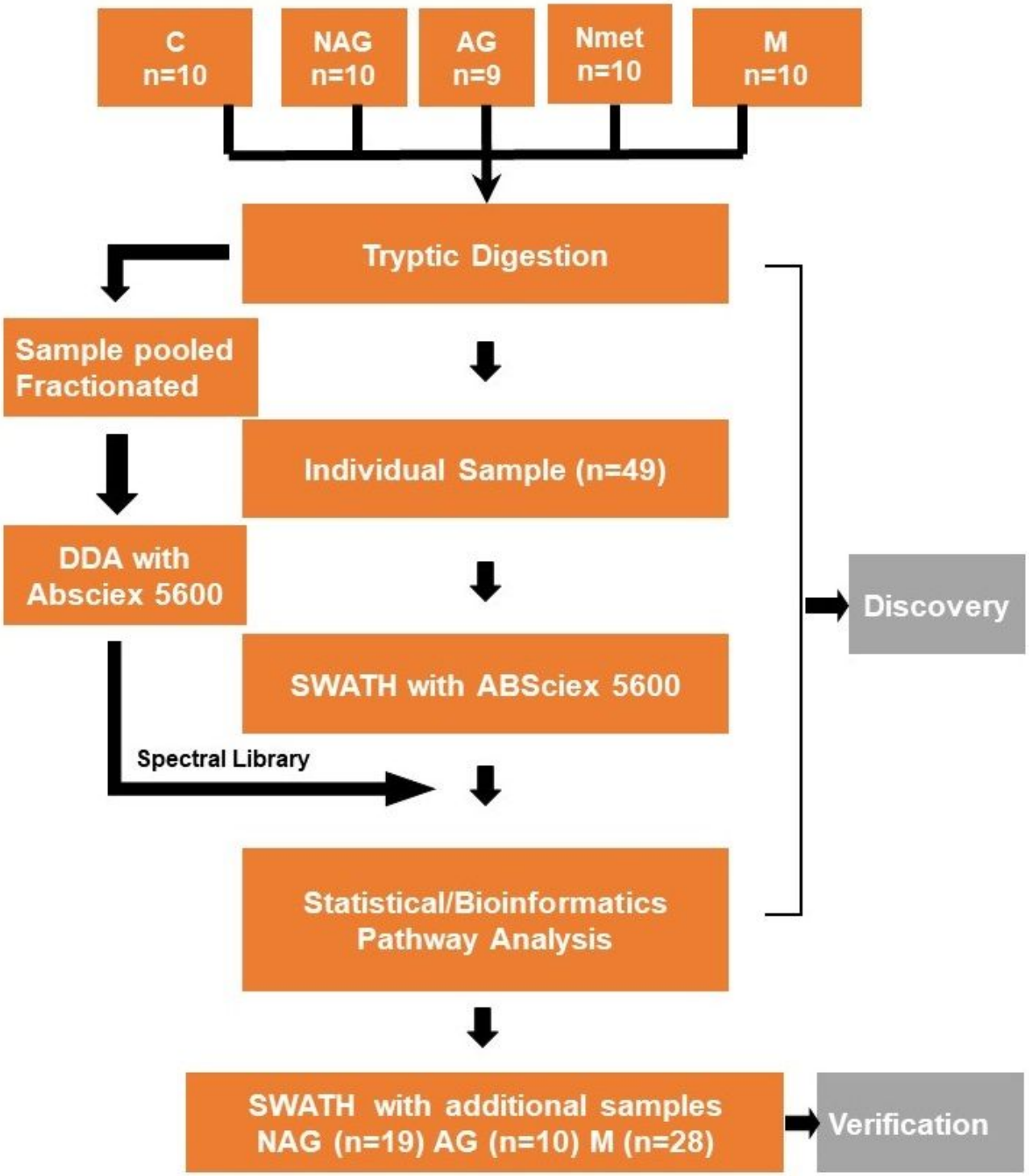


Figure 1

Schematic diagram of the workflow. Frist, we characterized the protein signatures of 49 cases in the discovery cohort, including 19 primary tumors (10 NAG and 9 AG), 20 metastatic tumors (10 Nmet and 10 M) and 10 benign controls (C). Next, we further validated our findings by the targeted re-examination of SWATH-MS maps using independently collected cohort, including 29 primary PCa (19 NAG and 10 AG)

and 28 metastatic tumors (M). C: benign control, NAG: non-aggressive PCa, AG: aggressive PCa, Nmet: metastasis without androgen deprivation therapy; M: metastasis with androgen deprivation therapy.

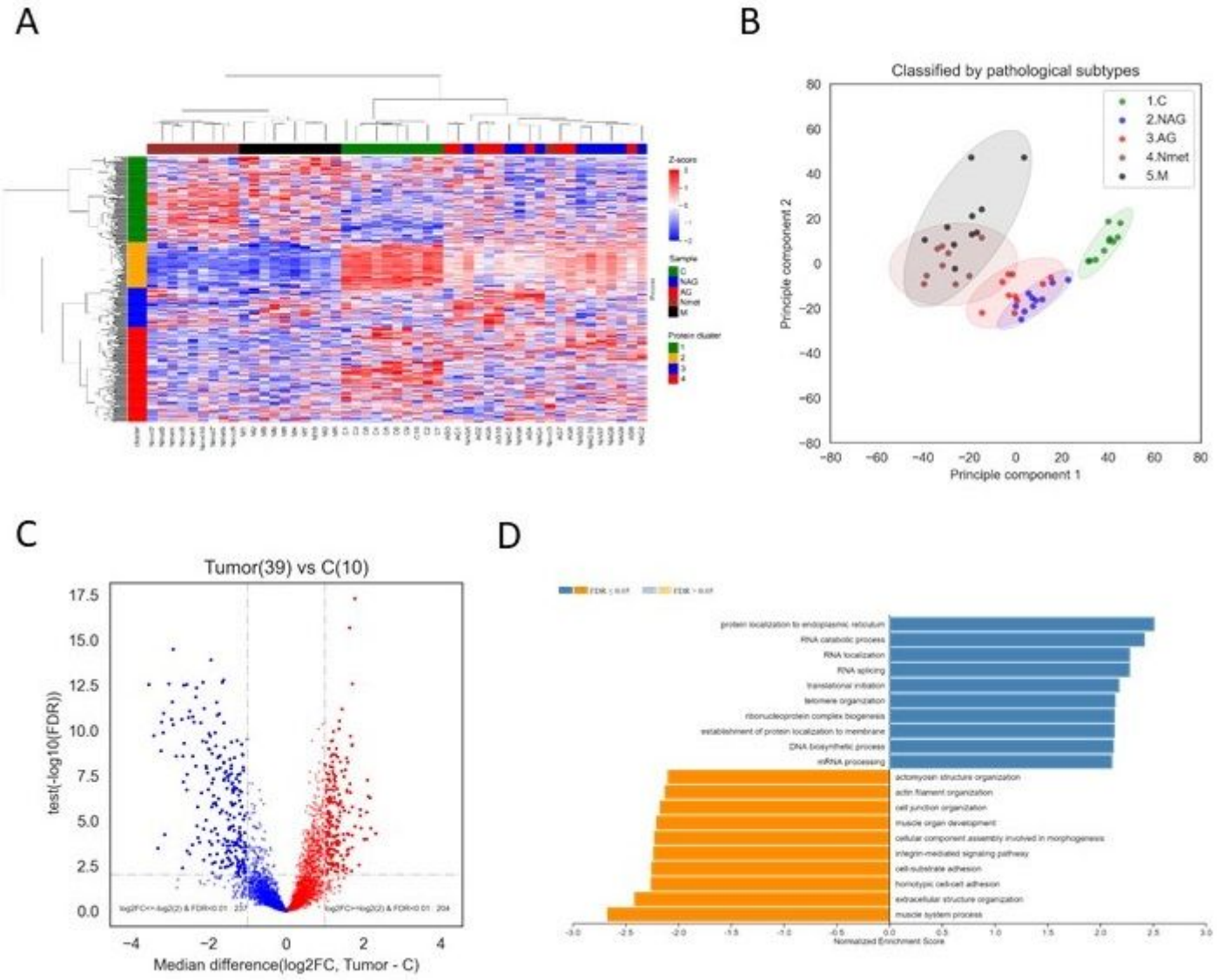


Figure 2

SWATH-MS analysis of PCa. (A) non-supervised hierarchical clustering on top 500 most variant proteins. Sample types were annotated by colors beyond the heat map as: NAG (blue), AG (red), Nmet (brown), M (black) and C (green). Protein clusters were labelled by colors on the left side of the heat map as: cluster 1 (green), cluster 2 (orange), cluster 3 (blue), and cluster 4 (red). (B) PCA analysis. (C) Volcano plot of significantly expressed proteins in tumors (NAG, AG, Nmet, and M groups) and controls C. (D) Gene set enrichment analysis (GSEA) of differentially expressed proteins based on biological process terms in gene ontology database via WebGestalt.

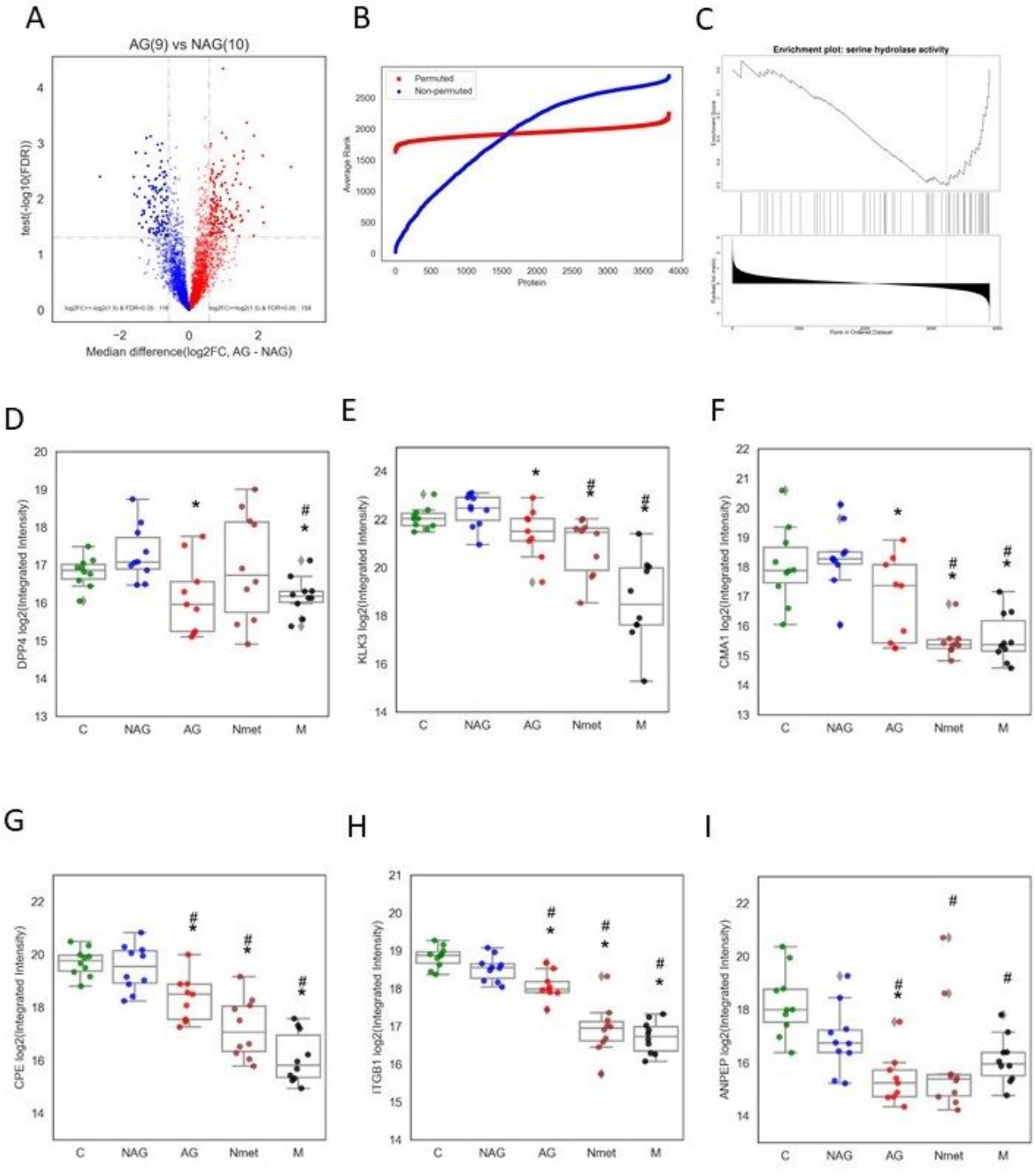


Figure 3

Functional analysis of different subtypes of PCa. (A) Volcano plot of differentially expressed proteins between NAG and AG groups. (B) Permutation test for the selection of stable proteins discriminating between NAG and AG groups. (C) Functional enrichment of proteinases suggested by GSEA analysis via WebGestalt. (D) Expression patterns of DPP4, (E) Expression patterns of KLK3, (F) Expression patterns of

CMA1, (G) Expression patterns of CPE, (H) Expression patterns of ITGB1, and (I) Expression patterns of ANPEP in different tumor subtypes. *p < 0.05 (NAG vs other subtypes). #: p < 0.05 (C vs other subtypes).

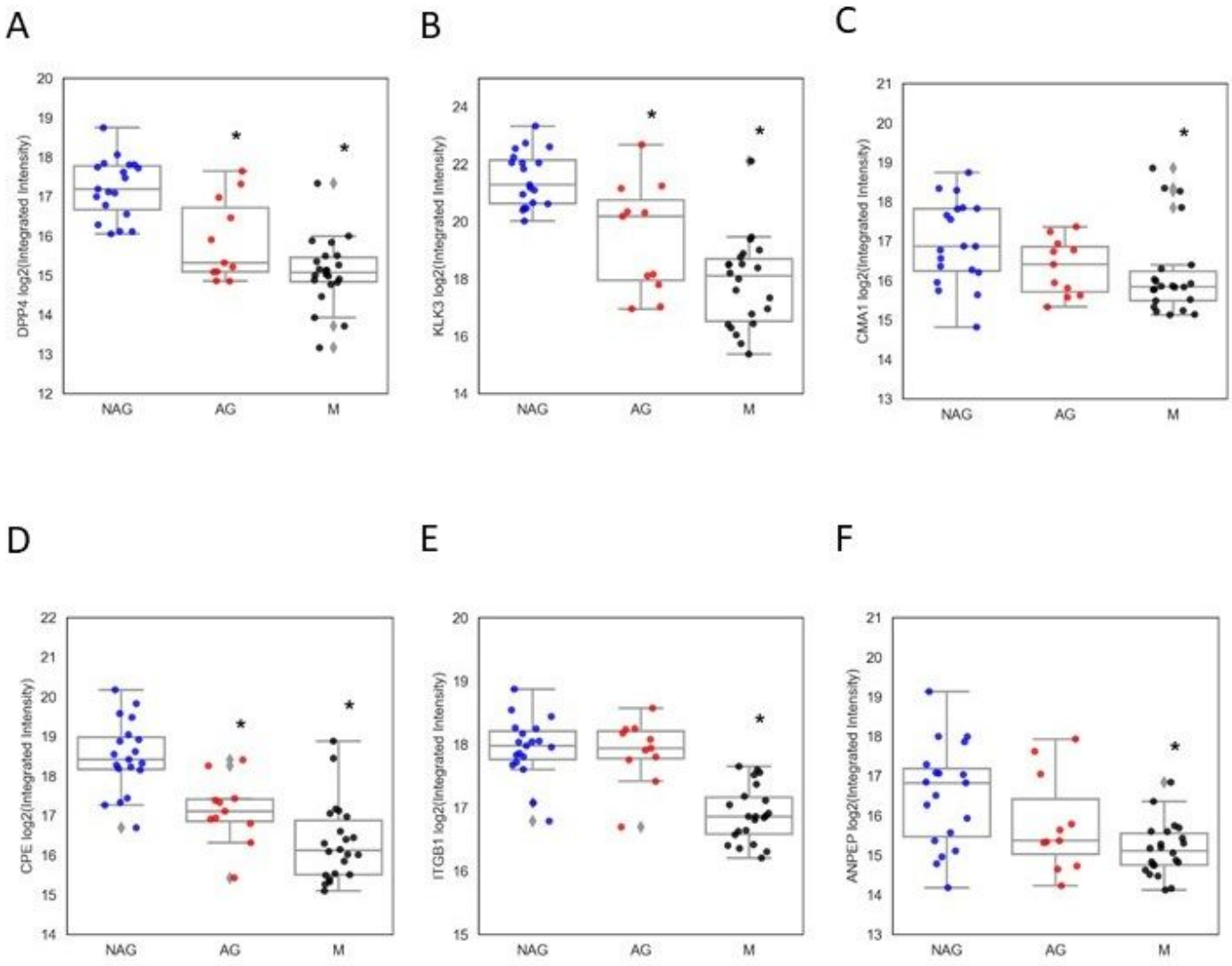


Figure 4

Verification of dysregulation of proteinases using independently collect cohort. Proteinase expression pattern of DPP4 (A), KLK3 (B), CAM1 (C), CPE (D), ITGB1 (E), and ANPEP (F) in different subtypes of tumors, including NAG, AG and metastatic PCa (M). *p < 0.05.

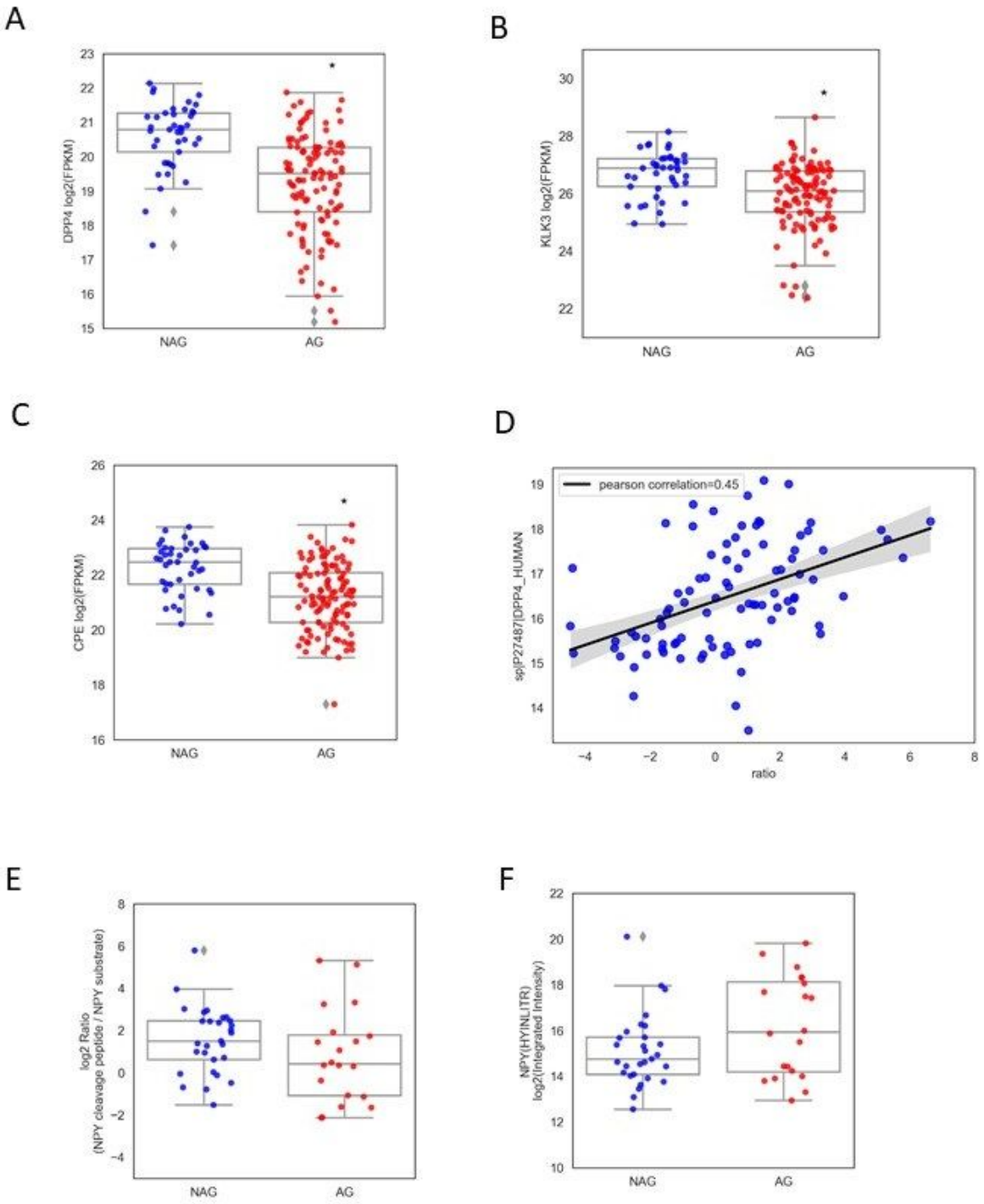
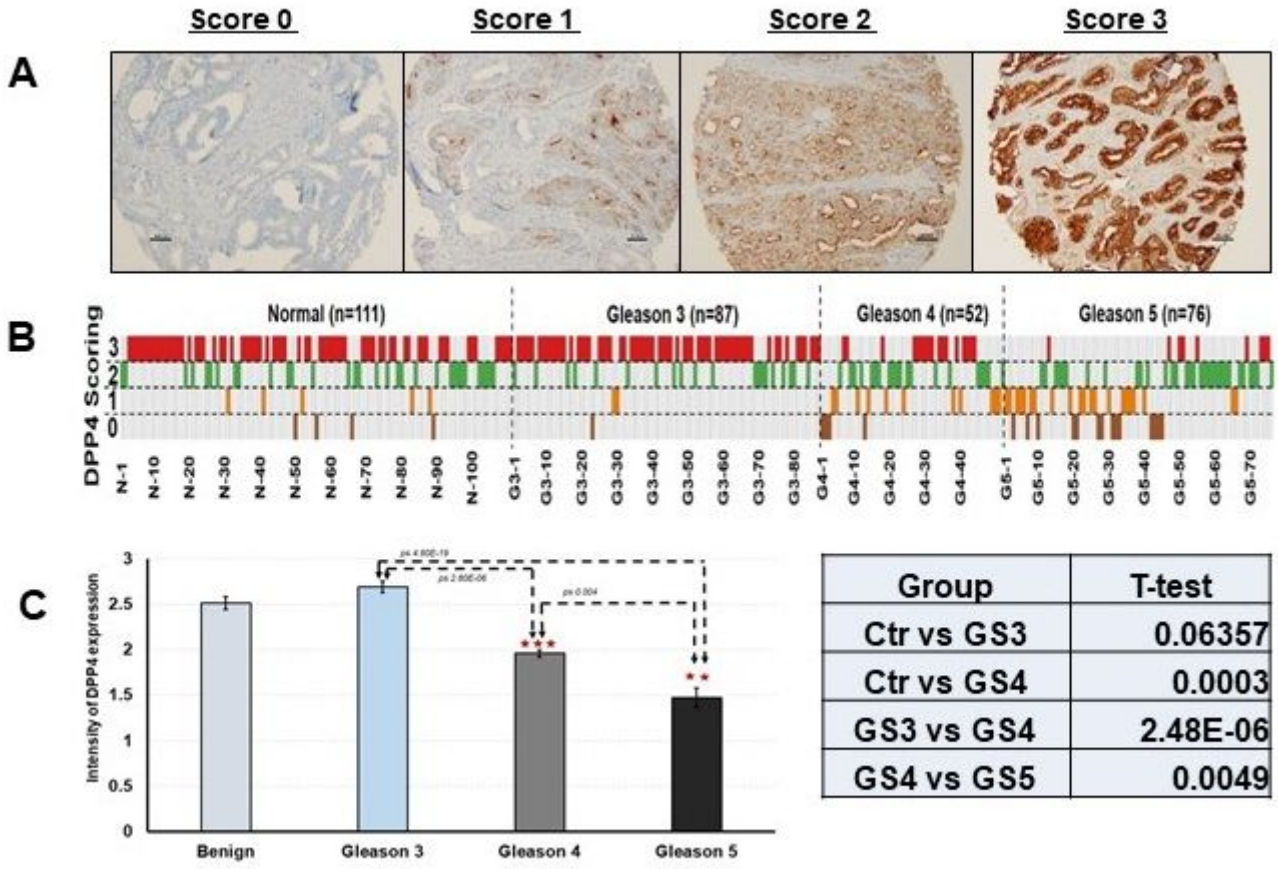


Figure 5

Loss of DPP4 transcription and activity in AG tumors. (A) DPP4 mRNA expression of AG tumors from TCGA date set. (B) KLK3 mRNA expression of AG tumors from TCGA data set. (C) CPE mRNA expression of AG tumors from TCGA data set. (D) Correlation of DPP4 expression with the ratio of NPY cleaved peptide to intact substrate peptide. (E) The ratio of NPY cleaved peptide to intact substrate peptide in AG tumors. (F) Total NPY peptide expressions between NAG and AG tumors. *p < 0.05.



Group	T-test
Ctr vs GS3	0.06357
Ctr vs GS4	0.0003
GS3 vs GS4	2.48E-06
GS4 vs GS5	0.0049

Figure 6

IHC staining of DPP4 expression in PCa tumor tissue microarray (TMA). (A) A semi-quantitative 4-tire scoring system: 0 (0%, no staining), 1 (<10%, weak and focally staining), 2 (10-50%, medium and focally staining), or 3 (>50%, strong and diffusely staining) in tumor cells. (B) Correlation of DPP4 expression with Gleason scores in tumors and tumor-matched benign controls. (C) Comparison of DPP4 expressional pattern in tumors and tumor-matched benign controls. *p < 0.05; **p < 0.01, ***p < 0.001.

Supplementary Files

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

- SupplementaryTable1Proteinidentification.xlsx
- SupplementaryTable2Clusteringandgenesetenrichment.xlsx
- SupplementaryTable3DifferentialanalysisonNAGandAG.xlsx
- SupplementaryTable4TCGAclinicalinformation.xlsx