NOVEL URINARY AND SEROLOGICAL MARKERS OF PROSTATE CANCER USING PROTEOMICS TECHNIQUES: AN IMPORTANT TOOL FOR EARLY CANCER DIAGNOSIS AND TREATMENT MONITORING

HENRY ADEMOLA ADEOLA

Cancer Genomics Group, International Centre for Genetic Engineering and Biotechnology and the Faculty of Health Sciences, University of Cape Town.

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

IN THE DIVISION OF CHEMICAL AND SYSTEMS BIOLOGY, DEPARTMENT OF INTEGRATIVE BIOMEDICAL SCIENCES

Supervisor: A/Professor Luiz. F. Zerbini

Co-Supervisor: Professor Jonathan M. Blackburn

JANUARY 2016

DECLARATION

I Henry Ademola Adeola, hereby declare that: (1) the work contained in this thesis is my own original unaided work apart from the normal guidance of my supervisors and where acknowledgements are duly made by reference; (2) Neither any part nor this whole work has been submitted in the past, is being submitted, or will be submitted in candidature for another degree in this university or any other university; (3) I authorize the university to reproduce both partly and in whole the content thereof for research purposes.

Signed by candidate

Signature Removed

Signature: ……

Henry Ademola Adeola

January, 2016

DEDICATION

To God, my creator, my inspiration and the true fountain of Knowledge.

To my family for all your encouragements and support.

To all my teachers and mentors from the cradle to postgraduate medical training; you all participated in various ways to shape the PhD journey.

ACKNOWLEDGEMENTS

My heart is full of gratitude towards everyone who helped to turn the dream of a PhD degree into reality; many of who are far too many to mention. In no particular order, I would like thank the following individuals who helped in different ways to ensure the completion of this work:

- To my supervisor A/Professor Luiz Fernando C. Zerbini, thank you so much giving me the best guidance during my PhD journey and also offering advice that would last me beyond the four walls of a university. Without your insight and brilliant counsel, completion of this work would be impossible and I certainly have you to thank always for this.
- To Professor Jonathan M Blackburn, who co-supervised this work, I am most grateful to you for initiating this Journey and for helping me identify good opportunities along the way. Thanks for all the intelligent suggestions during the progress of this work, and always inspiring me to find a way past difficulty and keep going on; for this I am highly thankful.
- To friends, staff and colleagues at ICGEB Cape Town Component; Irene, Nurhaan, Claudia, and others in the main office, I appreciate your moral, administrative and stationery support. I thank all ICGEB Cape Town Fellows, both past and present, for sharing your experiences; and birthday and farewell cakes with me.
- To member of the Cancer Genomics Group at ICGEB Cape Town past and present, under the leadership of A/Prof. Luiz Fernando C. Zerbini. To Juliano, Kristal, and Deidre, I appreciate your assistance with building the sample repositories and our many journeys to New Sommerset Hospital and Eerste Rivier Hospital; you guys made the journey worthwhile.
- To members of the Applied Proteomics and Chemical Biology Group at UCT past and present, under the leadership of Prof. Jonathan M. Blackburn. I really appreciate all your peer review contributions and suggestions during the progress of this work, particularly Nelson for proteomics workflow, Shaun Garnett for R statistical tutorial; and Brandy and Bridget for being my dependable “urine buddies” and not to forget Muneerah for her enduring support during my protein microarray project.
- To Members of the Urology Team at Grootes Schuur Hospital for their immeasurable assistance with participant selection and obtaining ethical consent from participants.
- To International Centre for Genetic Engineering and Biotechnology (ICGEB) for funding me through a Pre-Doctoral Fellowship during my PhD studies.

- To my Oral Pathology/ Dental colleagues at Tygerberg Hospital, I really appreciate your support and keeping me in touch with contemporary pathology throughout my PhD period.
- To Professor Johannes J. Hille, for believing in me and always challenging me to go the extra mile; and to Professor L.X.G. Stephen and Professor Y. Osman for giving me an enabling opportunity to develop myself in Oral and Maxillofacial Pathology.
- Finally, to my darling wife Omolola Oluwatoyin Adeola and My lovely children Oluwadamilola John, Oluwatomilola Mary and Ifeoluwa Joanna; you guys were a solid rock of prayers and support. I really appreciate your understanding and all the sacrifices you made towards the realization of this goal. You made the whole journey altogether lovely.

PUBLICATIONS

JOURNAL ARTICLES

- **Adeola Henry A**, Soares Nelson C, Paccez Juliano D, Kaestner Lisa, Blackburn Jonathan M, Zerbini Luiz F (2015). Discovery of novel candidate urinary protein biomarkers for prostate cancer in a multiethnic cohort of South African patients via label-free mass spectrometry. ***Proteomics Clinical Applications 2015 Jun;9(5-6):597-609. Article ID#: 25708745 (PMID)***
- **Henry A. Adeola**, Bridget Calder, Nelson C. Soares, Lisa Kaestner, Jonathan M. Blackburn, Luiz F. Zerbini (2016). In silico verification and parallel reaction monitoring prevalidation of potential prostate cancer biomarkers. ***Future Oncology 2016 Jan;12(1):43-57. doi: 10.2217/fon.15.296. Article ID#: 26615920 (PMID)***
- **Henry A. Adeola,** Muneerah Smith, Lisa Kaestner, Jonathan M. Blackburn, Luiz F. Zerbini (2016). Novel potential serological prostate cancer biomarkers using CT100+ cancer antigen microarray platform in a multi-cultural South African cohort. ***Oncotarget 2016 Oncotarget. 2016 Feb 12. doi: 10.18632/oncotarget.7359. [Epub ahead of print] Article ID#: 26885621 (PMID)***

BOOK CHAPTERS

- **Henry Adeola**, Ryan William Goosen, Paul Goldberg and Jonathan Blackburn (2014). Prospects of 'Omics Based Molecular Approaches in Colorectal Cancer Diagnosis and Treatment in the Developing World: A Case Study in Cape Town, South Africa. ***Colorectal Cancer - Surgery, Diagnostics and Treatment, Dr. Jim Khan (Ed.), ISBN: 978-953-51-1231-0.*** InTech DOI: 10.5772/57485

CONFERENCE PROCEEDINGS

Poster Presentations

- **Henry Adeola**, Juliano Paccez, Lisa Kaestner, Nelson Soares, Jonathan Blackburn, Luiz Zerbini. Poster Title: Urinary Proteomics for Prostate Cancer Biomarker Discovery in a South African Cohort. ***3rd Biennial science of global prostate cancer disparity in black***

men conference, 5-8 November 2014, Montego Bay, Jamaica. Abstract ID# Session 1(06/11/2014)

- **Henry Adeola**, Juliano Paccez, Lisa Kaestner, Nelson Soares, Jonathan Blackburn, Luiz Zerbini. Poster Title: Urinary Proteomics for Prostate Cancer Biomarker Discovery in a South African Cohort. ***University of Cape Town Research Day, Faculty of Health Sciences, September 2015.***

Oral/Podium Presentation

- **Henry Adeola**, Juliano Paccez, Lisa Kaestner, Nelson Soares, Jonathan Blackburn, Luiz Zerbini. Presentation Title: Urinary Proteomics for Prostate Cancer Biomarker Discovery in a South African Cohort. ***3rd Biennial science of global prostate cancer disparity in black men Conference, 5-8 November 2014, Montego Bay, Jamaica***. Session ID#: Session 7 Late Breaking Topics: Basic Research (8/11/2014)

ABBREVIATIONS

5-AR: 5- ALPHA REDUCTASE

ABC: AMMONIUM BICARBONATE BUFFER

ACPP: PROSTATIC ACID PHOSPHATASE

AJCC: AMERICAN JOINT COMMITTEE ON CANCER

AL: ANTERIOR LOBE OF PROSTATE GLAND

AMACR: ALPHA-METHYLACYL-COA RACEMASE

ANOVA: ANALYSIS OF VARIANCE

APC: ADENO-POLYPOSIS COLI

AUC: AREA UNDER CURVE

BCA: BICINCHONINIC ACID

BCCP: BIOTIN CARBOXYLASE-BIOTIN CARBOXYL CARRIER PROTEIN

BPH: BENIGN PROSTATIC HYPERPLASIA

BSA: BOVINE SERUM ALBUMIN

CD: CLUSTER OF DIFFERENTIATION

cDNA: COMPLEMENTARY DNA

CGG: CANCER GENOMICS GROUP

CID: COLLISION INDUCED DISSOCIATION

COG: CLUSTERS OF ORTHOLOGOUS GROUPS

CRAN: THE COMPREHENSIVE R ARCHIVE NETWORK

CTA: CANCER –TESTIS ANTIGEN

CTC: CIRCULATING TUMOUR CELL

ctNA: CIRCULATING NUCLEIC ACID

CV: COEFFICIENT OF VARIATION

DC: DISEASE CONTROLS

DHT: DIHYDROTESTOSTERONE

DNA: DEOXYRIBONUCLEIC ACID

DOH: DECLARATION OF HELSINKI

DRE: DIGITAL RECTAL EXAMINATION

DTT: DITHIOTHREITOL

EBV: EPSTEIN–BARR VIRUS

ECD: ELECTRON CAPTURE DISSOCIATION

ECM: EXTRACELLULAR MATRIX

EDTA: ETHYLENEDIAMINETETRAACETIC ACID

ELISA: ENZYME-LINKED IMMUNOSORBENT ASSAY

ERH: EERSTE RIVIER HOSPITAL

ESI: ELECTROSPRAY IONIZATION

ETD: ELECTRON TRANSFER DISSOCIATION

FA: FORMIC ACID

FASP: FILTER AIDED SAMPLE PREPARATION

FDR: FALSE DISCOVERY RATE

FFPE: FORMALIN-FIXED PARAFFIN-EMBEDDED

FTIC: FOURIER TRANSFORM ION CYCLOTRON

FWER: FAMILY-WISE ERROR RATE

GO: GENE ONTOLOGY

GSH: GROOTES SCHUUR HOSPITAL

H&E: HEMATOXYLIN AND EOSIN

HBV: HEPATITIS B VIRUS

HCD: HIGH-ENERGY COLLISION DISSOCIATION

HEPES: 4-(2-HYDROXYETHYL)-1-PIPERAZINEETHANESULFONIC ACID

HGPIN: HIGH GRADE PROSTATIC INTRAEPITHELIAL NEOPLASM

HHV8: HUMAN HERPES VIRUS 8

HIV: HUMAN IMMUNODEFICIENCY VIRUS

HPA: HUMAN PROTEIN ATLAS

HPV: HUMAN PAPILLOMA VIRUS

HREC: HUMAN RESEARCH ETHICS COMMITTEE

IARC: INTERNATIONAL AGENCY FOR RESEARCH ON CANCER

iBAQ: INTENSITY BASED ABSOLUTE QUANTIFICATION

ICGEB: INTERNATIONAL CENTRE FOR GENETIC ENGINEERING AND BIOTECHNOLOGY

IMRT: INTENSITY-MODULATED RADIATION THERAPY

IPI: INTERNATIONAL PROTEIN INDEX

ITH: INTRA-TUMOUR HETEROGENEITY

iTRAQ: ISOBARIC TAG FOR RELATIVE AND ABSOLUTE QUANTITATION

KEGG: KYOTO ENCYCLOPEDIA OF GENES AND GENOMES

KLK: KALLIKREIN

LFQ: LABEL FREE QUANTIFICATION

LGPIN: LOW GRADE PROSTATIC INTRAEPITHELIAL NEOPLASM

LH: LUTEINIZING HORMONE

LHRH: LH- RELEASING HORMONE

LMIC: LOW AND MIDDLE INCOME COUNTRIES

lncRNA: LONG NON-CODING RNA

MALDI: MATRIX-ASSISTED LASER DESORPTION/IONIZATION

ML: MEDIAN LOBE OF PROSTATE GLAND

MR: MEDICAL RECORDS

MRI: MAGNETIC RESONANCE IMAGING

MWCO: MOLECULAR WEIGHT CUT-OFF

NAPPA: NUCLEIC ACID PROGRAMMABLE PROTEIN ARRAYS

NC: NORMAL HEALTHY CONTROLS

NCE: NORMALIZED COLLISION ENERGY

NE: NEUROENDOCRINE

NHLS: NATIONAL HEALTH LABORATORY SERVICES

NSH: NEW SOMMERSET HOSPITAL

PASSEL: PEPTIDEATLAS SRM EXPERIMENT LIBRARY

PATR: PEPTIDEATLAS TRANSITIONS RESOURCE

PBS: PHOSPHATE BUFFERED SALINE

PBST: PBS IN TRITON-X

PCA: PROSTATE CANCER

PCR: POLYMERASE CHAIN REACTION

PEP: POSTERIOR ERROR PROBABILITY

PG: PROSTATE GLAND

PHI: PROSTATE HEALTH INDEX

PISA: PROTEIN IN-SITU ARRAYS

PL: POSTERIOR LOBE OF PROSTATE GLAND

PMT: PHOTOMULTIPLIER TUBE

POC: POINT-OF-CARE

PREST: PROTEIN EPITOPE SIGNATURE TAGS

PRM: PARALLEL REACTION MONITORING

PSA: PROSTATE SPECIFIC ANTIGEN

PSD: POST SOURCE DECAY

PTP: PROTEOTYPIC PEPTIDE

QC: QUALITY CHECK

QQQ: TRIPLE QUADRUPOLE

RFU: RELATIVE FLUORESCENCE UNITS

RNA: RIBONUCLEIC ACID

RPM: REVOLUTIONS PER MINUTE

RTqPCR: REAL-TIME QUALITATIVE PCR

SDS-PAGE: SODIUM DODECYL SULFATE POLYACRYLAMIDE GEL ELECTROPHORESIS

SEREX: SEROLOGICAL ANALYSIS OF EXPRESSION CDNA LIBRARIES

SILAC: STABLE ISOTOPE LABELING BY AMINO ACIDS IN CELL CULTURE

SNP: SINGLE NUCLEOTIDE POLYMORPHISM

SNR: SIGNAL-TO-NOISE RATIO

SRM: SELECTED REACTION MONITORING

STRING: SEARCH TOOL FOR THE RETRIEVAL OF INTERACTING GENES/PROTEINS

SVA: SURROGATE VARIABLE ANALYSIS

SVD: SINGULAR VECTOR DECOMPOSITION

SWATH: SEQUENTIAL WINDOW ACQUISITION OF ALL THEORETICAL FRAGMENT ION SPECTRA

TAA: TUMOUR-ASSOCIATED ANTIGEN

TCA: TRICHLOROACETIC ACID

TCEP: TRIS (2-CARBOXYETHYL) PHOSPHINE

TIFF: TAG IMAGE FILE FORMAT

TMA: TISSUE MICROARRAY

TMPRSS2-ERG: TRANSMEMBRANE PROTEASE, SERINE 2- ETS FUSION GENES

TMT: TANDEM MASS TAG

TNM: TUMOUR, NODE, AND METASTASIS

TOF: TIME OF FLIGHT

TURP: TRAN-URETHRAL RESECTION OF PROSTATE

TRUS: TRANS-RECTAL ULTRASOUND

UA: UREA BUFFER

UGS: UROGENITAL SINUS

UHPLC: ULTRA HIGH PERFORMANCE LIQUID CHROMATOGRAPHY

UT: UROLOGY TEAM

VEGF: VASCULAR ENDOTHELIAL GROWTH FACTOR

WHO: WORLD HEALTH ORGANIZATION

TABLE OF CONTENT

LIST OF FIGURES

Figure 3.4: Box- whisker plot to assess the spread of data in each study group (BPH, NC, PCa &DC). On the x-axis are the different pooled groups and on the y-axis are log2 -transformed IBAQ values.

Figure 3.5: Q-Q plot used to assess data spread in PCa, BPH, DC and NC by comparing the experimental quantiles with theoretical quantiles. On the x-axis is the normal theoretical quantile while on the y-axis is the sample quantiles. The data points can be seen to be within acceptable distance of the reference line (red).

Figure 3.6: House-keeping protein profile revealed that β2-microglobulin (B2M) and β-actin were stably expressed across all 4 pooled groups. Nine common housekeeping proteins from literature were selected. On the x-axis are the housekeeping proteins and on the y-axis are the log2-transformed IBAQ intensities for the protein groups.

Figure 3.7: Protein profile shows β-actin to be stably expressed across duplicate pooled sample runs. On the x-axis are the pooled sample replicates and on the y-axis are log2-transformed IBAQ values.

Figure 3.8: Venn diagram showing unique and overlapping protein sets in BPH, DC, NC, and PCa.

Figure 3.9: Volcano plot of differentially expressed (±2SD) protein groups between PCa and NC.

Figure 3.10: Hierarchical clustering of PCa, BPH, NC and DC.

Figure 3.11: Power plot showing that 43 samples were required to attain a statistical power of 0.8. This is a power curve generated automatically from G*Power using statistical parameters generated from the initial pooling experiment.

Figure 3.12: Protein profile plot demonstrated β-2-microglobulin to be stably expressed across multiple individual samples.

Figure 3.13: Protein profile plot demonstrates β–actin to be stably expressed across multiple individual samples.

LIST OF TABLES

ABSTRACT

In Africa, Prostate cancer (PCa) is the most frequently diagnosed solid organ tumour in males and use of prostate specific antigen (PSA) is presently fraught with diagnostic inaccuracies. Not least, in a multi-ethnic society like South Africa, proteome differences between African, Caucasian and Mixed-Ancestry PCa patients are largely unknown. Hence, discovery and validation of affordable, non-invasive and reliable diagnostic biomarkers of PCa would expand the frontiers of PCa management. We have employed two high-throughput proteomics technologies to identify novel urine- and blood-based biomarkers for early diagnosis and treatment monitoring of prostate cancer in a South African cohort as well as elucidate proteome differences in patients from our heterogeneous cohort. We compared the urinary proteomes of PCa, Benign Prostatic Hyperplasia (BPH), disease controls comprising patients with other uropathies (DC) and normal healthy controls (NC) both by pooling and individual discovery shotgun proteomic assessment on a nano-Liquid chromatography (nLC) coupled Hybrid Quadrupole-Orbitrap Mass Spectrometer platform. In-silico verification of identified biomarkers was performed using the Human Protein Atlas (HPA) as well as SRMAtlas; and verified potential biomarkers were experimentally prevalidated using a targeted parallel reaction monitoring (PRM) proteomics approach. Further, we employed the CT100+ antigen microarray platform to assess the differential humoral antibody response of PCa, DC and BPH patients in our cohort to a panel of 123 tumour-associated cancer antigens. Candidate antigen biomarkers were analyzed for ethnic group variation in our cohort and potential cancer diagnostic and immunotherapeutic inferences were drawn. Using these approaches, we identified 5595 and 9991 non-redundant peptides from the pooled and individual experiments respectively. While nine proteins demonstrated ethnic trend, 37 and 73 proteins were differentially expressed by pooled and individual analysis respectively. All 32 verified biomarkers were prevalidated with parallel reaction monitoring. Good PRM signals for 12 top ranking biomarker was observed, including PSA and prostatic acid phosphatase. We also identified 41 potential diagnostic and immunotherapeutic antigen biomarkers. Proteogenomic functional pathway analyses of differentially expressed antigens showed similar enrichments of biologic processes. We identified herein novel urinary and blood-based potential diagnostic biomarkers and immunotherapeutic targets of PCa in a South African PCa Cohort using multiple proteomics approaches.

CHAPTER 1: GENERAL INTRODUCTION

1.1. Functional Anatomy of the Prostate Gland

The prostate gland (PG) is a compound tubule-alveolar accessory exocrine organ of the male genitourinary system. Its name is derived from the Greek word "προστάτη", which stands for "the one standing in front" (http://etymonline.com/?term=prostate) due to its position below the bladder in the pelvis (**Fig. 1.1**). It is typically the size of a walnut/chestnut in healthy individuals, and weighs about 7-16g with a mean of ca. 11g [1]. It secretes and stores a moderately alkaline fluid (seminal plasma) which provides about 50-75% of the volume of the human semen (~3mL). This fluid is important for preserving the lifespan of the spermatozoa as well as liquefaction of the semen. The function and development of the PG is regulated by androgen hormones [2, 3] such as dihydrotestosterone (DHT), testosterone,

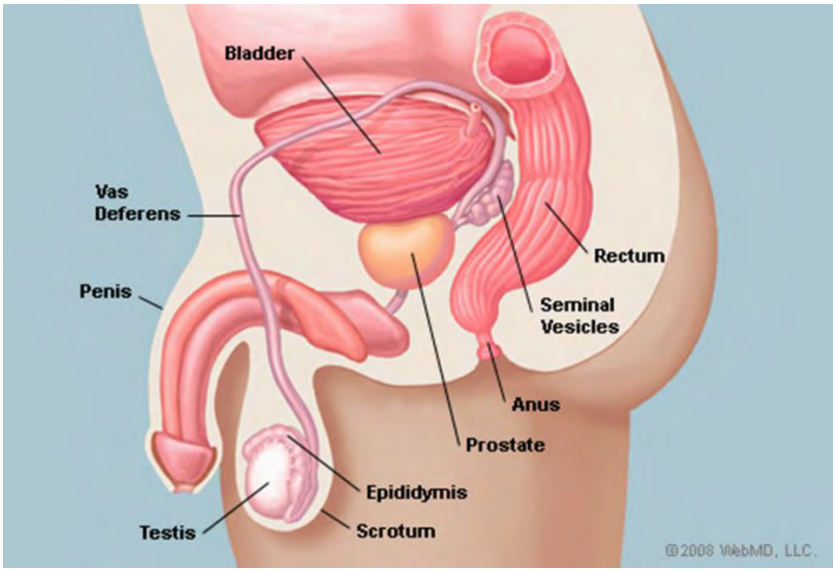


Figure 1.1 Sagittal section of the male pelvis showing the position of the prostate gland (http://www.webmd.com/)

dehydroepiandrosterone, and dihydrotestosterone (**Fig. 1.2**).

Anatomically, it is a cone shaped gland made up of smooth muscles, epithelium and connective tissue [4]. The PG base is located cephalad against the neck of the bladder and its apex is directed backward towards the urogenital diaphragm. Its posterior surface is separated from the rectum by the recto-vesical septum. The anterior aspect is bounded by the retro-pubic space (containing a venous plexus) which separates it from the pubic symphysis. The levator ani muscle and a venous plexus faces its lateral wall [4]. The prostatic urethra passes through the PG and the seminal vesicles and bulbourethral glands are proximally linked to it as well. The front of the PG shows the urethral crest, seminal colliculus and the prostatic sinus into which the prostatic duct opens.

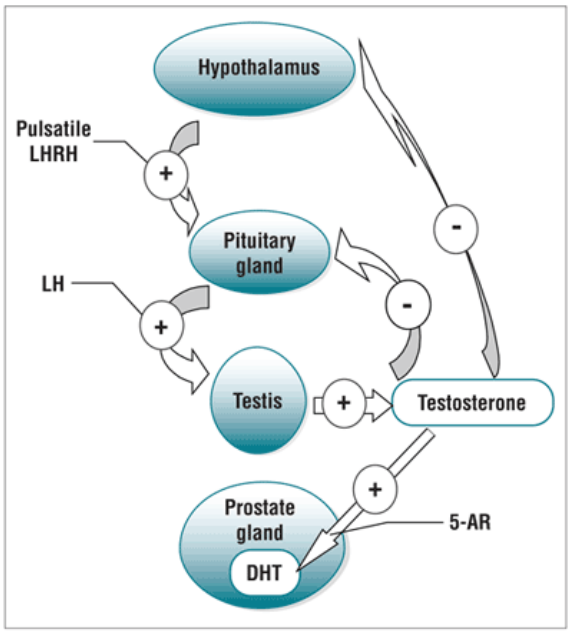


Figure 1. Hypothalamic-pituitary-testicular axis. DHT: dihydrotestosterone; 5-AR: 5-alpha reductase; LH: luteinizing hormone; LHRH: LH-releasing hormone.

Figure 1.2. Peripheral and central hormone effect on the prostate gland (prostatesupplementguide.com)

The PG can be divided anatomically into lobes [5] *viz*. two lateral lobes that extends to all zones, a posterior lobe, an anterior lobe also referred to as the "isthmus" which is anterior to the prostatic urethra, and a median lobe which is posterior to the prostatic urethral and flanked by the two ejaculatory duct on either sides. The posterior lobe (PL) corresponds approximately to the peripheral zone, while the median (ML) and anterior (AL) lobes correspond to the central

and transitional zones respectively (**Table 1.1**). However, the lobar classification is not well delineated in adult prostate tissue. Hence, in recent times, zonal anatomic classification of the prostate gland has also emerged. Using the zonal anatomic model by McNeal [6], the PG is divided into 3 glandular zones with an anterior fibromuscular component without glandular components. The three major zones are: peripheral, central and transitional zones. The peripheral zone accounts for about 70% of the gland, while the central and transition zones account for about 25% and 5% respectively. Nearly 70-80% of prostate cancer (PCa) occurs in the peripheral zone and about 10-20% in the transitional zone. Only about 2.5% of PCa occurs in the central zone, and tend to be more aggressive. Most benign prostatic hyperplasia (BPH) occurs in the transitional zone which is known to grow throughout life.

Zones	% Gland involved	% in PCa
Peripheral (PL)	~ 70	~ 70-80
Central (ML)	~ 25	~2.5 (more aggressive)
Transition (AL)	~5	~10-20 (responsible for BPH, grows throughout)
Anterior Fibromuscular	~5	No glandular component

Table 1.1 Lobes and zones of the prostate gland

The role of numerous secreted components of the seminal plasma is unclear except for a few enzymes (including Kallikrein (KLK) 3, also known as prostate-specific antigen (PSA)) which are known to participate in the seminal plasma enzymatic clotting and lysis cascade [7]. PSA is a serine protease which is known to cleave the clotting protein semenogelin (I & II) [7], which is an important component of the seminal coagulum. Prostatic fluids have also been found to contain prostaglandins, zinc, potassium, citric acid and several proteins and poly-/oligo-peptides [8]. These substances have been found to possess antimicrobial properties as well as improve sperm motility and viability.

Inadvertent resection of the bladder neck involving the distal urethral sphincter or the periprostatic neuro-vasculature may result in urinary incontinence and

impotence respectively [9]; albeit the PG is not an essential organ to life and can be totally resected during radical prostatectomy without life-threatening implications [10].

Histologically, there is no characteristic separation of acinar and ductal structures as found in most other glands of the body. The normal prostatic epithelium is typically divided into two cellular layers comprising of an androgen dependent secretory luminal layer and an androgen independent basal epithelial layer [11]. These two layers are scantily interspersed by a small population of cells called the transit-amplifying or neuroendocrine (NE) cells [12] which have sensory and endocrine attributes and are hypothesized to originate embryologically from prostatic stem cells and are able develop into luminal or basal cells and hence are involved in growth and differentiation of the PG. Aggressive prostate cancer (PCa) growth or progression can lead to differentiation of PCa cell into a more androgen-refractory "NE-like" phenotype [13, 14]. This malignant phenotype promotes anti-apoptotic properties by perturbation of *survivin* as well as neoangiogenesis via vascular endothelial growth factor (VEGF) augmentation [15]. Under benign conditions, the basal epithelial layer separates the luminal layer from the basement membrane and the absence of basal layer is highly suspicious of malignant invasion of cancer cells. In addition, the adult PG demonstrates a branching acinar-ductal system within a dense fibro-muscular connective tissue stroma.

Immunohistochemically, the luminal layer is positive for PSA, prostatic acid phosphatase (ACPP), androgen receptor (AR), vimentin, CK8, and CK18 [16]; while the basal layer is negative for all these immunohistochemical markers and NE cells are positive for ACPP and chromogranin-A. In addition, distinct cell types with unique immunohistochemical properties in comparison with benign or normal prostatic epithelium are found during cancer development in the PG.

Embryogically, the dual origin of the PG stems from the fact that contiguous structures such as the ejaculatory ducts, seminal vesicle, epididymis, and vas deferens; as well as the central zone of the PG are hypothesized to originate from the mesodermal Wolffian duct, while the rest of the PG is derived from

endodermal invagination into surrounding mesenchyme [17]. The primordium for the future PG, the urogenital sinus (UGS) develops around the 9-11th week of embryonic life and the primitive testis stimulates endodermal bud ingrowths into the UGS [18]. Androgens then cause the UGS epithelium to differentiate into prostatic ducts around the 13th week [19], while peri-ductal smooth muscles are formed by the urogenital mesenchyme. Around the 20-30th week of gestation, a solid bud which is made of central spindle cells and peripheral columnar cells appears at the distal end of the ducts. Between 30-35th weeks of gestation the gland undergoes tubulo-alveolar differentiation and develops a lobular architecture [19, 20]. There is a resting phase for PG enlargement after birth, albeit budding morphogenesis and ductal proliferation does not cease; androgen dependent glandular enlargement and maturation is resumed again in the peri-pubertal teenage years.

1.2. Basic Concepts in Cancer

Cancer is a multifactorial disease which has a variable incidence, mortality, and prevalence rate in different populations globally. Chemical carcinogens such as beryllium, arsenic, benzene, asbestos, cadmium, chromium, nickel, ethylene oxide, radon and vinyl chloride have been shown to demonstrate reasonable evidence of involvement in cancer development [21, 22].

Cancers can also be inherited in an autosomal dominant manner [23]. A few examples of inherited cancer syndromes include: Li Fraumeni Syndrome (p53), retinoblastoma (Rb), hereditary non-polyposis colon cancer (MSH2, MSH6, MLH1), Familial adenomatous polyposis coli (APC) and breast cancer (BRCA1 and BRCA2). Autosomal recessive inheritance of DNA repair defects includes: Bloom syndrome, Fanconi anaemia, xeroderma pigmentosum and ataxia-telangiectasia [23].

Normal growth regulatory genes are typically proto-oncogenes, tumour suppressor genes, DNA repair genes and Apoptotic genes [23]. Tumour suppressor genes can be promoters (e.g. p53 and Rb) or caretaker genes which

preserve the integrity of the genome and prevent the development of mutator phenotypes. Two landmark papers that described the major hallmarks of cancer were published by Hanahan and Weinberg in 2000 and 2011. Their earlier paper described the six major hallmarks of cancer cells which were: limitless replicative potential, evading apoptosis, insensitivity to anti-growth signals, self-sufficiency in growth signals, sustained angiogenesis, and tissue invasion/ metastasis [24]. In their second paper, four additional emerging hallmarks of cancer cells were added *viz*. deregulating cellular energetics, avoiding immune destruction, tumour-promoting inflammation and genome instability/ mutation [25]. Clearly these 10 hallmarks (**Fig. 1.3**) do not occur in all cancer cell types at all times [26] and changes in the tumour microenvironment and potential for stromal metabolic reprogramming of tumour cells needs to be carefully considered [27].

Another important consideration is intra-tumour heterogeneity (ITH) or inter-biopsy heterogeneity which suggests that multiple genetic mutation and phenotypes of cancer can be found in the same cancerous lesion. This was well demonstrated by Gerlinger et al [28] using multi-region sequencing of renal cancer biopsies. ITH in clonal expansion of cancer stem cells was shown by these authours to follow a Darwinian evolution pattern. Cancer progression and development of ITH is dependent on factors like growth rate, invasiveness, hormones, metastatic potential, karyotype and responsiveness to anticancer drugs. An important theory for cancer cell expansion is the "cancer stem cell" theory, which presupposes that only a fraction of the cancer cells, known as stem cells drive tumour expansion [29]. Cancer biology is indeed enigmatic and multiple pathways and processes are involved in its pathogenesis.

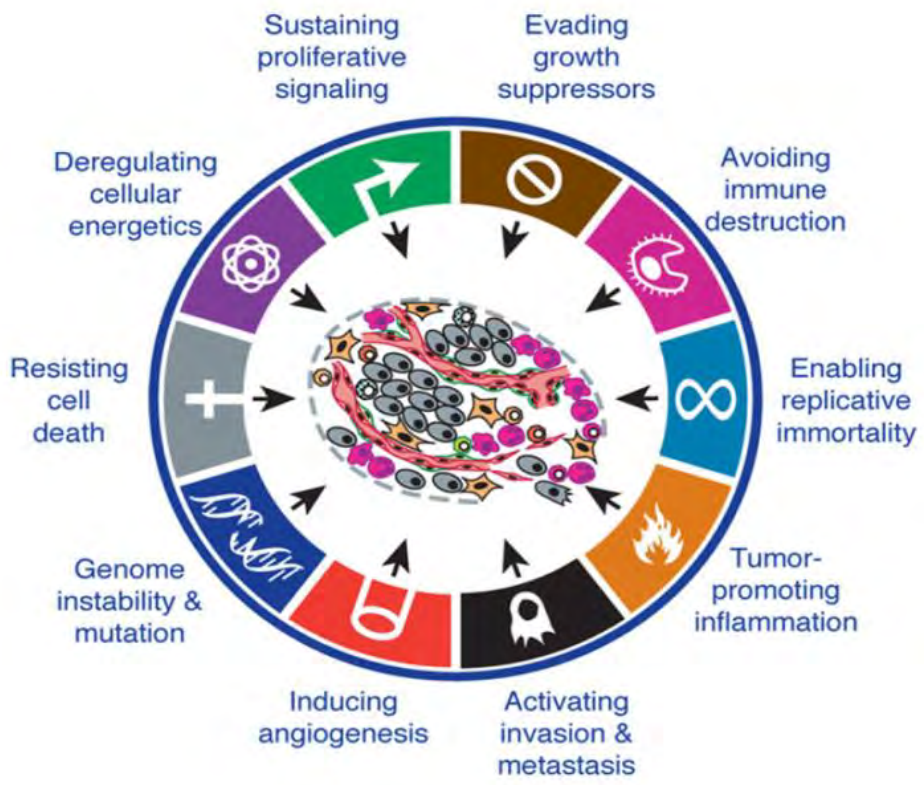


Figure 1.3. Ten essential hallmarks of Cancer are: limitless replicative potential, evading apoptosis, insensitivity to anti-growth signals, self-sufficiency in growth signals, sustained angiogenesis, tissue invasion/ metastasis, deregulating cellular energetics, avoiding immune destruction, tumour-promoting inflammation and genome instability/mutation. Although all ten do not occur in all cancer cell types at all times. (adapted from [25])

1.3. Prostate Cancer

Analogous to cancer at any other site in the body, prostate cancer (PCa) is an uncontrolled, abnormally disordered malignant growth of tissue in the PG, which is found only in men. It could be a small growth limited to the PG in which case it is referred to as localized; or it could gradually enlarge and invade adjacent or distant structures (**Fig. 1.4**). Early detection and treatment of PCa while it is still localized is important for better disease outcome. Some prostate cancers grow very slowly and may not be problematic for many years after its diagnosis, while some may run a very aggressive course and spread to other organs shortly after detection. Factors responsible for indolent or aggressive phenotypes of PCa are poorly understood. The risk of developing PCa significantly increases after the

age of 40 years [30]. Even without any form of treatment, PCa often runs a protracted natural history and many men die with it rather than from it [31].

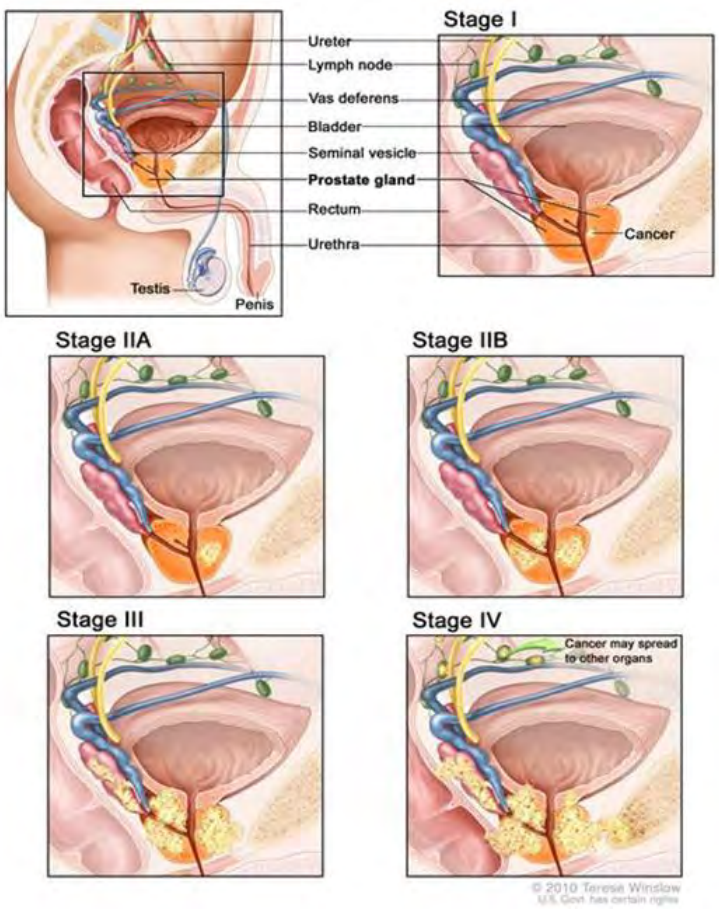


Figure 1.4. Stages of progression of prostate cancer (www.cancer.gov)

1.3.1. Prostate Cancer Epidemiology

PCa is a leading cause of death in elderly males globally. Over a million new cases are reported annually according to the GLOBOCAN/ IARC 2012 databases, second only to lung cancer. According to this database, PCa is the fifth leading cause of cancer death in men globally, accounting for up to 307,000 deaths annually [32]. It was reported by this global cancer statistics database to have the highest incidence (59,493), mortality (42,802) and 5 year prevalence rate (155,028) in Africa.

This high incidence, mortality and 5-year prevalence trend is similar whether you look at sub-Saharan Africa, Southern Africa or the Republic of South Africa (http://globocan.iarc.fr/). In contrast to the situation in the developed world, where high incidence and low mortality reflect the efficacy of the early diagnosis and prompt treatment; most of cases are diagnosed at advanced stages in Africa and the incidence and mortality rates are almost at similar levels in Middle, Western, and Eastern Africa (**Fig. 1.5**).

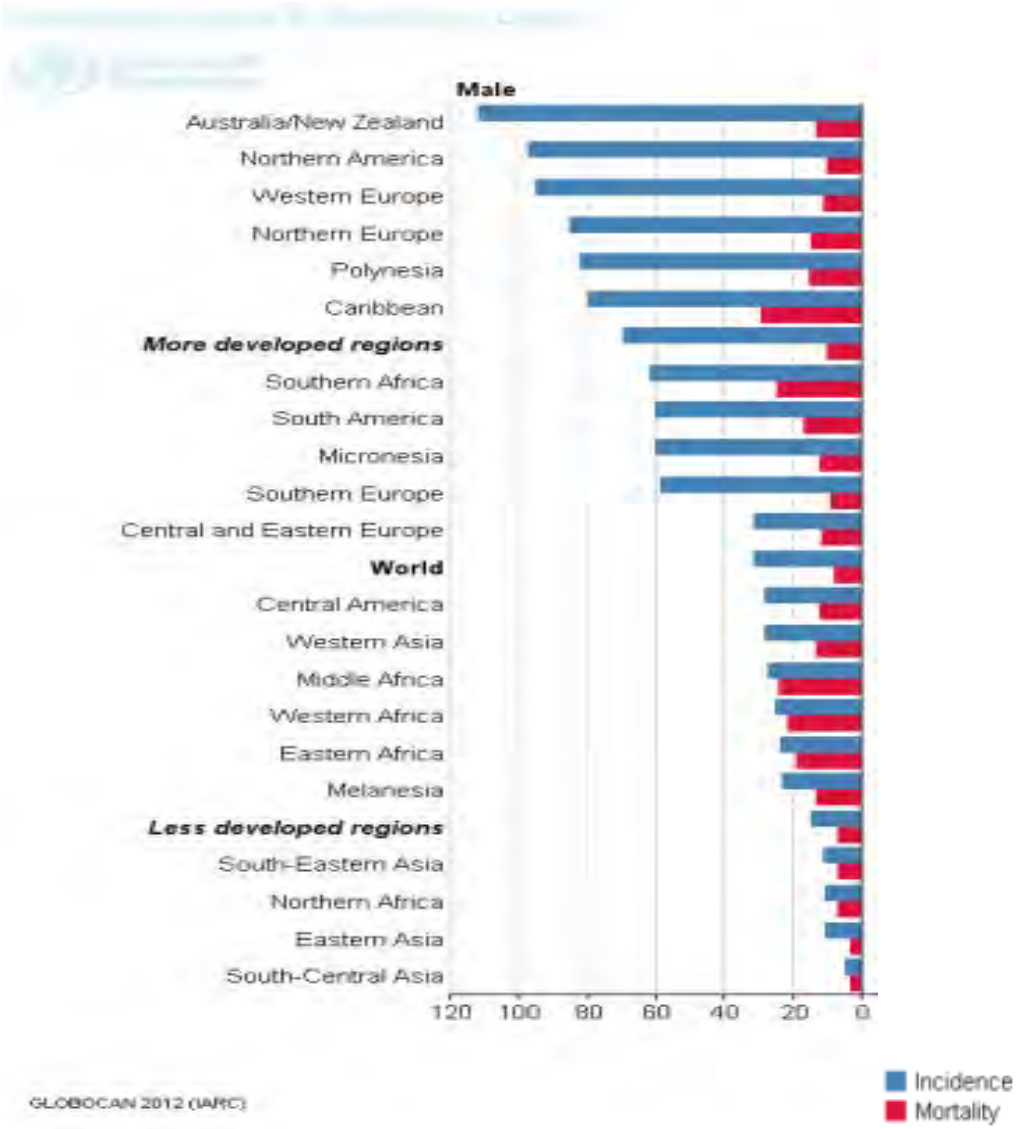


Figure 1.5. Bar graph of Global Incidence and Mortality of PCa (http://globocan.iarc.fr/)

Possibly due to a relatively higher level of development and infrastructure in South Africa, PCa incidence is quite high compared to the rest of Africa (**Fig. 1.6**); but unfortunately, mortality rates in this region are high as well (**Fig. 1.7**).

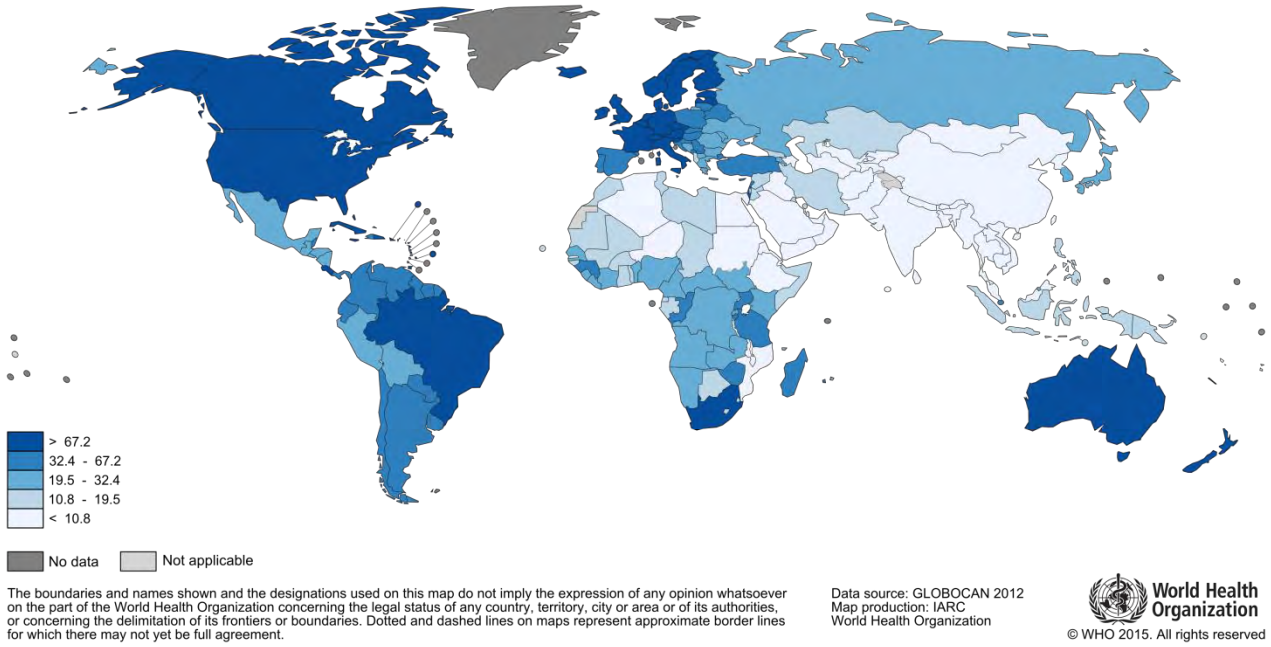


Figure 1.6. Map of Global Incidence of PCa (http://globocan.iarc.fr/)

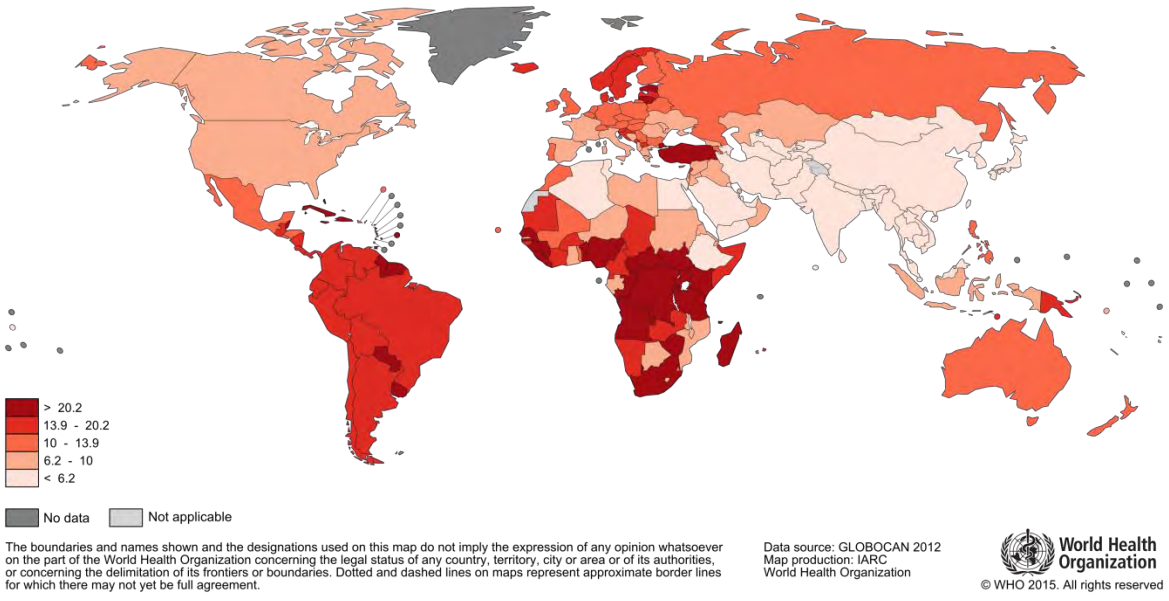


Figure 1.7. Map of Global Mortality of PCa (http://globocan.iarc.fr/)

The fact that South Africa has a high incidence rates of PCa similar to those in North America, Western Europe and Australia (**Fig. 1.6**) but a high mortality rate comparable to most other sub-Saharan African countries (**Fig. 1.7**) may indicate that despite having a relatively better diagnostic infrastructure; manpower and resources are lacking to manage the huge burden of diagnosed PCa cases.

1.3.2. Aetiological/ Risk factors

Even though the main causative factors for prostate carcinogenesis are poorly understood, several aetiologic and risk factors have been suggested to be closely associated with PCa development. Three important non-modifiable risk factors suggested for PCa are positive family history of PCa, age and racial profile. On the other hand, modifiable risk factors include obesity, smoking, alcohol consumption, androgens, diet, and diabetes mellitus, *inter alia*. Men over the age of 40 and men of African descent are at greater risk of developing PCa compared to their Caucasian counterparts. Men of Asian origin have the least risk of developing PCa, albeit their risk increases when they migrate to North America [33], as most North American diets are richer in meat and animal fat compared with fruit and vegetables.

Large studies have shown that red meat consumption and total fat intake significantly increases the risk of developing PCa [34]. Dietary consumption of red tomatoes rich in antioxidant carotenoid lycopene [35]; cruciferous vegetables rich in isothiocyanate sulforaphane, which has chemoprotective properties [36]; and other antioxidant nutrients such as vitamin E [37] and selenium [38] have been found to correlate with reduced PCa risk. High consumption of soy products rich in isoflavinoids in eastern Asia has been suggested to be responsible for reduction in PCa risk compared to the western world [39]. Body mass index has been show to correlate with risk of advanced PCa, albeit no correlation was found for localized disease [40]. Many other studies have investigated the effect of various diets on PCa risk including green tea [41], fish

[42] and dairy products [43], *inter alia*. Even though PCa has been shown not to be related to alcohol consumption by some studies [44], binge drinking was shown to increase its risk [45]. Smoking, which exposes individuals to polycyclic aromatic hydrocarbons has been demonstrated to increase the mortality risk in PCa by 14% in a meta-analysis of twenty-four cohort studies [46].

Many studies have indicated that there is an inverse relationship in the risk of developing PCa and type II diabetes mellitus, despite the presence of hyperinsulinemia [47]. Hereditary and genetic factors have been suggested to play a role in PCa development. An inherited PCa susceptibility gene locus on 1q24–25 (HPC1) has been well researched. Some other susceptibility genes that have received PCa research attention include MSR1 (8p22), RNASEL (1q25), CYP17, ELAC2, Androgen Receptor (AR) and steroid-5-α-reductase type II (SRD5A2) [48]. Hypermethylation of CpG island surrounding GSTP1 gene has been suggested to result in somatic defects in PCa leading to loss of "caretaker" effect of GSTP1 in PCa [49]. Other genes widely studied for somatic defects in PCa include PTEN, CDKN1B, and NKX3.1.

Inflammatory prostatitis, including those that result from sexually transmitted infections (regardless of infective agent) have been reported to increase the risk of developing PCa [50, 51].

1.3.3. Histopathology of Prostatic Diseases and Prostate Cancer

Pathologic processes that affect the PG can be broadly divided into three *viz* inflammatory processes, benign hyperplastic processes, and malignant tumours. Inflammatory processes include acute or chronic prostatitis, granulomatous prostatitis and chronic non-bacterial prostatitis. Acute bacterial prostatitis can be causes by *Escherichia Coli* (E.Coli), *Staphylococci*, *Enterococci* and other gram negative rods that cause urinary tract infections [52]. Granulomatous prostatitis can be caused by fungal or mycobacterial infection in immunocompromised patients. Non-bacteria prostatitis can be caused by bicycle riding, backflow of

urine into the prostate, irritation from chemicals, nerve problems and viruses. Benign hyperplastic processes include benign nodular hyperplasia and benign prostatic hyperplasia (BPH). The major suggested aetiologies for benign hyperplastic processes, demonstrated in canine PGs involves imbalance between the normal ratio of oestrogen and androgens [53], which leads to autocrine stimulation of stroma cells and paracrine stimulation of neighbouring epithelial cells by testosterone.

Prostatic tissue that begin to demonstrate signs of cellular atypia are potentially malignant or precancerous and are generally referred to as prostatic intraepithelial neoplasia (PIN), which can either be low (LGPIN) or high grade (HGPIN). Generally speaking, potential for cancer development in LGPIN and high HGPIN are low and high respectively, albeit PIN is a well-established precursor of PCa [54]. Invasive carcinomas of the prostate are the commonest malignant lesions of the prostate; and acinar adenocarcinomas constitutes ca. 90-95% of all adenocarcinomas [55].

Common acinar variants include pseudohyperplastic, oncocytic, mucinous, foamy, atrophic, lymphoepithelioma-like and signet ring; while non-acinar variants include squamous and adenosquamous carcinomas, urothelial carcinomas, sarcomatoid carcinomas, basal cell carcinomas, neuroendocrine carcinomas, and ductal carcinomas [55].

Carcinomas of the prostate can be diagnosed as an unexpected finding during autopsy in which case it is referred to as latent; can be an incidental finding after endoscopic trans-urethral resection of prostate (TURP) for a BPH; can present with evidence of metastasis without a discernible primary lesion in which case it is called occult; and it can be clinically evident and detectable by routine investigation.

1.3.4. Histologic Classification and Grading of Prostate Cancer

Grading and classification of disease enables clinicians to understand the severity or aggressiveness of disease and allows for international standardization

of nomenclature. In addition, it aids in the determination of the best treatment option for disease. The Gleason's grading is the most widely used grading system for PCa, and more recently, the World Health Organization (WHO) has developed an international histologic classification system for tumours of the prostate.

1.3.4.1. Gleason's Grading and Scoring System

Based on architectural features of hematoxylin and eosin (H&E) slides under the light microscope at low power magnification, Donald F. Gleason developed a grading system for PCa in 1966 using a small cohort of PCa patients; and this was further revised in 1974 and 1977 using a larger cohort [56]. In this grading, PCa samples fell into one of five grades (1-5). Increasing grade corresponded to severity of malignancy and degree of differentiation (**Fig. 1.8**).

Nine architectural patterns of PCa were merged into 5 major groups [57]. The Grade 1 is the best differentiated, while grade 5 is the worst differentiated histopathologically; and other grades fall in between these extremes. Gleason score is derived from this grading system by adding the Gleason grades of the most common and the second most common patterns found in a tumour sample.

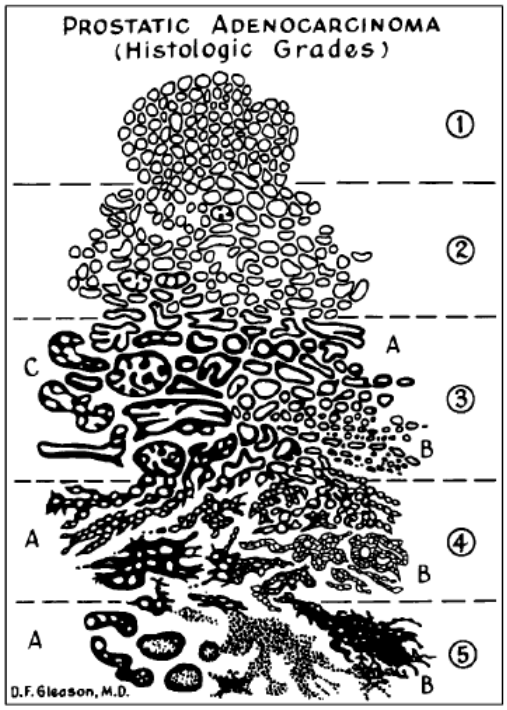
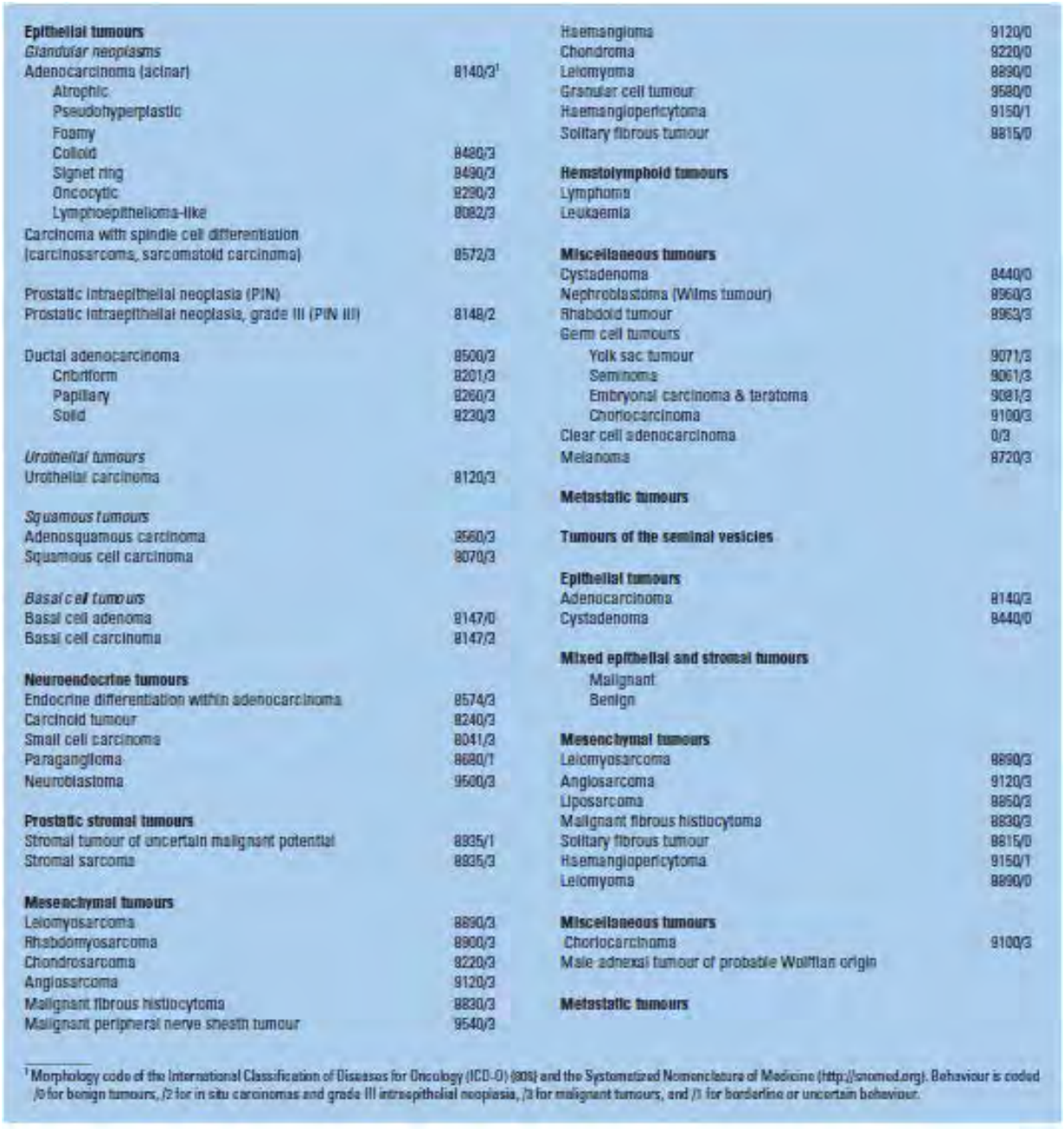


Figure 1.8. Gleason's grading system for prostate cancer (adapted from [57])

1.3.4.2. WHO Histologic Classification

In 2004, the WHO histologically classified tumours of the prostate to include various primary and metastatic tumours. Different classes of tumours were included as depicted in **Figure 1.9**.



Epithelial tumours	
Glandular neoplasms	
Adenocarcinoma (acinar)	8140/3[1]
Atrophic	
Pseudohyperplastic	
Foamy	
Colloid	8480/3
Signet ring	8490/3
Oncocytic	8290/3
Lymphoepithelioma-like	8082/3
Carcinoma with spindle cell differentiation (carcinosarcoma, sarcomatoid carcinoma)	8572/3
Prostatic intraepithelial neoplasia (PIN)	
Prostatic intraepithelial neoplasia, grade III (PIN III)	8148/2
Ductal adenocarcinoma	8500/3
Cribriform	8201/3
Papillary	8260/3
Solid	8230/3
Urothelial tumours	
Urothelial carcinoma	8120/3
Squamous tumours	
Adenosquamous carcinoma	8560/3
Squamous cell carcinoma	8070/3
Basal cell tumours	
Basal cell adenoma	8147/0
Basal cell carcinoma	8147/3
Neuroendocrine tumours	
Endocrine differentiation within adenocarcinoma	8574/3
Carcinoid tumour	8240/3
Small cell carcinoma	8041/3
Paraganglioma	8680/1
Neuroblastoma	9500/3
Prostatic stromal tumours	
Stromal tumour of uncertain malignant potential	8935/1
Stromal sarcoma	8935/3
Mesenchymal tumours	
Leiomyosarcoma	8890/3
Rhabdomyosarcoma	8900/3
Chondrosarcoma	8220/3
Angiosarcoma	9120/3
Malignant fibrous histiocytoma	8830/3
Malignant peripheral nerve sheath tumour	9540/3
Haemangioma	9120/0
Chondroma	8220/0
Leiomyoma	8890/0
Granular cell tumour	9580/0
Haemangiopericytoma	9150/1
Solitary fibrous tumour	8815/0
Hemolymphoid tumours	
Lymphoma	
Leukaemia	
Miscellaneous tumours	
Cystadenoma	8440/0
Nephroblastoma (Wilms tumour)	8960/3
Rhabdoid tumour	8963/3
Germ cell tumours	
Yolk sac tumour	9071/3
Seminoma	9061/3
Embryonal carcinoma & teratoma	9081/3
Choriocarcinoma	9100/3
Clear cell adenocarcinoma	0/3
Melanoma	8720/3
Metastatic tumours	
Tumours of the seminal vesicles	
Epithelial tumours	
Adenocarcinoma	8140/3
Cystadenoma	8440/0
Mixed epithelial and stromal tumours	
Malignant	
Benign	
Mesenchymal tumours	
Leiomyosarcoma	8890/3
Angiosarcoma	9120/3
Liposarcoma	8850/3
Malignant fibrous histiocytoma	8830/3
Solitary fibrous tumour	8815/0
Haemangiopericytoma	9150/1
Leiomyoma	8890/0
Miscellaneous tumours	
Choriocarcinoma	9100/3
Male adnexal tumour of probable Wolffian origin	
Metastatic tumours	

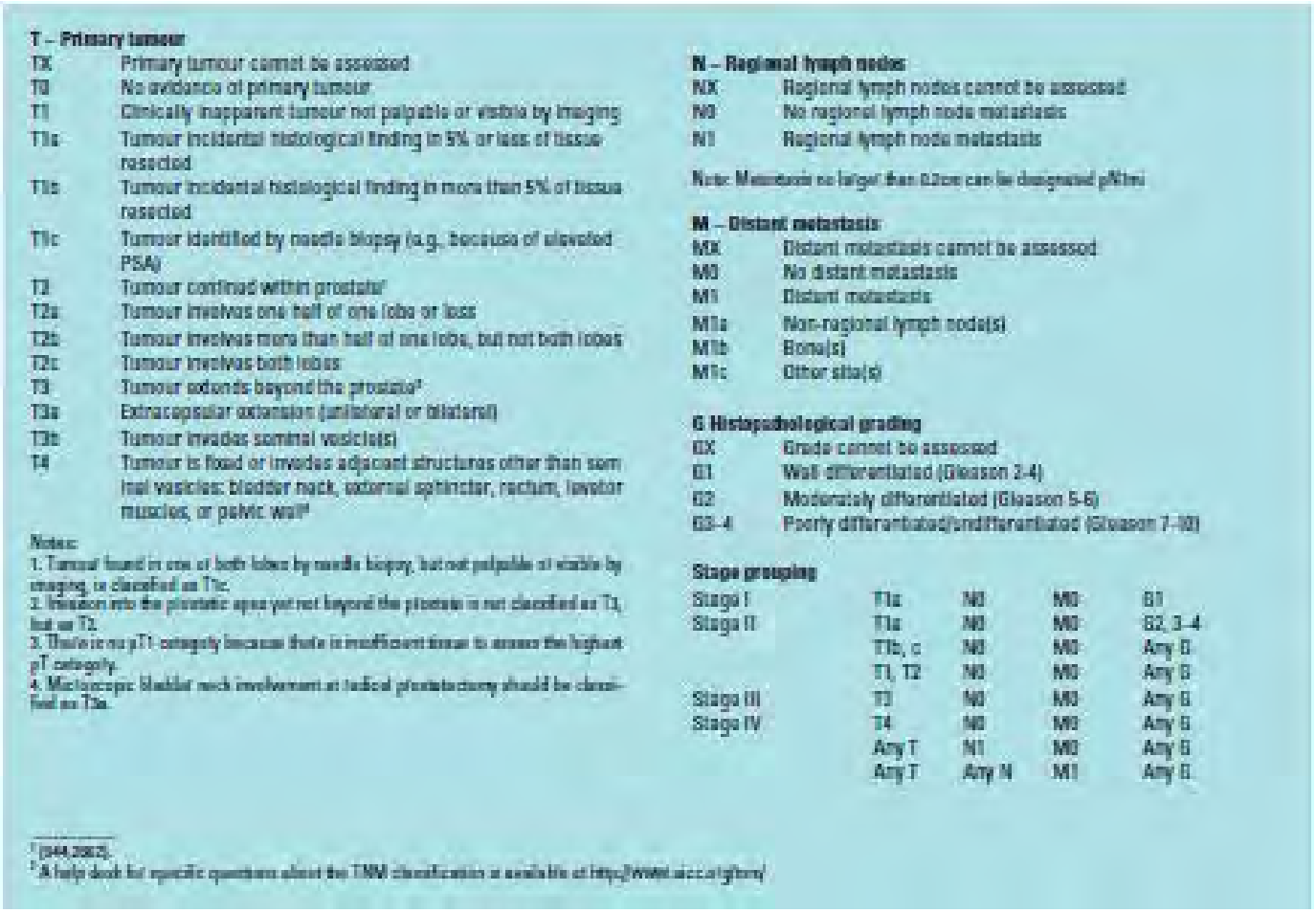
[1] Morphology code of the International Classification of Diseases for Oncology (ICD-O) (WHO) and the Systematized Nomenclature of Medicine (http://snomed.org). Behaviour is coded /0 for benign tumours, /2 for in situ carcinomas and grade III intraepithelial neoplasia, /3 for malignant tumours, and /1 for borderline or uncertain behaviour.

Figure 1.9. World Health Organization's histological classification of tumours of the prostate 2004 (www.iarc.fr)

1.3.5. Clinical Staging of Prostate Cancer

Beside the use of pathological staging/grading, clinical staging which is often employed by clinicians, can be used to predict the outcomes and to formulate the appropriate treatment for various stages of disease. Older methods of clinical staging such as Jewett and Whitmore scheme (a.k.a ABDC) is still in use by

some clinicians. Currently, the most common clinical staging system for PCa developed by the WHO is the TNM staging system. "T" stands for the primary tumour, while "N" stands for nodal involvement, and "M" stands for distant metastasis. A summary of the TNM classification is shown in **Figure 1.10.**



T – Primary tumour

TX	Primary tumour cannot be assessed
T0	No evidence of primary tumour
T1	Clinically inapparent tumour not palpable or visible by imaging
T1a	Tumour incidental histological finding in 5% or less of tissue resected
T1b	Tumour incidental histological finding in more than 5% of tissue resected
T1c	Tumour identified by needle biopsy (e.g. because of elevated PSA)
T2	Tumour confined within prostate[1]
T2a	Tumour involves one half of one lobe or less
T2b	Tumour involves more than half of one lobe, but not both lobes
T2c	Tumour involves both lobes
T3	Tumour extends beyond the prostate[2]
T3a	Extracapsular extension (unilateral or bilateral)
T3b	Tumour invades seminal vesicle(s)
T4	Tumour is fixed or invades adjacent structures other than seminal vesicles: bladder neck, external sphincter, rectum, levator muscles, or pelvic wall[3]

Notes:
1. Tumour found in one or both lobes by needle biopsy, but not palpable or visible by imaging, is classified as T1c.
2. Invasion into the prostatic apex (at not beyond the prostate) is not classified as T3, but as T2.
3. There is no pT1 category because there is insufficient tissue to assess the highest pT category.
4. Microscopic bladder neck involvement at radical prostatectomy should be classified as T3a.

N – Regional lymph nodes

NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Regional lymph node metastasis

Note: Metastasis no larger than 0.2cm can be designated pN1mi

M – Distant metastasis

MX	Distant metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastasis
M1a	Non-regional lymph node(s)
M1b	Bone(s)
M1c	Other site(s)

G Histopathological grading

GX	Grade cannot be assessed
G1	Well differentiated (Gleason 2-4)
G2	Moderately differentiated (Gleason 5-6)
G3-4	Poorly differentiated/undifferentiated (Gleason 7-10)

Stage grouping

Stage I	T1a	N0	M0	G1
Stage II	T1a	N0	M0	G2, 3-4
	T1b, c	N0	M0	Any G
	T1, T2	N0	M0	Any G
Stage III	T3	N0	M0	Any G
Stage IV	T4	N0	M0	Any G
	Any T	N1	M0	Any G
	Any T	Any N	M1	Any G

[1] (UICC, 2002).
[2] A help desk for specific questions about the TNM classification is available at http://www.uicc.ch/tnm/

Figure 1.10. TNM staging of prostate cancer (www.iarc.fr)

As shown, various combinations of the T, N and M components have been used to classify PCa into clinical stages I to IV. The American Joint Committee on Cancer (AJCC) has revised the TNM staging, putting into consideration the Jewett-Whitmore staging, Gleason score and Prostate-Specific Antigen (PSA) level (**Table. 1.2**). This classification was advanced to create a multiparametric improvement to staging of men with PCa.

	AJCC Prostate Cancer Stage Groupings													
	Stage I		Stage IIa					Stage IIb			III	Stage IV		
Jewett-Whitmore stage	A1		A2, B0-2								C1-3	D1		D2
TNM stage	T1a-c N0M0	T2a N0M0	T1a-c N0M0	T1a-c N0M0	T2a N0M0	T2a N0M0	T2b N0M0	T2c N0M0	T1-2 N0M0	T1-2 N0M0	T3a-b N0M0	T4 N0M0	Any T N1M0	Any T Any N M1
Gleason score	≤ 6	≤ 6	7	≤ 6	≤ 6	7	≤ 7	Any	Any	≥ 8	Any	Any	Any	Any
PSA level (ng/ml)	< 10	< 10	< 20	10-19.9	10-19.9	< 20	< 20	Any	≥ 20	Any	Any	Any	Any	Any

Table 1.2. AJCC staging of Prostate Cancer (http://prostatecancerinfolink.net/)

1.3.6. Diagnosis of Prostate Cancer

An accurate diagnosis of PCa depends on a combination of good clinical history taking and appropriate investigations. The major clinical investigations that have aided the diagnosis of PCa tremendously are digital rectal examination (DRE), trans-rectal ultrasound (TRUS), biopsy, and measurement of blood levels of PSA. A DRE is done routinely as part of medical screening in men. The physician inserts a lubricated finger through anus of the patient being examined to palpate the posterior lobe of the prostate for lumps, hardness or any irregularities (**Fig. 1.11**).

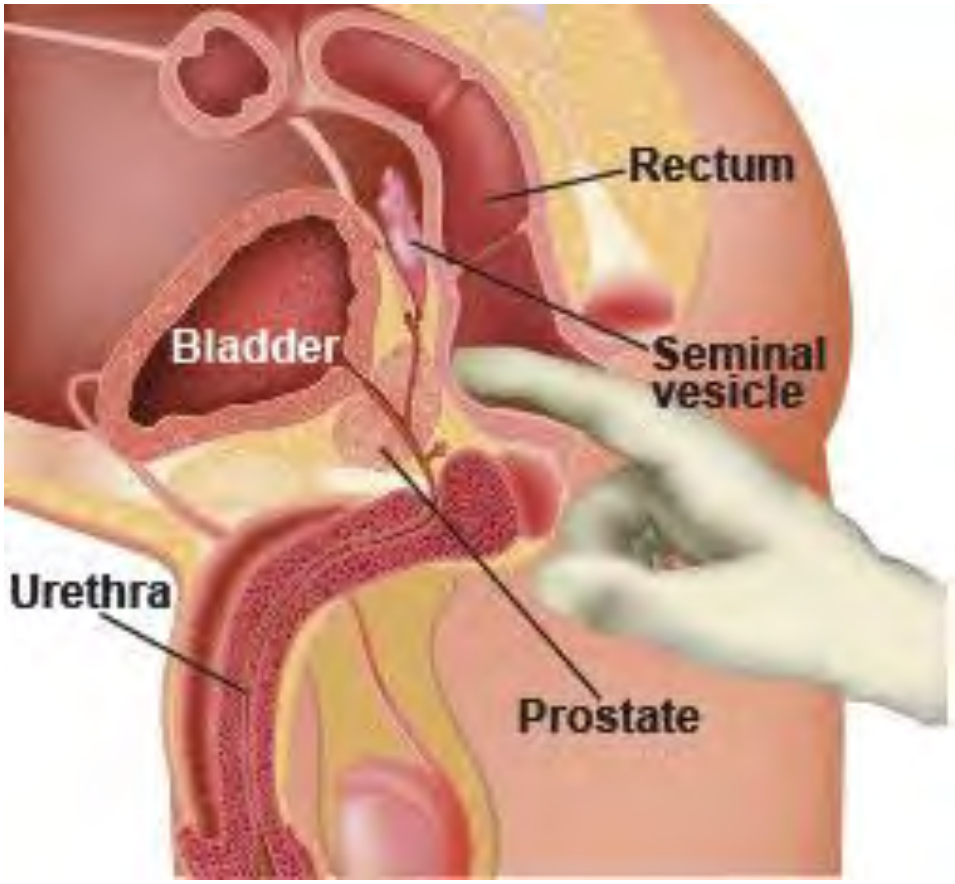


Figure 1.11. Digital rectal examination of the prostate gland (http://www.prostatespecialist.co.uk/)

TRUS uses an ultrasound probe to examine the PG and adjacent tissue tissues through the wall of the patient's anus (**Fig. 1.12**). Both procedures involve insertion of either the lubricated finger or the ultrasound transducer into the anus of the patient. Biopsy procedures are carried out either trans-rectally, through the urethra, or the perineum (under ultrasound guidance); although trans-rectal route is the commonest. Needle cores are used to take multiple samples from different aspects of the PG for histopathologic diagnosis. Between 8 and 10 needle cores are recommended for adequate sampling of the PG. Biopsy procedures can be complicated by rectal bleeding, hematuria, prostatitis, hematospermia and epididymitis [58]. Serum PSA level measurement has been an important adjunctive biochemical test for the diagnosis of PCa. Generally, levels of PSA exceeding 4.0ng/mL may indicate the presence of PCa in a patient [59].

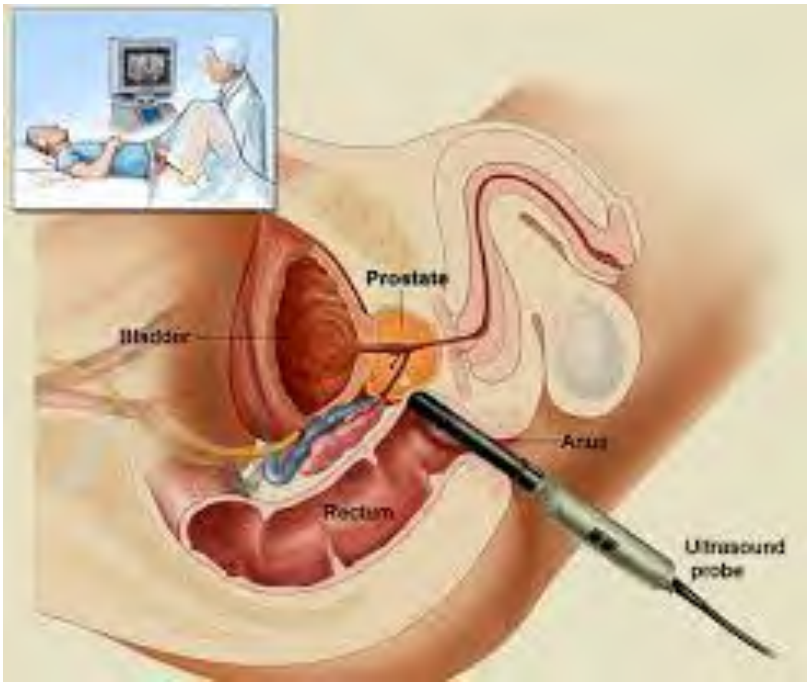


Figure 1.12 Trans-rectal ultrasound for prostate examination (http://www.eastcoasturology.com.au/)

1.3.7. Treatment of Prostate Cancer

Low risk (typically Gleason score 2-6, PSA level <10.0) or clinically localized PCa treatment options may be palliative or curative. Palliative measure may include watchful waiting and hormonal therapy; while curative measure includes active surveillance, radical prostatectomy and radical radiotherapy. Radical radiotherapy may include interstitial brachytherapy, intensity-modulate radiation therapy (IMRT), external beam therapy, proton beam therapy and 3D conformal radiation therapy. Active surveillance involves repeated biopsy and monitoring of PSA level particularly in PCA patients with a short PSA doubling time [60, 61]. PSA doubling time is the number of months it would take for PSA level to increase by two-fold [60]; and PSA doubling time of < 12 months is regarded as high risk [62]. Even though often used interchangeably with watchful waiting which does not involve frequent PSA testing; the goal of active surveillance is curative.

Intermediate risk PCa (typically Gleason score 7, PSA level 10-20) is clinically localized PCa with extensive involvement of greater than half of one lobe of the prostate gland, or even bilaterally without extraprostatic or seminal vesicle spread. In these cases, dosage of radiation and hormonal deprivation therapy (which may involve surgical castration/ orchiectomy or chemotherapy) is often escalated, and combination therapy is sometimes employed. Intermediate grade PCa can also be treated surgically by radical prostatectomy with pelvic lymph node dissection.

High risk PCa (typically Gleason score 8-10, PSA level >20), which involves extensive prostatic with extracapsular and seminal vesicle involvement is treated by a combination of radiotherapy and androgen deprivation therapy. High risk PCa can also be surgically treated with radical prostatectomy provided there is no fixation of tumour. Other methods of PCa treatment include cryosurgery [63], immunotherapy/vaccine therapy [64], and bone directed therapy [65]. An overview of the treatment of various stages of PCa is provided below (**Table. 1.3**).

Progression of PCa is characterized by changing from an androgen dependent PCa to an androgen refractory or castration resistant phenotype [66]. A few drugs currently approved for the treatment of metastatic castration resistant PCa includes: Docetaxel a taxoid mitotic inhibitor; Abiraterone an androgen synthesis inhibitor; Carbazitaxel a microtubule inhibitor; and Sipuleuce-T, an autologous immunotherapeutic agent raised against prostatic acid phosphatase (PAP).

Emerging drugs in phase III clinical trials for PCa treatment include: MDV3100 (enzalutamide) which is an androgen receptor antagonist; Orteronel (TAK-700) which blocks androgen synthesis by inhibiting the 17,20-lyase enzyme; Dasatinib (srcKI) an Src family kinase (SFK) inhibitor; clusterin chaperone protein inhibitor (Custirsen); and anti-angiogenic agent Bevacizumab [67, 68].

Medscape® www.medscape.com

Localized (T1, T2)	▪ Observation/watchful waiting ▪ Surgery: Radical retropubic Radical perineal Laparoscopic Da Vinci robotic-assisted ▪ Radiation therapy: Brachytherapy ▪ External beam radiation therapy (EBRT) combination ▪ Cryosurgery
Locally advanced	▪ Surgery ▪ Radiation therapy ▪ Combination therapy ▪ Clinical trials (hormones + EBRT) Chemotherapy + surgery Surgery + chemotherapy Surgery + EBRT Hormones + EBRT
Metastatic	▪ Surgical castration (bilateral orchiectomy) ▪ Medical castration with luteinizing hormone releasing hormone (LHRH) agonist with or without anti-androgen medication
Rising PSA following definitive treatment (surgery or radiation therapy)	▪ Observation ▪ Radiation to prostatic fossa (after surgery) ▪ Medical or surgical castration ▪ Clinical trial
Hormone resistant	▪ Continue LHRH agonist ▪ Second-line hormone therapy ▪ Clinical trial

Adapted from Daw & Peereboom (2001); Presti (2004); Zippe & Kedia (2000).

Source: Urol Nurs © 2004 Society of Urologic Nurses and Associates

Table 1.3. Overview of PCa treatment (www.medscape.com)

1.4. Biomarkers and Prostate Cancer

Biomarkers are measurable biologic, chemical, or physical characteristics or objective indicators of a normal biologic state, progress of disease, or response to drug therapy [69]. They can be nucleic acids, proteins or metabolites measurable in blood or other body tissues. An ideal biomarker should be safe, highly specific, sensitive, accurate in its detection ability, easily measurable and consistent across ethnicity and gender. The currently reported biomarkers of PCa

are very limited in terms of test characteristics; and in general there is a paucity of good biomarkers for early detection or prognosis of PCa. Even though the classification of biomarkers is currently under ongoing debate, cancer biomarkers can be classified into the following domains; diagnostic, prognostic, predictive, chemoprevention, screening, risk stratification, and post-treatment surveillance [70].

Pharmacologic classes of cancer biomarkers include; target verification, pharmacodynamics assay, treatment selection for clinical trial, surrogate end point in drug approval and early compound selection [70]. Several biomarkers have been used for detection and treatment monitoring of PCa. A few previously described biomarkers of PCa includes: PTEN, PI3K, PCA7 gene panel, PSGR, MME, PSCA, PCA3, TMPRSS2-ERG gene fusion, CD98, EPCA, CD276, prostate-specific membrane antigen (PSMA), caveolin-1, EN-1, and annexin A3 [71]. Some traditional, and a few emerging biomarkers of PCa are briefly described below:

Acid Phosphatase (ACPP): This is one of the oldest biomarkers used for PCa diagnosis in serum, however the drawback of this biomarker is that ACPP is expressed by both normal and malignant prostatic tissues, as well as extraprostatic tissues [72].

Prostate-Specific Antigen (PSA): Also known as kallikrein 3 (KLK3), PSA is the most widely used biomarker of PCa and has tremendously improved the diagnosis of PCa. However, recent evidence indicates that PSA falls short in its diagnostic ability in the lower reference ranges (2-10 ng/mL). Even though highly sensitive, it is not so specific and has led to a lot of false positives, false negatives and overtreatment of PCa patients. To improve the diagnostic ability of PSA, several related parameters such as PSA doubling time, PSA velocity, free-to-total PSA ratio and prostate health index (PHI) have been explored [73].

Prostate cancer antigen 3 (PCA3): Otherwise known as DD3 or differential display clone 3, PCA3 is a noncoding mRNA which is found to be highly abundant in malignant prostatic tissues in comparison to benign [74]. It is the most widely used non-PSA based biomarker for prostate cancer diagnosis [75].

The PCA3 assays was developed to detect the urinary level of RNA transcripts which are contained in cells shed into the urine during micturition [76]. One limitation of this test is that it is dependent on the urinary PSA transcript expression level.

TMPRSS2-ERG gene fusion: Transmembrane Protease, Serine 2– ETS fusion genes are members of the ETS family of genes, noted to be highly expressed in malignant prostatic tissue and not expressed in benign prostatic tissues. It was hypothesized to be a good indicator for PCa aggressiveness, however its expression in only about 50% of all PCa patients in Western studies [77], makes it a better marker of disease heterogeneity rather than a predictive or prognostic marker. It suffers the same drawback as PCA3 in that it depends on urinary PSA transcript levels for meaningful interpretation of result.

Alpha methylacyl-CoA racemase (AMACR): Is a highly sensitive and specific diagnostic biomarker often used in PCa tissues. Low levels of AMACR in biopsy tissues have been associated with biochemical recurrence and PCa metastasis [78].

1.5. Personalized/ Omics-based Approaches in Prostate Cancer

Omics based approaches seek to use high-throughput techniques to answer biologic questions in a systems-oriented manner. It interrogates the full complement of biomolecules in a cell, tissue or body fluid and examines their interaction rather than in isolation. Examples of omics based approaches include genomics, proteomics, metabolomics, lipidomics, epigenomics and pharmacogenomics, *inter alia*. Many of these approaches have been used to develop useful biomarkers for PCa; and it is expected that biomarkers generated from omics based approaches would benefit personalized or individualized diagnosis and treatment of PCa in no small measure.

For example, a recently described classification of the seven possible genotypes of prostate cancer [79] based on TMPRSS2:ERG translocations (**Fig 1.13**) may

represent a more useful molecular classification of PCa (in terms of therapeutic options) than histologic classification.

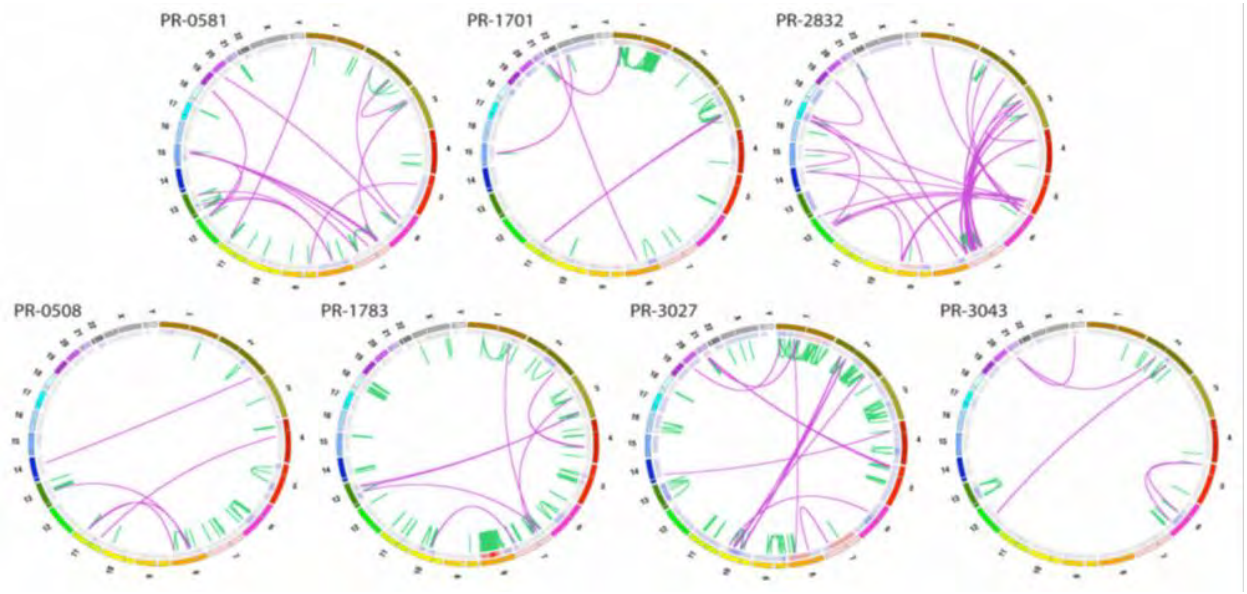


Figure 1.13. Graphical circos plot representation of seven PCa genomes based on TMPRSS2-ERG translocation. The outer ring and inner rings depict the genomic location and chromosomal copy number respectively. In the inner ring, red and blue indicates copy number gain and loss respectively. Interchromosomal translocations are represented in purple, while intrachromosomal rearrangements are shown in green. This picture shows genomic organization based on the absence (bottom row) and presence (top row) of the TMPRSS2-ERG fusion genes (adapted from [79]).

1.6. Proteomics and Prostate Cancer Biomarkers

The field of proteomics encompasses omics-based high throughput techniques for large-scale identification of the full complement of proteins in an organism, tissue, cells or body fluid. The field is also potentially able to investigate the functional states of proteins including, post-translational modifications, protein-protein interactions; and protein interaction with other biomolecules such as carbohydrates, lipids and other metabolites. Proteomics can provide insight into protein structures, alternative splicing events, as well as aiding genome annotation. It is important to note, that there can be variations to the proteome of a cell depending on the time point, stage of disease, diet and a host of other factors. Currently, proteomics has been employed to identify clear, reproducible cancer-related signatures for early detection, diagnosis, or prognosis between

disease and healthy cohorts of patient [80-85]. However, the lack of standardization of validation methods among researchers; as well as non-concordant experimental outcomes from independent research groups [75], has underscored the urgent need for systematic validation of proteomics biomarkers [86-88]. Notably, the most common proteomics methodologies are mass spectrometry-based proteomics and protein microarray technology based proteomics. Using these methodologies, a gamut of proteomics biomarkers of PCa have already been identified [89-95], and some were demonstrated to potentially predict progression and aggressiveness of prostate cancer [96-98]. However, successful application of omics based approaches is heavily dependent on available bioinformatics and computational biology resources.

1.7. Liquid Biopsy in Prostate cancer diagnosis

Tissue biopsy and surgical procedures are invasive and can potentially be accompanied by various complications. To minimize these complications, liquid/fluid biopsy technology has become an emerging hot spot in cancer research; because it is a less invasive and can isolate biomarkers from body fluids such as urine, saliva, ascites, blood, semen, cerebrospinal fluid (CSF) and pleural effusion [99]. Examples of biomarkers that can be isolated by the liquid biopsy approach include: proteins, microRNA, circulating tumour nucleic acid (ctNA), circulating tumour cells (CTC), exosomes, and long noncoding RNA (lncRNA) [100-102].

This method is highly beneficial for personalized treatment of patients, and provides access to important information, even with a small amount of test sample. It also permits frequent monitoring of treatment response without the added cost of repeated surgical procedures. An important caveat is that liquid biopsy requires sophisticated equipment such as ultracentrifuges, DNA-sequencing machines, and diagnostic kits. Not least, results generated are sometimes not reproducible between different research groups as a result of tumour heterogeneity [103-106]. Liquid biopsy has been used for diagnosis of

various cancer types [107-111], as well as PCa [112, 113]. Additionally, this technique has been found useful in cancer patient stratification, monitoring and screening [114].

1.8. Urine as a Biomarker Source

As an ultrafiltrate of blood, urine possesses analogous protein profiles with blood and provides a usable catalog of proteins for interpretation of pathophysiologic events in the human body [115]. Sampling urine, as compared to blood or prostatic tissue biopsy is a less invasive approach for PCa diagnosis and treatment monitoring; and the service of a skilled personnel is not required for urine sample collection. In addition to this, urine is abundantly available and permits repeated sampling [116], unlike the blood which has a fixed total body volume of approximately five liters [117].

Using this approach, the risk of blood-borne infections like human immunodeficiency virus (HIV) and hepatitis are minimal. Notably, the urinary proteome is less complex compared with the blood proteome; and not least, there is very little or no possibilities of proteolysis since urinary proteins are quite stable, having undergone all the possible proteolysis during storage in the bladder [116]. Hence there is little or no benefit of using protease inhibitors during storage [116]. Despite the fact that urine is a very promising disease biomarker source, a caveat is that its composition is variable depending on the time of the day, dietary intake and the state of health of the individual. Many urinary biomarkers of various diseases have been previously described [118], albeit only a few of these biomarkers make it to clinical use.

1.9. Blood as a Biomarker Source

The human blood is the most vital body fluid that is involved in the transport of oxygen and essential nutrients to different organs, tissues and cells as well as transporting metabolic waste away from them. Considering the extensive contact

of the blood with body structures, it is an attractive source of biomarker discovery. In addition, the increased discohesive nature of cancerous tissues sometimes makes tumour cells more mobile and allows them to be transported in the blood stream as tumour markers. Even though limited in volume, it is the most frequently used sample for clinical diagnosis of many disease conditions.
Blood samples have been used to identify biomarkers of many human diseases including: Alzheimer's disease [119], Parkinson's disease [120], breast cancer [121], preeclampsia [122], and prostate cancer [74]. Moreover, it is also known that tumour-associated autoantigens (TAA) from cancerous conditions are able to stimulate the humoral immune response leading to the generation of auto-antibodies [123]; and this has been demonstrated to drive cancer-specific autoimmunity in different human cancers including renal [124], colorectal [125], lung [126], and prostate cancer [127-131]. These autoantibodies have been found to have potential theranostic utilities for cancer diagnosis, immunotherapy and therapeutic vaccine target development.

1.10. Historical Perspectives on Prostate Cancer in Men of African Descent

The currently accepted hypothesis is that aggressive phenotypes of PCa are more common in men of African descent; albeit underlying factors for such aggressiveness are poorly understood. Beyond various suggested sociocultural issues such as poor funding [132], insufficient manpower and skilled health personnel [133], poor access to healthcare [134], religious and cultural beliefs [135], lack of well-updated cancer registries [136], poor research and healthcare infrastructure [137], low educational level [138], prevalence of infection [139, 140], poor governance structures and fiscal policies [141]; the underlying genetic, hereditary and environmental basis of PCa aggressiveness in men of African descent still warrants further research.

Migration patterns of people in Africa may provide a clue to the high burden and aggressiveness of PCa in this population. Historically, humans appear to have lived longest on the African continent and preserved fossils of early men discovered at the "cradle of humankind" which is a

world heritage site in South Africa [142-145], alludes to that theory. Ever since the human hunter-gatherer populations in the late Pleistocene epoch were replaced by the human farmer or agro-pastoralist populations in the Holocene epoch [146, 147], mankind continues to evolve genetically, based on environmental exposure and necessity for survival. Unlike the nomadic hunter-gatherers who had low resistance to infections due to poor exposure, farmers had the advantage of developing immunity to disease breakouts and epidemics in their crowded farming settlements [147].

Hence, a deeper consideration of the complex ancestry of the culturally and linguistically diverse African male population would improve our understanding of disease pathogenesis in this population. Albeit, it appears plausible that a complex interplay *inter alia,* of genetics, infection and environmental factors underlies the pathogenesis and aggressive phenotypes of PCa found in men of African descent, resulting in a disproportionately high PCa burden in sub-Saharan Africa [148]. Western studies have revealed that African-Americans have up to 20% of their genetic makeup mixed with Caucasian ancestry [149]; however, literature on the ancestral characterization of the highly heterogeneous indigenous African populations is scarce. Considering that even within sub-Saharan African populations, disparities exist in incidence and mortality of PCa in different countries [150].

Many of the indigenous Africans who presently live in South Africa have been predicted to originate from people of Southern Bantu ancestry who migrated downward from the original Bantu Homeland [151] which is approximately between eastern Nigeria and western Cameroon [152]. Various genetic and phenotypic events have been attributed to the Bantu expansion [153]. Even though hunter-gatherers were lighter in skin tone, have child-like bodies and were often monogamous in lifestyle; agro-pastoralist tended to be polygynous [154]. In fact, polygamy rate as high as 20-50% has been reported in marriages across sub-Saharan Africa [155]. In such agro-pastoralist societies, intense competition for female partners could result in Darwinian style sexual selection [156] in which females would prefer stronger, bigger and more masculine males. This may explain why men of African descent are often quite robust anthropometrically [157-160]; albeit there are other sexually selected attributes that makes a mate attractive in humans such as intelligence [161-163], hair color [164], personality [165], immune function [166], natural eye gaze [167], and facial appearance [168] , *inter alia*.

Hormonal influences are hypothesized to play a part in male body morphology and behavioral patterns [169-171]. Hunter-gatherer descendants were observed to have lower levels of androgen hormones dihydrotestosterone (DHT) and total testosterone as compared to agro-pastoralist descendants in a study carried out in the Southern African region [172]. Besides evidence from Africa, it has been shown that the Ache hunter-gatherer men from Paraguay have a lower salivary testosterone level in comparison to western values [173]. Also, the Tsimane hunter-gatherer populations from lowland Bolivia were shown to demonstrate a spike in their traditionally low testosterone level as a result of the introduction of horticultural tasks into their daily routine [174]. Evolutionary population variations in testosterone level and PCa disparities has been well discussed in a synthesis by Alvarado *et al* [175] . Young men of African descent have been reported to have a higher level of testosterone than their Caucasian counterparts [176] and Asian men have been reported to have a low level of 5α-reductase activity [177]. Also, variations have been reported in androgen receptor activity between Africans, Caucasians and Asians [178]. (A comprehensive table of androgen level racial variation is provided in **Annexure V**) These may suggest some of the possible reasons for high incidence and mortality of PCa among men of African descent worldwide. Many questions around PCa remain unanswered. For example, it is difficult to tell with any currently available diagnostic aid/test if PCa would be hormone refractory, aggressive or metastatic [179].

1.11. Biorepository development in Africa

Most ground breaking projects in the field of molecular biology such as: the Human Genome Project (HGP) [180]; The Cancer Genome Atlas (TCGA) [181]; Human Proteome Project (HPP) [182]; and Chromosome Centric Human Proteome Project (CHPP) [183]; have benefitted immensely from specimen biorepositories. The National Cancer Institute has described a biorepository as human specimen collection including relevant data for the purpose of research, and subject to relevant processes, ethics and policies [184, 185]. There are various standards that must be met in order to create a good biorepository. These include: proper annotation of specimens (including clinical information); ensuring ethical conduct and patient confidentiality; regulated centralized access to samples; and data-driven, *in-silico* research enablement of information generated from samples. Despite the gamut of biorepositories established in the Americas, Europe, Asia and Australia [186-188], very few such biobanks have been established in sub-Saharan Africa [189].

Most biorepositories in Africa are established within an investigator's research group and most are yet to be standardized and centralized. There is a growing need for collaborative efforts amongst researchers in Africa to smoothen the task of standard biorepository establishment. Several challenges have plagued biorepository development and regulation in sub-Saharan Africa [189-192]; albeit modest progress has been made in a few sub-Saharan African countries. The emergence of the H3Africa consortium, has improved the centralization and standardization of biospecimen collection in Africa [192]. Standardization of sample annotation in various African studies would permit researchers across Africa to collaborate.

1.12. Statement of problem

Low and Middle Income Countries (LMICs) receive only around 5% of the global spending on cancer, despite carrying up to 80% of the Global cancer burden [132]. The burden of PCa is on the increase both globally and particularly in South Africa. According to the International Agency for Cancer Research on Cancer (IARC) in 2012, there were 1.1 million and 9,957 new cases of PCa reported globally and in South Africa respectively; making PCa the most diagnosed cancer in South Africa. During this period, there were 307,000 and 3,539 PCa deaths reported worldwide and in South Africa respectively; making PCa the second leading cause of cancer deaths in South Africa.

Estimated age-standardized incidence and mortality rates per 100,000 in all ages is high in South Africa as compared to the rest of Africa. Men of African descent have been known to carry aggressive phenotypes [193-195] of the disease and factors responsible for this racial disparity in PCa are poorly understood. In addition, there is significant under-reporting of PCa cases due to background sociocultural issues within the South African population. Not least, it is unclear what factors underpin the progression of PCa, i.e. "which PCa lesions will remain indolent, and which ones would evolve to a more aggressive phenotype?" Factors that make androgen dependent tumours transform into androgen resistant phenotypes are poorly understood. In addition, it is unclear why metastatic phenotypes of PCa have a predilection for osseous tissues in comparison to other tissues [179]. Use of PSA for diagnosis of PCa can be fraught with false negative, false positives and overtreatment especially in the reference range of 2-10ng/ml.

It is unknown if there is a difference between the proteomes of the whites, blacks and mixed ancestry PCa patients in South Africa and putative urinary biomarkers are yet to be validated in any African cohort of PCa patients. Thus, further research is required to elucidate the biology and characteristics of PCa, particularly within a heterogeneous cohort of Southern African PCa patients; and

novel diagnostic and therapeutic biomarkers of PCa urgently need to be developed.

1.13. Study Hypotheses

- Based on proximity, PCa cells may secrete proteins into the genitourinary tract which may be detected in urine.
- There may be a variation in the humoral autoantibody response generated to PCa by different ethnicities within the South African population.
- There may be subtle differences in the proteomics profile of the different racial groups within the South African PCa patient cohort based on environmental exposure.

1.14. Project Aim

"Urinary and serological biomarker discovery and prevalidation for prostate cancer biomarkers in South African patients using proteomics techniques"

1.15. Specific objectives

- Analyse the urinary proteome of South African patients for PCa biomarker discovery.
  - Use mass spectrometry based methods to investigate the urinary proteomes of PCa patients in South Africa.
  - Identify potential urinary biomarkers of prostate cancer in our South African cohort.
  - Compare the urinary proteomics profile of the different racial groups in South Africa within our cohort.
  - Use in-silico methods to verify identified candidate biomarkers and prevalidate using targeted parallel reaction monitoring experiments.

- Analyse the serum autoantibody repertoire of South African patients for PCa biomarker discovery.
  - Use a novel cancer-testis antigen protein microarray platform based methods to investigate the blood proteomes of PCa patients in South Africa.
  - Identify potential serological biomarkers of prostate cancer in our South African cohort.
  - Compare the serological proteomics profile of the different racial groups in South Africa within our cohort.
  - Identify functional pathways involved in autoantibody response to PCa.

CHAPTER 2: GENERAL MATERIALS AND METHODS

2.1. Prostate Cancer Biorepository Development

The study was carried out in collaboration with the Urology Department at the Grootes Schuur Hospital (GSH), a tertiary teaching hospital connected to the University of Cape Town Medical School. GSH urology unit operates in two satellite hospitals which are the New Sommerset Hospital (NSH) and the Eerste Rivier Hospital (ERH). One of the major goals of the project is to develop for research purposes; a South African cohort of patients biodata and biospecimen, drawn from a wide background of patients including indigenous Africans, Caucasians and Mixed Ancestries.

2.1.1. Study Staff

Being a multi-departmental study, staff in all units involved in the study participated in the development of this biorepository. Academic staff included specialists and lecturers in Urology, Anatomic Pathology and Medical Biochemistry; as well as registrars in Urology, post-and pre doctoral research fellows at the International Centre for Genetic Engineering and Biotechnology (ICGEB). Non-academic staff included Nurses at GSH, Research Technicians at ICGEB, and Medical Records (MR) staff. The whole project was conducted under the coordination and oversight of the principal investigator (PI) and Group Leader of the Cancer Genomics Group (CGG) at ICGEB.

2.1.2. Cohort Recruitment Workflow

A high level of coordination and communication between study staff in the ICGEB laboratory and the hospital is needed. Phone calls, text messages, social media messaging and emails were used as means of communication between the urology team, medical record team and the CGC during weekly schedules. A summarized overview of the patient recruitment and sample storage process is presented below (**Figure 2.1)**. Regular meetings were held between the leaders and members of the different teams on how to improve the cohort size, patient recruitment and sample collection process. Periodic reorientation meetings were held with the academic and non-

academic staff to update them on the purpose of the study. Patient information was kept confidential and encrypted throughout the duration of the study.

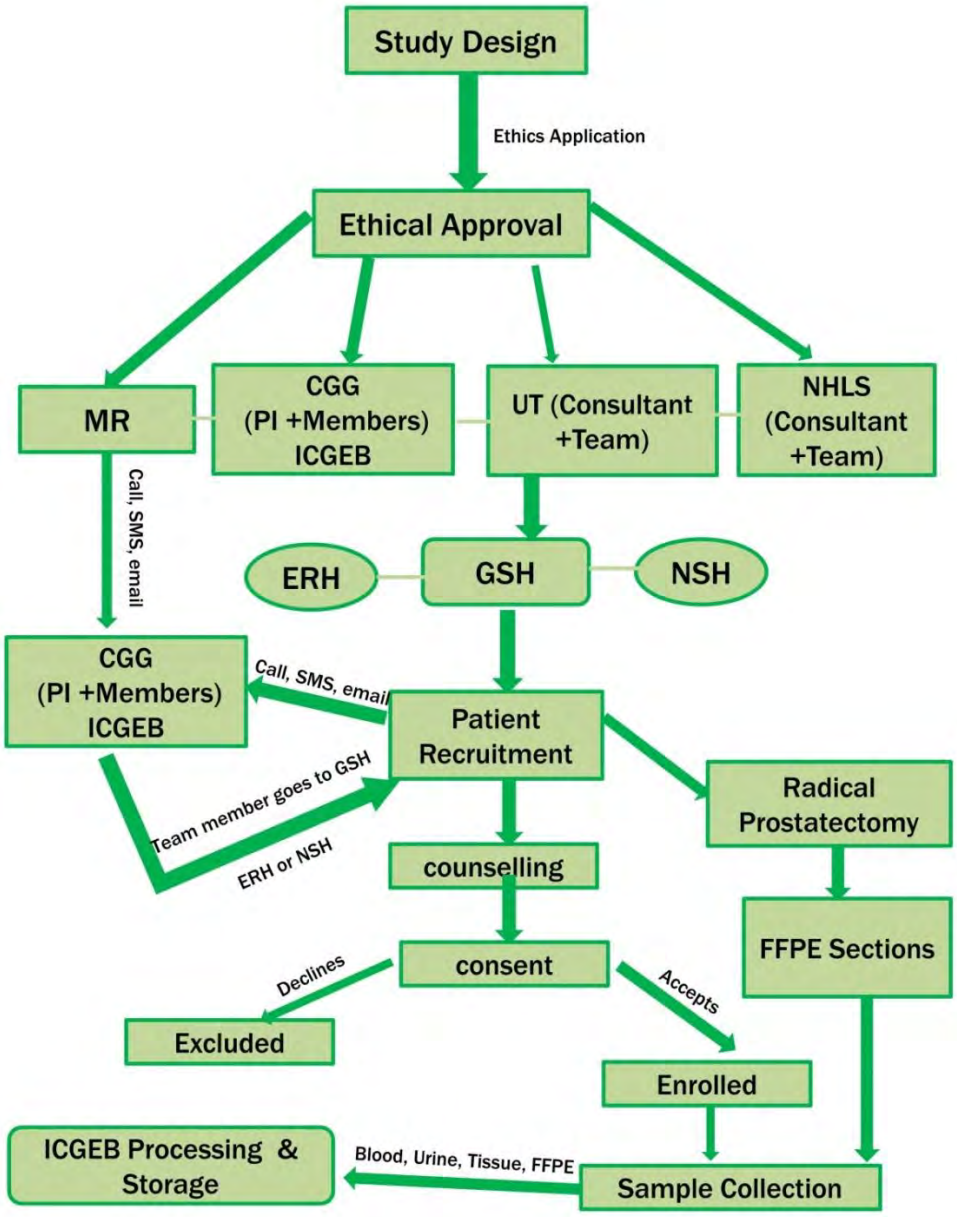


Figure 2.1. Overview of the patient recruitment and sample storage process in the study

2.1.3. Study Protocol and Ethical Consent

Study protocol clearly stated the purpose and objectives of the study as well as criteria for patient selection. The protocol describes the method of participant selection and how their safety and confidentiality would be ensured throughout the duration of the study. Detailed explanation of how and for how long samples would be collected and stored is included in the protocol; as well as the potential benefit of the study. Study was carried out only after full ethical approval by the Health Research Ethics Committee of the Faculty of Health Sciences, University of Cape Town. The ethics approval number is ***#HREC 454/2012***. Study was carried out following standard ethical guidelines as embodied in the declaration of Helsinki [196-198]. Consent forms were made available in local languages like Xhosa, Zulu, Afrikaans and English.

2.1.4. Patient Selection

All patients were initially seen at Grootes Schuur Hospital before some were referred to the satellite hospitals for surgery. Consecutive patients scheduled to undergo Transurethral resection of prostate (TURP) or prostatectomy are approached for enrollment after counseling by the nursing staff. After full explanation of the content of the consent form, patients were allowed to ask questions and sign the consent form thereafter. Patients who were unwilling to participate after going through the consent form were excluded from the study. Inclusion criteria for PCa patients in this study were; patients who were being diagnosed with PCa for the first time, who (1) have not been previously treated for PCa, (2) have not had surgical orchiectomy or hormonal treatment (3) do not have any comorbidities such as arterial hypertension or diabetes, and (4) do not have history of any other cancer at any site in the body. Patients who decline to participate in the study for religious, cultural or personal reasons and those who do not have adequate clinical information were excluded from the study. Other patients recruited for the study were patents diagnosed with benign prostatic hyperplasia (BPH), and those diagnosed with other uropathies (DC) such as urethral stricture, kidney stones and prostatitis. In addition, urine samples were collected from age matched normal healthy individuals as control cohort (NC) after proper consent taking. Indigenous Africans, Mixed Ancestry patients and Caucasian African patients with PCa were enrolled in this study. Clinicopathologic information for each enrolled patient was

accessed through the National Health Laboratory Services (NHLS) database at the hospital. Weekly tasks were scheduled around preexisting clinical activities in the Urology unit; and a draft timetable of the weekly activities of the team is as described in the table below (**Table 2.1)**. Establishment of standardized guidelines for sample processing and data storage aided good annotation of the cohort. Also, access to the progress of the study and "opt out" option was provided for all participants without any penalties.

	8am-10am	**11am-12pm**	**1pm-2pm**	**3pm-4pm**	**Later than 4pm**
Monday	ICGEB calls registrar and medical records at GSH, ERH and NSH				Sample processing and storage at ICGEB
		ICGEB goes to GSH to consent patients			
	ICGEB goes to NSH to collect TURP/ Prostatectomy samples				
	ICGEB goes to ERH to collect TURP/ Prostatectomy samples				
	Sample processing and storage at ICGEB				
Tuesday	ICGEB goes to GSH to collect TURP/ Prostatectomy specimen				Sample processing and storage at ICGEB
	Sample processing and storage at ICGEB				
Wednesday	ICGEB goes to GSH to collect TURP/ Prostatectomy specimen				Sample processing and storage at ICGEB
	Sample processing and storage at ICGEB				
Thursday	Sample database update				Sample processing and storage at ICGEB
	Sample processing and storage at ICGEB				
Friday	ICGEB calls registrar and Interns at NSH/ GSH				Sample processing and storage at ICGEB
	ICGEB goes to GSH to consent ERH and GSH patients and collect urine and blood samples at GSH				
	ICGEB goes to NSH to consent patients and collect urine and blood samples				
	Sample processing and storage at ICGEB				

Table 2.1. Draft time table of patient recruitment, sample collection and sample storage in the study.

Upon notification on Fridays from the hospital that a suitable participant has been identified for treatment at ERH or NSH, delegated member of the CGG drive about 16 minutes down to NSH which is about 8.8 km away from GSH (**Figure 2.2**); to consent patients and collect blood and urine samples. On Mondays, tissue samples are collected from the NHS surgical theatre for these patients, by CGG members who drive down to NSH with a dry ice box.

Figure 2.2. Distance from GSH to NSH is about 8.8km and about 16 minutes of driving (www.google.co.za/maps)

Blood and urine samples from the ERH patients are usually collected on Friday at the GSH clinic. On Monday, to collect the surgical tissue specimen, the CGG member then drives down to ERH which is about 29.6 km from GSH and takes about 29 minutes to drive (**Figure 2.3**). Samples are collected and preserved in a dry ice box on the journey back to GSH. All collected samples are immediately stored at the recommended temperature after appropriate processing (See **section 2.2** below).

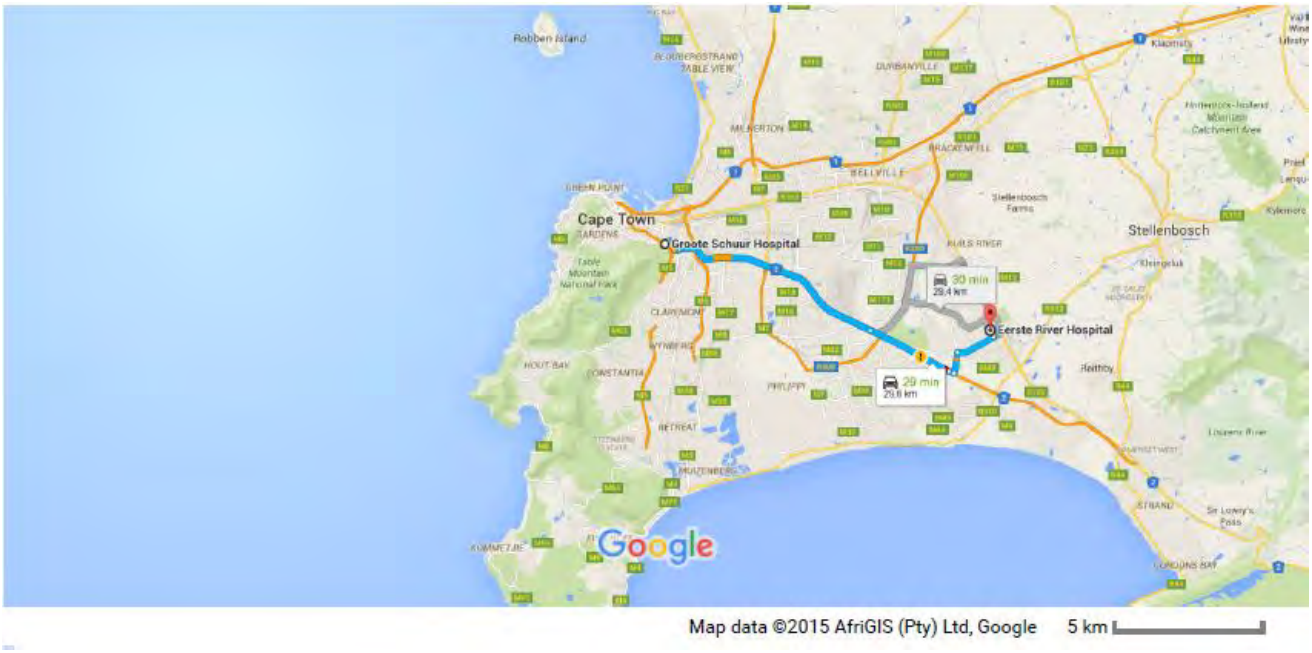


Figure 2.3. Distance from GSH to ERH is about 29.6km and about 29 minutes of driving (www.google.co.za/maps)

2.1.5. Cohort Characteristics, Demographics and Physical Samples

African, Caucasian, and Mixed Ancestry patients who met the inclusion criteria of this study were enrolled and their ages ranged between 30-80 years and the cohort is slightly dominated by mixed ancestry population. Only urine samples were collected from normal healthy individuals who were drawn from a healthy population. Patients that declined participation for personal, religious or cultural reasons were excluded. The physical samples that were generated for the biorepository were blood, urine, surgical tissue, and FFPE samples. Samples were processed as described below and quality checks were performed to ensure that samples were of adequate size and are in good condition. For example, blood, tissue and urine samples with inadequate volumes/size were excluded. Samples that were drawn before arrival of the CGG member or that were collected and not appropriately stored were excluded as well. Samples that were collected in the wrong sample bottles were not banked and collected samples were carefully annotated with clinical information and the database is backed-up and periodically updated. Patient information is kept confidential and encrypted in order to eliminate bias and to comply with

international standard of study data confidentiality. Samples with incomplete annotation were excluded as well.

2.2. Urine and Blood Sample Collection and Preparation

In this study, both urine and blood samples were collected from patients. Urine was collected by voiding into a 50 mL collection tube, or collected directly from a catheter bag. Collected urine samples were stored in an ice box and transferred rapidly after collection to a -80°C freezer facility at the ICGEB within the premises of the Faculty of Health Sciences, University of Cape Town. Blood samples were intravenously drawn and collected in a 5ml K_2-EDTA bottle and stored on ice before being transferred to a -80^{0}C freezer facility as described above. Following the collection of urine and blood samples, they are processed as described below before proceeding to the analytical phases of the study. Quality control checks and assay optimization were ensured at every phase of the sample preparation process. Any sample that does not meet the standard in terms of sample volume and quality were excluded from further processing.

2.2.1. Urine Sample Preparation

Urine samples were centrifuged at 2000×g for 30 minutes; to remove cellular debris and particle precipitates [199, 200]. The supernatant was collected and protein extraction and digestion was performed on it.

2.2.1.1. Protein Extraction

Other contents of urine, such as lipids, peptides, salts, mucopolysaccharides, and oligosaccharides may interfere with the electrophoretic mobility of urinary proteins [201-204]; hence there is the need for concentration and purification of urinary proteins before analysis. There are various methods for urinary protein extraction including ultrafiltration, and organic solvent precipitation using acetone, isopropanol, ethanol, methanol/chloroform, tricholoracetic acid (TCA), and occasionally, a combination of these methods. The protein extraction method used here was the ultrafiltration method where a series of molecular weight cut-off (MWCO)

filters were used to concentrate urine samples. Solutions and reagents required are described below in **Table 2.2**.

Solutions and Reagents	Preparation Description
10% SDS	**Add 2.5g of SDS + 25ml of water**
0.5M Tris-(2-Carboxyethyl)phosphine (TCEP) (Sigma-Aldrich, St. Louis, MO, USA)	**Use as prepared**
0.1M Tris HCl	**Add 0.78795g of Tris HCl + 50ml of water (pH 8.5)**
0.55M iodoacetamide (IAA) (Sigma-Aldrich, St. Louis, MO, USA)	**Add 40.7mg of iodoacetamide + 400ul of UA. Prepare immediately and store at in the dark at 5˚C.**
proteomics grade trypsin (Sigma-Aldrich, St. Louis, MO, USA)*	**2ug aliquot**
1M Tris HCl	**Add 7.8795g of Tris HCl + 50ml of water (pH 8.5)**
Lysis Solution	**Add 60ul of 10% SDS + 15ul 1M Tris HCl + 30ul of 0.5M TCEP + 45ul of water**
UA	**Add 24.024g of urea + 50ml of 0.1M Tris HCl**
50mM Ammonium bicarbonate buffer (ABC)(Sigma-Aldrich, St. Louis, MO, USA)	**Add 98.8mg of NH4HCO3 + 25ml of water. Store at 5˚C**

Table 2.2. List of Reagents and Solutions used for protein extraction and digestion (* the trypsin to protein ratio is 1:50)

Steps involved in urinary protein extraction

1. A 15 ml 50KDa MWCO filter is conditioned 3X with UA.
2. A 15 ml 3KDa MWCO filter is conditioned 3X with UA.
3. First run the samples through the 50KDa MWCO filter in a refrigerated centrifuge at 3500 x g to filter off high molecular weight proteins like uromodulin and human serum albumin.

4. Then concentrate the proteins to about 1mL by running the flow through from the 50KDa MWCO filter through a 3KDa MWCO filter.
5. For steps 3 and 4, you can top up with 15 ml of UA and re-spin.
6. Then denature the protein concentrate using the lysis solution (75µL lysis buffer/mg starting protein)
7. Incubate at 95^{o}C for 3 minutes
8. Pulsed indirect sonication was the performed for a minute at vibrational amplitude setting of 1-2 on a high intensity ultrasonic bath (Analab, Bischheim, France)
9. Spin at 3500 x g for 10 minutes.

2.2.1.2. Polyacrylamide Gel Electrophoresis

Gel electrophoresis is one of the most widely employed techniques in traditional biochemistry and molecular biology to separate and investigate proteins and nucleic acid using an electric field. Typically, agarose gels are used to separate nucleic acids while polyacrylamide gels are used for proteins; albeit polyacrylamide gels can also be used for shorter fragments of DNA. One dimensional (1-D) polyacrylamide gel electrophoresis separates proteins based on their molecular weight, while two-dimensional (2-D) polyacrylamide gels further separates proteins based on molecular weight and isoelectric pH gradient. One dimensional Laemmli's Sodium Dodecyl Sulfate- Polyacrylamide Gel Electrophoresis (SDS-PAGE) method was employed to ensure the presence of proteins, particularly the known urinary proteins like albumin and uromodulin in urine sample concentrates. The Dual Xtra Prestained protein standards (Biorad, UK) were used as molecular weight markers in the gel. We used 12% acrylamide gels that were poured in the lab. The 1X running buffer contained 25mM of Tris-HCl, 200mM of glycine, and 0.1% (w/v) SDS.Urinary protein content were visualized with Bio-Safe^{TM} Coomasie blue 250G (Biorad, UK) staining after running 50µg, 100µg, 150µg on a 1D-SDS PAGE gel. To improve the resolution of the protein bands (especially albumin), lower concentrations of protein (i.e. 25µg, 20µg, 15µg, 10µg, 5µg, 2.5µg, 1.25µg, and 0.625µg) was loaded on duplicate gels, one was stained with Coomasie (**Fig. 2.4**) and the other was stained using the Pierce™ Silver

Staining kit (**Fig.2.5**). This basic protein gel experiments served as a foundation for my proteomics work.

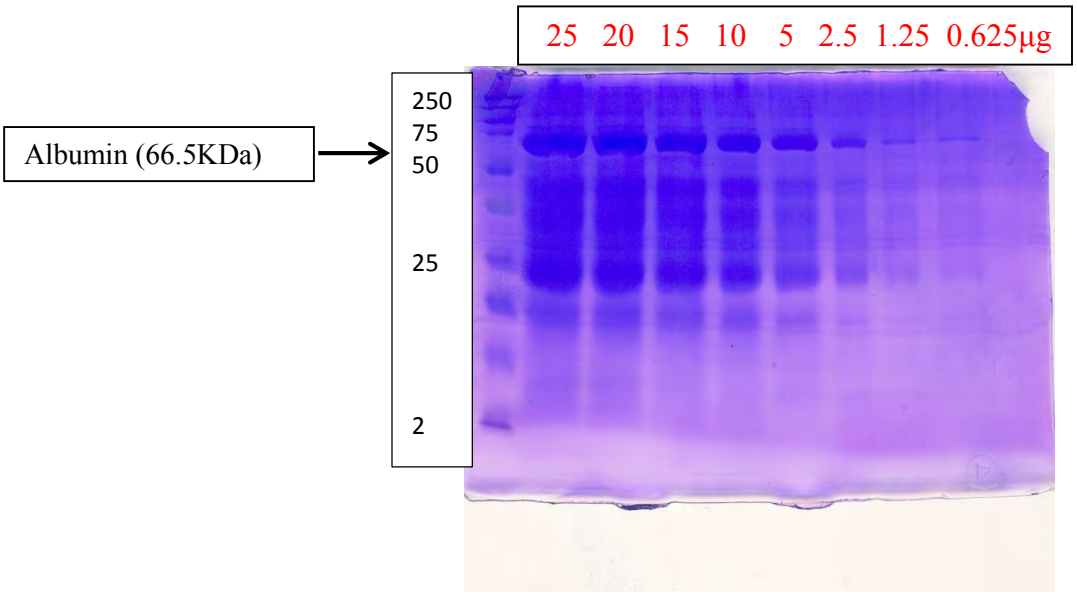


Figure 2.4. Coomasie blue staining of 1D- SDS-PAGE gel of serial dilution of urinary protein

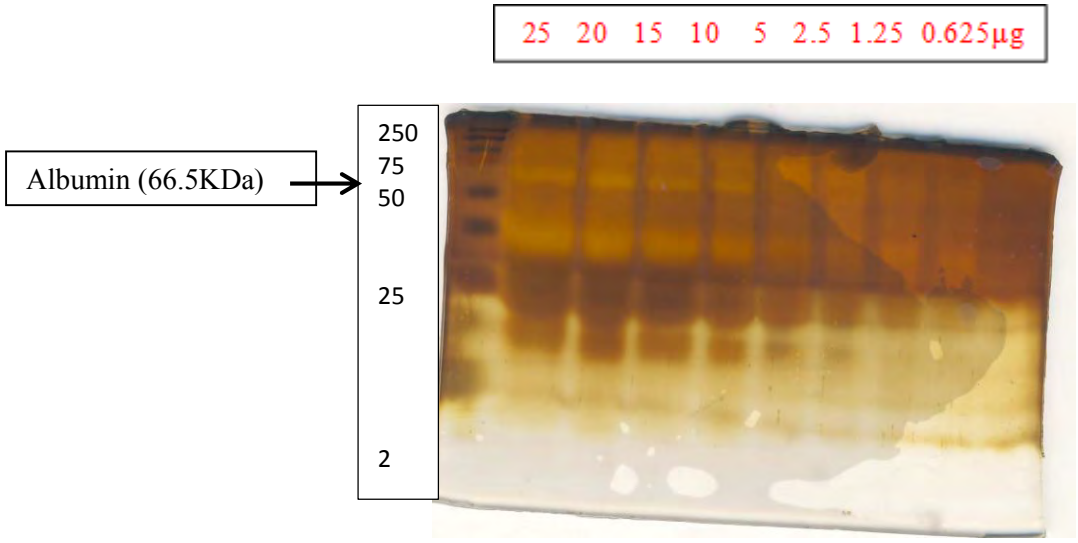


Figure 2.5. Silver staining of 1D- SDS-PAGE gel of serial dilution of urinary protein

2.2.1.3. Bradford Protein Quantification

The Bradford assay is a quick reproducible colorimetric protein concentration assay described by Marion M. Bradford in 1976 [205]. In this method, Coomassie Brilliant Blue G-250 dye binds to

protein components like basic (e.g. arginine) and aromatic residues and shifts its absorption from 465 to 595 nm. There are 3 forms of the dye *viz* neutral green, anionic blue and cationic red [206]. The dye is doubly charged (red 465nm) in acidic medium but becomes unprotonated (595nm blue) when bound to protein components [207]. Detection of this absorption shift using a spectrophotometer (cuvette) or microplate reader indicates the concentration of the protein being assayed. Detergents are capable of creating interference with the accuracy of this protein quantitation method but the use of adequately controlled and optimized experiment may eliminate this interference. Common standards used in this assay are bovine serum albumin (BSA) and bovine gamma-globulin. The protein standard is diluted serially (2, 1.5, 1, 0.75, 0.5, 0.25, 0.125 mg/ml) to generate multiple data points for comparative estimation of protein concentration. The Bradford protein assay kit used was purchased from Biorad, UK. Protein measurements were performed in triplicates and typically CV values were typically found to be less than 8%. Outlier samples were excluded from calculations and assay was repeated and re-measured if the CV values were greater than 10%. The Thermo-Labsystems Multiskan FC Absorbance 96-well Plate Reader as well as a clear polystyrene, Corning® Costar® 96-well Flat-Bottom EIA Plate (Sigma-Aldrich, St. Louis, MO, USA) were used for the Bradford assay. Bradford assay gives unreliable results in high concentrations of detergents like SDS, so the assay was modified by preparing the protein standards used in the same buffers that were used for sample preparation.

Steps involved in a Bradford assay on a microplate assay

1. Dilute the Bradford reagent (BioRad, UK) to 1:5 with deionized water in a tube.
2. Add 200µL of diluted Bradford reagent to into the test wells (standards and proteins of unknown concentrations) on the 96 well plate.
3. If protein is suspected to be too concentrated, it can be diluted x10 using phosphate buffered saline (PBS) or water.
4. Leave a blank well containing only the diluted Braford agent as control.
5. Add 10 µL of BSA standards (≥4) to the wells.
6. Then add 10 µL of the unknown proteins to the well.
7. Then insert into the spectrophotometer reader and read off in a slight agitation mode.

8. The data is then extracted in text format and analysed by plotting a standard curve (**Fig. 2.3**) to determine the concentration of the unknown proteins. The 595nm values read from the microplate reader are plotted on the *y*-axis, while the concentrations (mg/mL) are plotted in the *x*-axis. The concentration of the unknown protein can be deduced from the intercept equation ($y=ax+b$)) and R^2 value shown in the example in **Figure 2.6**. If any dilution was performed, correct the protein concentration result using the original dilution factor used.

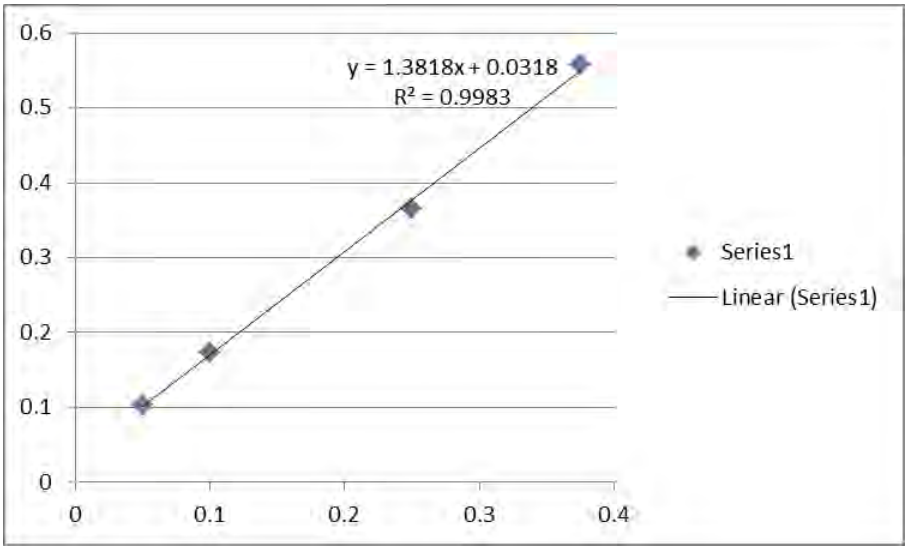


Figure 2.6. Standard curve derived from a Bradford assay.

2.2.1.4. Bicinchoninic Acid (BCA) Protein Quantification

This is another colorimetric total protein quantification method which is based on a Cu^{2+} - protein complex formation in an alkaline environment. Its advantage over Bradford assay is that detergents do not cause interference during the assay. However, reducing agents like dithiothreitol (>1mM) may interfere with the assay.

Steps involved in BCA protein quantification

1. Add 10 ul of each of the standards (1; 0.5; 0.25; 0.125; 0.0625mg/mL) to a well in a 96-well assay plate.

2. Add 10 ul of each unknown samples to the plate.

3. Working reagent is prepared by adding the BCA Solution (Pierce, ThermoFisher Scientific) to the Copper Solution at a ratio 50:1.

4. 200 µl of working reagent is then added to each well.

5. Plate is incubated for 45 minutes at room temperature (plate must be covered with a lid or parafilm).

6. Absorbance is read at 560 nm.

7. Extracted data is then analysed

2.2.1.5. In-Solution Protein Digestion

Proteins can be digested following quantification with in-solution digestion method. This entails the use of organic solvents to aid the precipitation and digestion of proteins into peptides. The materials (**Table 2.3**) and protocol for in-solution digestion are described below.

Solution and Reagent	Preparation Description
6 M Urea in Tris buffer (pH 7.8)	2.0 g of urea in a 15 ml + 1.25 ml of 0.4 M Tris stock. Adjust total volume to 5 ml with HPLC-grade water (Sigma-Aldrich, St. Louis, MO, USA).
200 mM iodoacetamide in Tris buffer (pH 7.8)	0.037 g of iodoacetamide + 750 µl of HPLC-grade water. Then add 250 µl of Tris stock and vortex.
200 mM Dithiothreitol (DTT) in Tris buffer (pH 7.8)	0.031 g of DTT + 750 µl of HPLC-grade water. The add 250 µl of Tris stock and vortex
Tris stock (0.4 M, pH 7.8)	12.1 g of Tris base + 200 ml of HPLC-grade water. Adjust the pH to pH 7.8 using 6 M HCl. Add HPLC-grade water to a final volume of 250 ml and store at 4°C.
Trypsin	25 µl of ice-cold Tris stock + 75 µl of ice-cold

	HPLC-grade water. Add solution to trypsin to a final concentration of 0.2µg/µL and store on ice

Table 2.3. List of reagents used for in-solution digestion and their preparation description.

Steps involved in in-solution digestion

1. After protein precipitation using methanol/ chloroform method, add 100µL of 6M urea.
2. Vortex the mixture and sonicate on and off for 2 minutes.
3. Dithiothreitol (5µL) is then added to the sample and vortexed.
4. Incubate at room temperature for 30 minutes to 1 hour.
5. Add 20µL of the alkylating agent iodoacetamide and vortex.
6. Incubate in the dark at room temperature for another 30 minutes to 1 hour.
7. Dithiothreitol (20µL) is added to the sample again and vortexed.
8. Incubate at room temperature for 30 minutes to 1 hour.
9. Dilute the solution with 775µL deionized water and vortex.
10. Add proteomics grade trypsin (Sigma-Aldrich, St. Louis, MO, USA) (1:50) and digest overnight at 37°C.
11. On the next day, stop the reaction by adding acetic acid to adjust the pH to 6.0.

2.2.1.6. Filter-Aided Sample Preparation (FASP)

Filter aided sample preparation (a.k.a FASP) is a relatively new technique of protein purification and digestion using molecular weight cut-off filters as described by Matthias Mann *et al* [208]. To summarize, important steps employed in FASP include: unfavourable low molecular weight species removal using urea containing buffer; thiol group carbamidomethylation; digestion of proteins with trypsin; and improved peptide elution.

Steps involved in FASP

1. Condition a 500µL 10 KDa MWCO (Amicon Ultra®, Millipore-Sigma, USA), with 500µL with UA and spin at 14,000 x g for 10 minutes and discard flow-through.

2. Add 200µL of UA and top up to 500µL with protein sample (100 and spin at 14,000 x g for 15 minutes.
3. Add another 200µL of UA and then repeat spin.
4. Add 100µL of IAA to filter and agitate gently.
5. Incubate in the dark for 20 minutes.
6. Then spin for 10 minutes at 14,000 x g.
7. Add 100µL of UA and spin at 14,000 x g for 15 minutes.
8. Repeat step 7.
9. Add 100 µL of ABC and spin at 14,000 x g for 10 minutes.
10. Repeat step 9.
11. Add 40µL of ABC + Proteomics grade Trypsin (Sigma-Aldrich, St. Louis, MO, USA) (1:50) and agitate gently.
12. Incubate in a wet chamber at 37°C for 16 to 18 hours.
13. Spin filters at 14,000 x g for 10 minutes.
14. Add 40 µL of ABC and repeat spin.
15. Acidify peptides by adding 0.1% TFA in water until pH is around 2-3 (using a pH test strip (Sigma-Aldrich, St. Louis, MO, USA)).

2.2.1.7. Desalting and Reverse Phase Extraction

Digested peptides were "cleaned up" by desalting/ reverse phase extraction which separates compounds based on hydrophobicity. The hydrophobic stationary phase we employed was made of C-18 material. In highly aqueous phase, peptides bind to the C-18 column and allow salts, impurities and undigested protein fragments to be washed off. However in high organic mobile phase, purified peptides can then be eluted for further analysis. Various solid phase extraction kits are commercially available but we used home-made stage-tips (described below) for desalting.

Steps involved in desalting

1. As described elsewhere [209], fabricate StageTips by making holes in the lids of irradiated 2ml tubes with clean sharp tweezers and inserting irradiated 200µL tip up to the hilt. Then place two discs of C18 material in each tip.
2. Equilibrate C18 with ≥5X C18 bed volumes of 80% HPLC-grade acetonitrile (in 0.1% TFA) and spin at 4000 x g for 1min.
3. Repeat step 2.
4. Wash the C18 column with ≥5X bed volumes of 2% HPLC-grade acetonitrile (0.1% TFA) and spin at 4000 x g for 1min.
5. Repeat step 4.
6. Load an equivalent of ≤15µg of peptides onto C18 column.
7. Wash with 50µL of 2% of HPLC-grade acetonitrile (0.1% TFA) and spin at 4000 x g for 1min to desalt.
8. Repeat step 7. (you may store peptides on column at 4°C at this stage for future elution)
9. Elute peptides by adding 50µL of 80% HPLC-grade acetonitrile (0.1% TFA) and spin at 4000 x g for 1min.
10. Repeat step 9 twice.
11. Lyophilize peptides on a SpeedVac (Thermo Fisher Scientific, USA) without heating (or low heat).
12. Re-suspend peptides in peptides in HPLC grade water containing 0.1% formic acid.

2.2.2. Blood Sample Preparation

Blood samples in Vacuette® K2EDTA venous blood collection set (Greiner Bio-One, North Carolina, USA) was centrifuged at 1,500 × g for 15 minutes to separate the plasma from cell fraction. The supernatant fraction was carefully removed; and 1.0 µl of plasma was diluted in 800 µl phosphate buffered saline in Triton-X-100 (PBST) for the CT100+ assay.

2.3. Mass Spectrometry-based Sample Processing

2.3.1. High Performance Liquid Chromatography (HPLC)

In HPLC, high pressure is introduced to force the analytes through a column for better chromatographic separation. HPLC permits column packing with finer chromatographic materials thereby improving the surface contact of the solvent with the stationary phase, and allows the use of much smaller quantities of analytes. Based on the stationary phase used and the relative polarity of solvent, HPLC can be normal phase or reverse phase. Compound separation is visualized by the means of a detector and the time taken for compound to travel through the column to the detector is known as the retention time. HPLC was performed using the specifications described below:

System specification for HPLC

1. Place re-suspended peptides in autosampler glass vials (labelled appropriately) and then queued up in the autosampler tray of the liquid chromatography system (Dionex UltiMate® 3500 RSnano LC systems (Thermo Fisher, San Jose, CA, USA).
2. Nanoflow ultra-high performance liquid chromatography (UHPLC) was performed online on a using a C-18 reverse phase precolumn trap (100µm×5cm; 5µM; 100Å) (Separations, South Africa) and analytic column (75 µm×50cm; 5µM; 100Å) (Separations, South Africa).
3. Gradient chromatography was performed at 23^0C with a flow rate of 300nL/min and eluted with a 5-80% gradient of water-acetonitrile for 180 minutes.
4. The binary mobile phase system was used as follows: buffer **A** contained water (990µL of HPLC grade water with 10µL of acetonitrile) and 0.1% formic acid, while buffer **B** contained acetonitrile (700µL of acetonitrile with 300µL of HPLC grade water) and 0.1% formic acid.
5. Gradient for peptides elution was 5% of **B** from 0-20minutes and increases to 80% of **B** at 180 minutes.

6. After completion of each run, the flow rate was increased to 450nL/minute and 50% **B** to equilibrate the analytic column and then dropped to 300nL and 5% **B** again prior to the next sample run.

2.3.2. Mass Spectrometry

Mass spectrometry (MS) is an analytical means of measuring the atomic or molecular weight of analytes. A mass spectrometer can be regarded as a highly sensitive and sophisticated weighing device that measures mass-to-charge (m/z) ratio of compounds. Its development made significant progress in the 1990's and current models have now attained femtomolar sensitivity. Depending on the type of mass spectrometer used, ionization of the molecules occurs as an integral part of the technique and is coupled with the type of resolution used. In principle, mass spectrometry involves placing ionized atoms in a magnetic or electric field and measuring its speed in relation to its *m/z* in a vacuum. Because it separates ions according to their measured *m/z*, it needs an ionization system and a mass analyzer. Two common ionization techniques employed by the mass spectrometer to ensure ionization of analytes into the gas phase includes electrospray ionization (ESI) and matrix assisted laser desorption and ionization (MALDI). A few commonly used mass analyzers are; quadrupole, Orbitrap, Fourier transform ion cyclotron (FTIC), ion trap, and time-of-flight (TOF). MS results are usually recorded as a series of peaks known as spectra formed by plotting the m/z on the x-axis and relative abundance/ intensity on the y-axis. The resolution of the spectra is defined as the width of the peak height typically measured at 50% peak height [210], it can be measured at any point. In mass spectrometry, the monoisotopic mass refers to the mass determined using the most abundant isotope of each atom in the ion, while the average mass is the mass derived from the abundance of all available isotopes of the same ion. Higher sensitivity of the MS system can be achieved using a nanospray ESI due to the higher surface to volume ratio [211]. There are generally three approaches to the analysis of MS based proteomics data which includes matching sequence tags to theoretical databases; *de novo* interpretation of peptide sequences; and assigning correlation scores to experimental-to-theoretical spectral match [212-214]. Peptide spectral matching scores can be assigned using q-values which are the minimal false discovery rates (FDR) or posterior error probabilities (PEP). The posterior error probability (PEP) score is the local probability that the observed peptide

spectral matching (PSM) is incorrect. It is calculated by the ratio of hits from true database vs decoy [215, 216]. FDR is the percentage of the global identification by PSM that is expected to be false; while the *q*-value is analogous to the *p*-value of the FDR that we are willing to accept [217-219].

2.3.3. Tandem Mass Spectrometry (MS/MS or MS^2)

Tandem MS which is achieved by using two mass analyzers in tandem and can give better insight into peptide sequence and its identity rather than merely depending on its molecular weight or m/z value. The first mass analyzer functions as a precursor/parent ion selector for fragmentation and the second mass analyzer functions as a fragment/daughter/product ion mass analyzer. Two types of mass analyzer may be coupled to a mass spectrometer in which case it is referred to as a hybrid e.g. Quadrupole- Orbitrap. Also, the same type of mass analyzer may be connected in series for MS^2 experiments e.g. Triple Quadrupole. Examples of MS^2 scanning modes are product ion, neutral loss and precursor ion scanning. There are different patterns of fragmentation which can be unimolecular or bimolecular [220]. Unimolecular fragmentation is also known as post source decay (PSD) and occurs when a single metastable (highly energetic) ion fragments suddenly causing peptide backbone fragmentation (*y*/*b* fragments). This type of fragmentation is common in MALDI setups due to bombardment of the matrix with a laser source [221]. However, MALDI TOF/TOF setups can also give rise to substantial internal fragmentation. (*x*/*a* or *z*/*c*). The bimolecular fragmentation also referred to as collision induced fragmentation (CID) occurs when precursor ions acquire sufficient fragmentation energy, due to multiple bimolecular events involving bombardment of the selected peptide with an inert noble gas in a collision cell, resulting in backbone fragmentation. Another bimolecular fragmentation method known as electron capture dissociation (ECD) or electron transfer dissociation (ETD) typically generates z/c fragmentation as a result of interaction with an electron donor, resulting in chemical fragmentation of the bonds [222, 223]. Traps typically use ETD while FTICs use ECD. Tandem mass spectrometry is a vital tool for proteomics which can employ a "bottom-up" (shotgun) approach or "top-down" approach. Shotgun proteomics is a discovery proteomics approach that determines protein identity from peptides generated from a mixture of digested proteins [224, 225]. However "top-down" approaches refer to intact protein analysis by mass

spectrometry [226]. There is relative quantitation and absolute quantitation proteomics. Relative quantitation is versus a control sample and no conclusions can be reached about concentration of the measured analytes. Absolute quantitation is determined using a standard curve to derive a measured value. Labelled and label-free methods can be used for relative quantitative proteomics which in principle attempts to quantify the identified proteins. Stable isotope labeling by amino acids in cell culture (SILAC) is an in-vivo method of labelling [227], while isobaric tag for relative and absolute quantitation (iTRAQ) [228], dimethyl labelling [229], and tandem mass tags (TMT) [230] are in-vitro approaches. Label free methods typically involve peak intensity measurement and MS1 spectral peak integration method using the "top three" approach and intensity based absolute quantification (iBAQ) method. Specification for our mass spectrometry runs are as described below.

Specifications of Mass spectrometry runs

1. QExactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher, San Jose, CA, USA) [231] was used for shotgun proteomics analysis in-line with a Dionex UltiMate® 3500 RSnano LC systems.
2. Flow injection analysis of samples introduced from the HPLC system was done with the Mass spectra (MS) acquisition resolution of 70,000 at a target value of 3×10^6 or maximum integration time of 250 milliseconds.
3. High-energy collision dissociation (HCD) and normalized collision energy set at 25 was used for peptide fragmentation.
4. Continuous MS^2 spectra acquisition resolution was set at 17,500 at a target value of 2×10^5 ions or maximum integration time of 120 milliseconds.
5. Automated data dependent full scan cycle was performed with automatic switching between MS/MS and MS scans at a scan range of 300-1650 m/z and a maximum injection time of 30 seconds.
6. The top 10 most abundant parent ions selected by the quadrupole during the initial scan is further fragmented using in-source HCD with normalized collision energy (NCE) at a pressure of 1.2millitorr and a dynamic exclusion time of 90 seconds.
7. Abundance threshold for ion selection was 0.001 and charge exclusion z =1 ions was used.

2.3.4. Proteotypic Peptide Selection

Targeted "hypothesis testing" proteomics focusses on a protein or peptide of interest, hence more efficient because they are dependent on adequate knowledge of the protein or peptide of interest [232]. Precursor peptide- fragment ion pairs known as "transitions" are used to confirm the identity of the targets. Transition selection is often difficult due to various technical and biologic factors which can be potentially addressed by using preexisting databases. A unique peptide for a target protein that present good MS responses is referred to as a proteotypic peptide [233]. In principle, these should be peptides detectable by the mass spectrometer and potentially capable of systematic incorporation into the available databases. The success of a targeted proteomics experiment is dependent on the successful selection of representative target transitions. Predictive machine learning algorithms have been previously suggested to overcome the potential bottlenecks of experimental optimization in proteotypic peptide selection [234]. Proteotypic peptides were selected prior to PRM assays. During proteotypic peptides selection, the following biologic and chemical criteria were considered for the targeted proteomics assay:

Selection criteria for proteotypic peptides

1. The goal was to select at least three transitions per target peptide (potential biomarker).
2. Shorter unmodified peptides that are unique to a single or specific isoform of the target protein with imino acid proline are better targets.
3. Poorly ionized longer peptides were avoided.
4. +2 and +3 charge states were selected for each proteotypic peptide.
5. +1 or > +3 were excluded from selection from further assay.
6. Peptides containing too many units of methionine (M) or tryptophan (W) were excluded due to the high propensity of artifactual side chain oxidation.
7. Based on adjacent sequences, asparagine (N) and glutamine (Q) were carefully considered due to possible chemical instability. For instance, N-terminal glutamine can change to pyro-glutamate when exposed to acidic treatment.
8. A maximum of two missed cleavages were permitted for each proteotypic peptide.
9. The possibility of having missed cleavages are identified to be increase in sequences containing two adjacent terminal basic amino acids such as –KR, -KK or -RR.

10. Because of low ion current as compared to their tryptic peptide counterpart, non-tryptic peptides were generally avoided [233]. Even though they can still be used if we are unable to find a good tryptic peptide.

2.3.5. In-silico SRMatlas Verification

In silico methods play a vital role in confirming the detectability of selected proteotypic peptides in previous targeted proteomics assays archived in targeted proteomics repositories. Selected reaction monitoring (SRM) is an important targeted proteomics method that permits the quantification of detected sets of protein in an unambiguous manner. Optimization and *a priori* characterization of target transition is required in this method [233]. A search for all our selected proteotypic peptides was performed against their target protein in the SRMAtlas database (http://www.srmatlas.org/). This is a publicly available online-based resource for choosing targeted proteomics experimental specifications. This relatively new database is derived from the PeptideAtlas projects which permits mining information on peptide transitions for target proteins. There is hierarchical integration of deposited data as well as bioinformatics predictions in a manner that generates a collection of the most suitable proteotypic peptide for each target protein in the proteome. Information contained in the database can also be repositioned in a manner that suits individual researcher's interest. There are two fundamental phases of the SRM build, which are firstly optimal peptide selection and secondly optimal transition selection [235]; enhanced by the robustness of integrated information from databases like PeptideAtlas SRM Experiment Library (PASSEL) [236] and PeptideAtlas Transitions Resource (PATR) [235] into the SRMAtlas database (**Fig. 2.7** and **Fig. 2.8**). PASSEL allows ease of submission of proteomic datasets generated from SRM by researchers and facilitates cross-analysis of such datasets. It also supports the review and optimization of SRM data collection [236]. PATR stores curated lists of transitions collected from community submissions as well as published articles. Peptide Atlas Best SRM Transition tool (PABST) algorithm used in PATR is a Peptide Atlas functionality that employs diverse sources of information in the production of peptides and fragment ions lists for use in SRMs. The PABST information is compiled into 'builds' that permits fast querying via a web interface. Three discrete steps are involved in the PABST build process *viz*; firstly, a list of the top 10 most suitable peptides for each target proteins; secondly, a

list of the top 8 potential fragment ions from each selected peptide; and finally the information generated is loaded into the Peptide Atlas database for end-user defined queries [237].

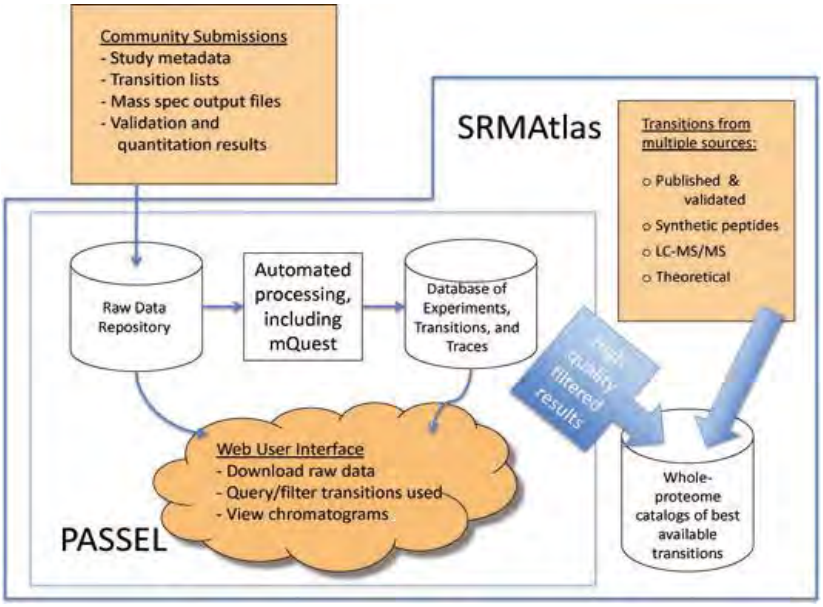


Figure 2.7. Integration of the PASSEL databases into the SRM database (adapted from [236])

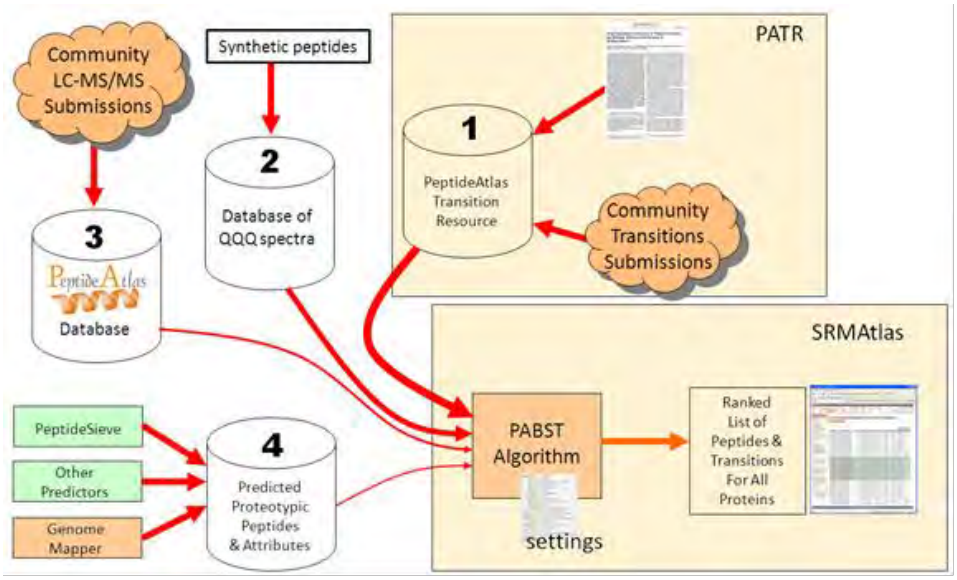


Figure 2.8. Integration of the PATR database into the SRMAtlas (adapted from [235])

The specified setting that was used for the SRM database search is specified below:

Specified setting for SRMAtlas database verification of potential biomarkers:

1. Protein accession identification number was filled in.

2. Transition source was selected as "ion trap" since an Orbitrap instrument was used.
3. The specified *m*/*z* range selected was 200-2000*m*/*z*.
4. Only *b*- and *y*- ions were permitted
5. Neutral loss MS^2 scanning mode was not permitted.
6. Target database selected for search were Ensembl, IPI and SwissProt.
7. Default setting were accepted for other parameters

2.3.6. Human Protein Atlas Verification

Biological screening of potential biomarkers was performed using the Human Protein Atlas database (http://www.proteinatlas.org/). The database is publicly available online and contains images generated from twenty different human cancers, 44 normal human tissue types, and 46 different human cell lines. The database is organized in a gene-centric manner and information regarding genome annotation of human protein-coding genes, protein expression level, as well as subcellular localization of protein is made available in this repository. The most recent version launched at the end of 2014 uses antibody profiling and transcriptomics analysis to provide information about various proteins in humans. Two important underpinning principles employed by the human protein atlas to generate information in a high-throughput manner are the improvement on the specificity and non-overlapping epitopes of the antibodies used as well as the use of tissue microarrays for immunohistochemical analysis. There are four sub-atlases in the human protein atlas *viz*; the normal tissue atlas, the cancer atlas, the cell line atlas and the subcellular atlas (**Fig. 2.9**). For verification, the cancer atlas and normal tissue sub-atlases were used. Normal and cancerous prostate tissue immunohistochemistry was assessed by searching either name or gene symbol of the investigated potential biomarkers. Immunohistochemical staining of target proteins was compared between normal and cancerous proteins. Proteins with differential expression between normal and cancerous tissue were documented as verified potential urinary biomarkers from shotgun proteomics experiments.

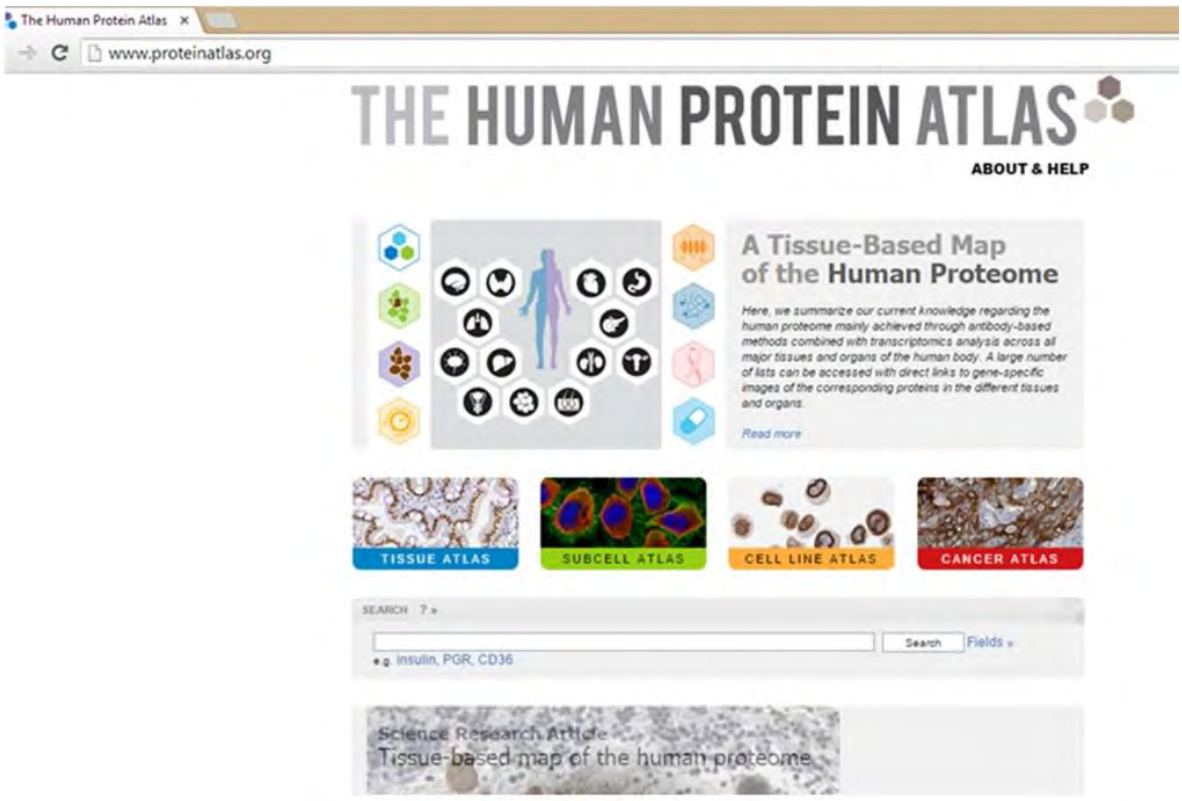


Figure 2.9. Screenshot of the query page of human protein atlas showing the four sub-atlases (www.proteinatlas.org) which are the normal tissue atlas; the cancer atlas, the cell line atlas, and the subcellular atlas.

2.3.7. Parallel Reaction Monitoring (PRM)

Parallel reaction monitoring is a quantitative high resolution targeted proteomics method. It is similar in some respect to selected reaction monitoring (SRM) which targets only a few transitions (precursor-product ion pairs) and is typically carried out on a triple quadrupole instrument. Parallel reaction monitoring involved the multiplexed/ parallel detection of all product ions in a single high resolution mass analysis; using a substitution of the third quadrupole mass analyzer with a high resolution and accurate mass Orbitrap or time-of-flight mass analyzer. Some advantages of the PRM method over SRM are that multiple transitions can be monitored to confirm the identity of the target precursor peptide; ion interference would be better tolerated by the full scan without affecting the quality of the spectra because of the high resolution of the instrument; less optimization is required in comparison to the SRM setup. On the other hand, sensitivity of the PRM would be less compared to the SRM because of the optimal duty cycle in SRM which employs an electron multiplier as compared to the trapped

image current detection system used in the Orbitrap instrument, although it has been suggested that due to high resolution and accurate mass, PRM may still be a better method for targeted survey of transitions [238], while SRM may be more accurate for absolute quantification. PRM was carried out on pooled healthy and prostate cancer patients samples using the settings described below:

Settings used for Parallel Reaction Monitoring Assay

1. PRM runs were performed on a QExactive^{TM} hybrid Quadrupole- Orbitrap Mass spectrometer.
2. A spectral library is first created using data from the original shotgun urinary proteomics assay.
3. An isolation list in Skyline (MacCoss Lab software) [239] is then generated.
4. Precursors are randomized in the isolation list generated to ensure that the following precursor m/z value is similar and also a randomized order of +2 and +3 of the precursor ions.
5. Isolation lists are uploaded into the system software connected to the mass spectrometer.
6. An isolation window of 2m/z units for each precursor ion was used; starting at an m/z of 80 and final *m*/*z* is then automatically derived based on the *m*/*z* of the proteotypic peptide precursor.
7. After generation of MS^2 data, raw files were imported into Skyline to choose the precursors with good signals.
8. The process is repeated using an isolation list that is ordered differently to ensure that interference from multiplexing does not create bias in our inferences [240].
9. Relative quantitative analysis was then performed to compare biomarker transitions between study groups.

2.3.8. Database Confirmation of Urinary Prostate Cancer Biomarkers

Identified top ranking potential biomarkers that have passed through three levels of stringency viz; SRMAtlas database, Human Protein Atlas and Parallel Reaction Monitoring were confirmed

in our shotgun urinary proteomics database as well as 14 other publicly available urinary proteomics databases.

2.4. Cancer Antigen Array-based Sample Processing

2.4.1. Streptavidin Derivatization for Cancer Antigen Array

Streptavidin is a 58.2KDa tetrameric protein that frequently used in immunologic hybridization assays. This protein which is purified from *Streptomyces avidinii* bacterium species [241], possesses an extraordinary high non-covalent affinity for vitamin B7 (biotin) with a dissociation constant (K_d) of ca. 4 x 10 $^{-14}$ [242]. The high binding strength makes it an attractive tool in molecular biology and nano- or bio-technology. Streptavidin biotin bonds are resistant to detergents, organic solvent as well as a wide variation in temperature and pH changes [243]. The bond strength remains a subject of research and scientific discussion in molecular biology and methods to strengthen and improved its efficacy have been well documented [242, 244-246]. For the cancer antigen array experiment, we performed in-house derivatization of a Nexterion-H slide with streptavidin. The Nexterion H-Slide demonstrates low intrinsic non-specific binding, making assays relatively independent of a blocking step. Its polymer coat has three important components, namely: an NHS ester reactive group, a terminal-amino modified amino acid and a surface exposed amine-group (**Fig. 2.10**). The properties of the Nexterion-H slides are as described below:

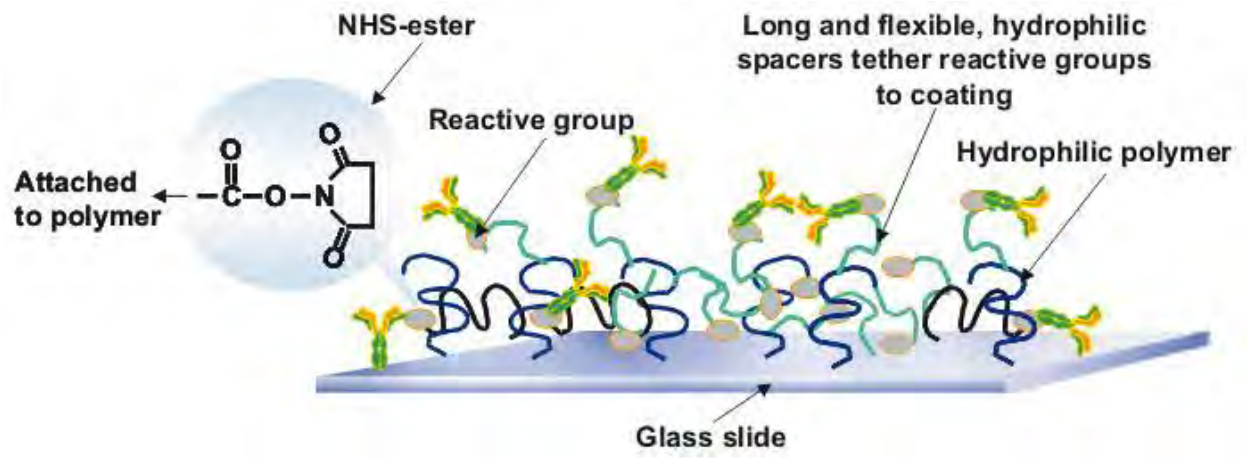


Figure 2.10. Properties of the polymer coating of Nexterion-H slides (http://www.schott.com/)

Properties of the Nexterion H-slides used:

1. Made in Germany
2. Contains 25 slides per pack
3. Thickness is 1.0 ± 0.05mm
4. Size is 75.6mm x 25mm x 1.0mm (L x W x H)
5. It is black barcoded
6. It is hydrogel coated (N-hydroxysuccinimide esters to provide covalent immobilization)
7. Should be stored at -20°C

The steps involved in the streptavidin derivatization are as follows:

Streptavidin derivatization protocol:

1. Slides were thawed for one hour at room temperature.
2. Streptavidin solution was prepared in 150mM Na_2HPO_4 buffer at pH 8.5.
3. Each slide was immersed in 5mL of 1 mg/mL streptavidin solution for one hour at room temperature.
4. Slides were then removed from the streptavidin solution (reusable).
5. The slides were washed for one hour in 10mL of 150mM Na2HPO4 buffer (pH 8.5) containing 50mM of the blocking agent (ethanolamine).
6. Each slide was washed thrice for 5 minutes in 10 mL of washing buffer.
7. It was then washed for 10 minutes in 10 mL of deionized water.
8. The slides were dried at room temperature by centrifugation at 1000 x g for 5 minutes.
9. For quality checking, the last slide per batch was incubated in 10µg/mL Cy5-biotinylated BSA in PBS and washed in washing buffer, and then scanned on Tecan LS reloaded microarray scanner to confirm even spread of streptavidin; and only slides with CV ≤5% were used for the assay.
10. The slides were then transferred to slide boxes and sealed in foil bags for storage at -20°C until use.

2.4.2. Cancer-Testis Antigen Protein Microarray Fabrication

Protein microarray technologies have been classified into two major groups, which are functional protein microarrays and analytic (protein-detecting) protein microarrays [247, 248]. Protein microarrays can be fabricated using recombinant proteins [249-253], reverse capture of native proteins [247, 248, 254-258], cDNA expression libraries [259-268] and in-situ expression techniques [269-274]. Native proteins are able to retain their post-translational modifications and complex structure as compared with recombinant proteins and do not necessitate a cloning process. However identifications generated via native proteins are often not reproducible. Using cDNA libraries like SEREX [268] is simple, can be combined with robotics and yields a vast library of proteins; however generation of truncated cDNA clones from *E.coli* system, poor folding of human proteins in *E. Coli* and lack of post-translational modifications or protein complexes are major drawbacks. In situ systems like Nucleic Acid Programmable Protein arrays (NAPPA) [267], in-situ puromycin capture, and protein in-situ arrays (PISA) [270] have the advantage that they use eukaryotic cell free system to translate arrayed cDNA to protein; however optimization across several cDNAs can be challenging, expressed protein may be misfolded, and expression is often not reproducible. An alternative system to overcome all these potential limitations is to express recombinant human proteins in "eukaryotic-like" insect cells. This would potentially overcome problems with folding, functionalization and reproducibility. It is important to note that post-translational modifications in insect cells are not the same as in humans. The "CT100+" cancer-testis antigen protein microarray employed in this study is based on recombinant protein expression in insect cells that employs the linearized baculovirus genome system [275, 276]. The steps involved in microarray fabrication are as listed below:

<u>Steps in CT100+ microarray fabrication</u>

1. Crude SF21 insect lysate containing individual, *in vivo* biotinylated cancer-testis antigens were printed on the streptavidin coated Nexterion-H slides using the robotic arrayer, Genetix QArray 2.
2. Biotinylated human IgG and serum were used as positive controls.
3. Crude cell lysates containing BCCP-tag without a recombinant fusion partner was used as negative control.

4. On the microarray, each subarray contained biotinylated Cy5-BSA on triplicate spots at 5, 10 and 15ng/mL.(this is used for slide orientation).
5. Before printing commences, slides were inspected and confirmed to be free of fingerprints and dust particles by microscopy (X40 magnification).
6. Microarray equipped with 300μM solid pin is employed to print CT100+ replica in a 4plex format.
7. Crude lysates were mixed with 40% sucrose and placed in triplicates in the array. Sucrose reduces protease release by stabilizing lysosomal membranes.
8. Positive and negative controls were prepared using lysis buffer containing 20% of glycerol, 50mM of KCl, Triton X-100, 25mM HEPES (pH 7.5), 50μM of biotin, 1mM of DDT, and supplemented with 20% sucrose.

For the Genetic QArray 2 robotic arrayer, the settings described below were used:

Settings for robotic arrayer

1. Row to column pitch was 562 microns.
2. Arraying pattern was 8 x 8 pins in 4 fields.
3. Number of stamps per spot was 1
4. Number of stamps per ink was 1
5. Stamp time was 0
6. Print depth adjustment was 150 microns
7. Inking time was 500 ms.
8. Water wash used was 60 s and dry time was 0 ms
9. Number of touch-offs was 0
10. Ethanol wash used was 10s and dry time was 10s

2.4.3. CT00+ Cancer Antigen Array Assay

In all, seventy samples were assayed including 3 quality control samples. The quality control samples comprised of negative control serum, positive control serum, and a mouse-anti-c-Myc-Cy3 antibody. The steps involved in the assay are highlighted below:

Protocol for CT 100 + Antigen array assay

1. CT 100+ microarray was washed with lysis buffer containing 20% Glycerol, 25mM HEPES, 50mM KCl, 0.1% Triton X-100, 1mM Dithiothreitol and 50 μM biotin for 5 minutes at room temperature.
2. Repeat step 1 twice.
3. Slides were dried by centrifugation at 16 x g for two minutes.
4. Slides were then placed in the automated hybridization station (TECAN HS4800 Pro) to isolate each array.
5. Each microarray was then incubated with diluted sera (1:800 in PBST) at 23°C for one hour.
6. After incubation, the microarray is washed with PBST for 5 minutes.
7. Repeat step 6 twice.
8. Microarray was then incubated with Cy5-goat anti-human IgG (diluted 1:100 in PBST) detection antibody at 23°C for 1 hour.
9. Slides were then washed four times for a minute with PBST.
10. Rinse with water for 30 seconds.
11. Individual arrays on the slide were then scanned with 10μm resolution on the Tecan LS Reloaded microarray scanner.
12. Fluorescence is detected at 650 nm.
13. Image is saved as a 16-bit TIFF file.

2.4.4. Data Extraction from Antigen Array

Information from the TIFF file was extracted using the steps described below:

Steps for data extraction from TIFF image

1. The TIFF images generated from the CT100+ assay were initially inspected using simple visual analysis to first ensure that the BSA spots (5, 10 and 15μg) were detectible in the subarrays of the microarray.
2. High signals for positive control which is human IgG detected by secondary antibody fluorescence was ensured and human anti IgG detected on addition of serum were observed in the images.
3. Arrays that show; speckles, high background, interfering dust particles and protein spot coalescing at inspection were excluded.
4. If array displayed attributes mentioned in step 3, sera was re-assayed on another fresh slide.
5. Data was then extracted using the ArrayPro Analyzer Software version 4.5 (Media Cybernetics, Inc, Maryland, USA).
6. Using a grid, identity of the antigens and control were aligned and encircled on the TIFF image.
7. Raw intensity of the individual spot was measured as mean pixel within the defined border to set maximum threshold intensity of 200 pixels.
8. Local background for each circumscribed spot was measured as adjacent area intensity.
9. The difference between raw mean intensity and local background was designated as the mean net intensity.
10. Data was extrapolated with ArrayPro Analyzer Software version 4.5 (Media Cybernetics, Inc, Maryland, USA).
11. Then data was filtered and normalized using our in-house developed CT100+ software [277, 278].
12. Data was finally analysed.

2.5. CT100+ Analysis Programme

The aim of our CT100+ programme was to provide customized normalization and qualitative sample clustering using the parameter described in the following sections.

2.5.1. Spot Homogeneity

Homogeneous signals are expected across all pixels within each spot. A deviation from this include what is referred to as the "doughnut effect" which is usually caused by inadequate pin height at print run and residual liquid on pin body during pin head immersion in source plate. This tends to result in uneven distribution of spots. When slides are inadequately stored and handled during the CT100+ assay, dust particles could settle on spot of interest and erroneously give a high pixel intensity thus skewing the actual signal on the array surface. Temperature increases and decreased humidity may result in printed spot evaporation; while humidity increase greater than 75% may result in condensation on the spots and account for unevenness of intensity. Coefficient of variation (CV) is employed to evaluate variations between replicates. This can be done by measuring the ratio of the mean intensity (%) and the standard deviation of all pixel intensities within the spot. Ideally, the CV is not expected to exceed 20%.

2.5.2. Spot-to-Spot Variation

In principle, signals of triplicate spots are expected to be similar; and spots should be uniform across the whole microarray. Practical deviation from this includes a phenomenon known as "spot bleeding" whereby two or more closely positioned spots run into each other. This may also occur as a result of using inappropriate spotting buffer and may potentially compromise the signal of all affected spots. When arrayer pins and printheads are inadequately cleaned between print runs, this causes erroneous pin calibration or pin sticking whereby some or all prints may fail or may even result in the generation of non-reproducible spots. If the washing step is inadequate throughout the assay, this may result in formation of large washing artefacts or speckles. This shows as a negative spot or additional smaller random spots across the array surface.

2.5.3. Signal-to-Noise Ratio

When evaluating multiple spots across arrays, net spot intensity (Foreground intensity minus background intensity) is expected to be higher than the neigbourhood or local background. Washing artefacts, speckles and dust particles may interfere with this expectation. To potentially overcome this, a signal-to-noise ratio ≥2 is accepted (where standard deviation of the background pixel is defined as the "noise")

2.5.4. Background Variation

When looking at multiple spots on an array, there should be minimal variation in the background between neighboring spots. Dust particles may cause variations in recorded background intensities across spots on the microarray.

2.5.5. Pixel Saturation

Saturated pixels above the slide scanner's reading capacity (limit of detection) should not be observed when evaluating multiple spots across the array. The reading capacity of the Tecan Reloaded scanner used in our laboratory is ca. 65,000 RFU. If such saturated spots are detected, the slide needs to be rescanned at a lower photomultiplier tube (PMT) gain setting until saturation is no longer detectible across the array [279].

2.5.6. Background Subtraction and Correction

Local background can be smoothened by correcting neigbourhood background. It can be done by reducing the noise and artefact effect in the neigbourhood background. This allows for accurate calculation of net intensities by removing the corrected neigbourhood background from the median local foreground pixel intensity value.

2.5.7. CT100+ Cancer Antigen Microarray Data Filtering

Quality controls checks are run after background corrections. This entails filtering out of defective or noisy array data prior downstream bioinformatics analyses. Filtering improves data quality and stringency of analysis. Low quality/ questionable arrays or individual spots are flagged. For example spots with foreground intensity close to 65,536 RFU are flagged and triplicate spots with high CV *vis-à-vis* spot intensities can also be flagged. In triplicate spots, if one of the spots is flagged, variation in the remaining two spots S1 and S2 can be calculated as follows: $(|S1 - S2|)/(S1 + S2)$. Net signal close to the noise level can be flagged using 2 standard deviation of the local background pixel intensity as a flag. In principle, negative controls on the array can allow filtering of low intensity spots because they depict cross reaction of detected antibodies with copurifying SF21 insect cells or the BCCP-tag. In reality antigen-to-antigen differences in expression level or access to BCCP-tag makes this approach less straightforward. A practical approach may be to establish baseline threshold intensity for individual antigens across a large healthy sample cohort.

2.5.8. Custom Microarray Data Normalization Methods

The goal of the normalization workflow is to provide efficient and robust methods of correcting technical and systematic bias in array to array and pin to pin discrepancies. This is performed bearing in mind the relatively small size of the positive control sample cohort. Assumption is made here that the positive controls share a common distribution across the arrays upon which they are printed. Hence, a combination of both intensity and quantile normalization modules is used to eliminate systematic bias in the composite normalization workflow. The quantile approach as described by Bolstad et al 2003 [280], identifies housekeeping spot intensities based on their underlying distribution and then reorganizes them based on the step described below:

<u>Steps involved in the quantile normalization module</u>

1. Assuming that for a positive control spot i on a chip j, S ij is the intensity.
2. The positive control spot intensities are loaded into an i x j matrix, represented as X.
3. The spot intensities are then sorted in each column j of X to get X_{sort}.

4. The means across each row i of X_{sort} is then taken to get iX, where iX is the distribution of spot intensities of positive control.

Total intensity based module described by Quackenbush et al 2001 [281] takes on the assumption that all spots have similar intensity; meaning that intensity sum for all positive controls must be constant. This is a scaling method that factors in flagged spots. If any positive control spot is flagged, the corresponding spot is flagged across all arrays to minimize discrepancies in the number of positive control spot used across the microarray.

2.6. Autoantigen Confirmation in Urinary Shotgun Proteomics Database

All antigens present on the CT100+ platform were searched against our urinary shotgun proteomics database. Antigens that were found to have high autoantibody titres were also checked in the database to assess if there were any correlations between autoantibody expression in blood and urine since most of the potential biomarkers that were discovered from the shotgun database were predicted membrane proteins and of extracellular matrix origin.

2.7. Functional Pathway Analysis

Functional pathway analysis is essential to identify common pathway involvement by the potential biomarkers or investigate functional relationship. Enquiry about enrichment for gene ontology (GO) categories such as biologic process, molecular function and cellular component can be performed on the annotation databases. There are several annotation databases available for pathway analysis, a few of which are; the REACTOME, Kyoto Encyclopedia of Genes and Genomes (KEGG), molecular signature database (MSigDB), UNIPROT, etc. Functional Pathway enrichment analyses were performed on the antigens with high autoantibody titres; using a free online gene interaction software GeneMANIA and enrichment was confirmed by functional protein association network software STRING (*version 9.1*).

2.7.1. GeneMANIA

The GeneMANIA web interface (http://www.genemania.org/) is a user-friendly tool for functional pathway analysis of gene sets. Hypotheses can be generated from this interface on gene list analysis; gene function and functional assay prioritization of genes; and assigning

predictive weights for genes in selected datasets [282]. For each query, GeneMANIA assigns weights to data-sets based on how useful they are. Each data-set is represented as a network, and each network is assigned a weight based on how well connected the queried genes are to each other compared to non-queried genes. The weighting network feature used by GeneMANIA identifies and down-weight redundant networks. Annotation or co-annotation based weighting may inflate network weights for an organism in the GO categories of biological process, molecular function, or cellular process hierarchies. However, high-throughput assays with annotation-based schemes are hardly affected by this potential inflation [282]. Gene list is entered and organism species is selected and queried against a list of selected datasets (**Fig. 2.11A**). Functionally similar genes to the original query list are then displayed in an interactive association network based on shared function (**Fig. 2.11B**).

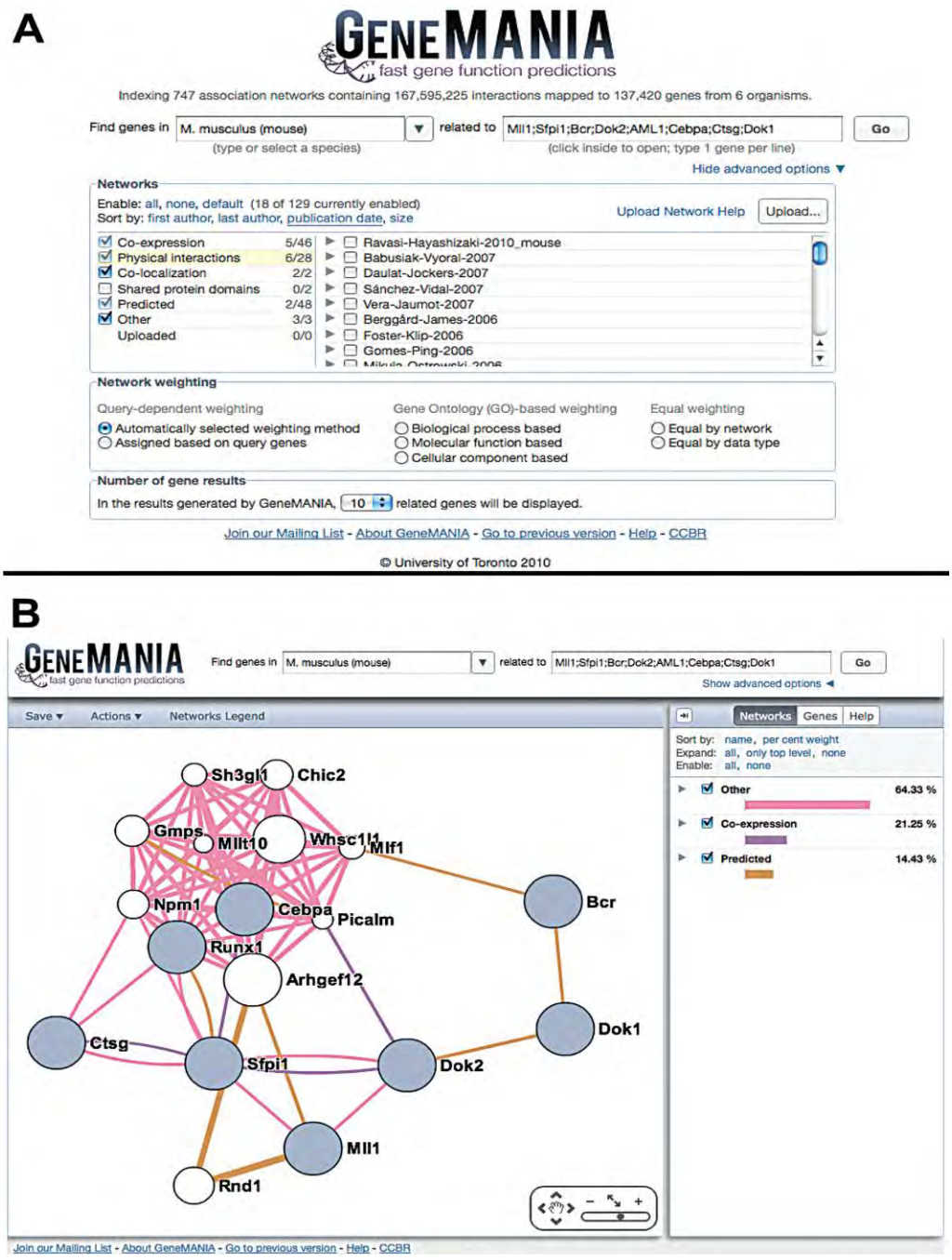


Figure 2.11. GeneMANIA web interface showing (**A**) the query page where user can input gene list and select search categories and (B) interactive functional network generated from query (adapted from [282])

2.7.2. STRING

Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) is a web-based resource (http://string-db.org) as well as a biological database, for interrogation of protein-protein

interaction, and functional pathway analysis [283, 284]. Data sources for STRING include curated literature, computational predictions and experimental data in physical interaction and curated biologic pathway databases [285, 286]. Data generated on STRING are also weighted and confidence scores are assigned to the predicted protein interactions [283]. Protein-mode and clusters of orthologous groups (COG)-mode can be used on this web tool. At least three lineages of orthologous paralogs are used for each orthologous protein in the COG-mode [287]. STRING permits a filtering process for functional genomics interaction by incorporation systems-oriented understanding of cellular processes. STRING predicts the specificity and biological meaningfulness of protein-protein functional association; and this makes it better than many other software. It attempts to predict the full complement of possible interaction for any fully sequenced organism, albeit only a subset would be possible in any given cell. The STRING database periodically execute interaction prediction algorithms that significantly increasing coverage for poorly studied organisms and assigns confidence scoring of coverage and accuracy to them. Currently, STRING is based on Adobe Flash Player which allows users to freely reposition nodes in, as well as applying clustering algorithm to the network; even while running a real time spring-embedded layout algorithm [283]. Most of the associations in STRING are derive from genome-based prediction algorithms or by pre-computed interolog transfer. Generally, a probabilistic confidence score is used to provide association in STRING. This score is derived from separately benchmarked groups of associations against the manually curated functional classification scheme of the KEGG database. The sources of interaction in STRING database includes: importing known experimental interaction from primary databases; parsing pathway knowledge from manually curated databases; automated text-mining of Medline abstract or full-text articles; de novo prediction of interactions by algorithms using genomic information; and systematically transferring interactions observed in one organism to another [286]. Selected gene or protein lists are entered on the query page and organism species is selected and queried (**Fig. 2.12**). Output displayed for network generated includes predicted co-expression, neigbourhood evidence, occurrence in similar biological pathway, and gene fusion-fission events (**Fig. 2.13**).

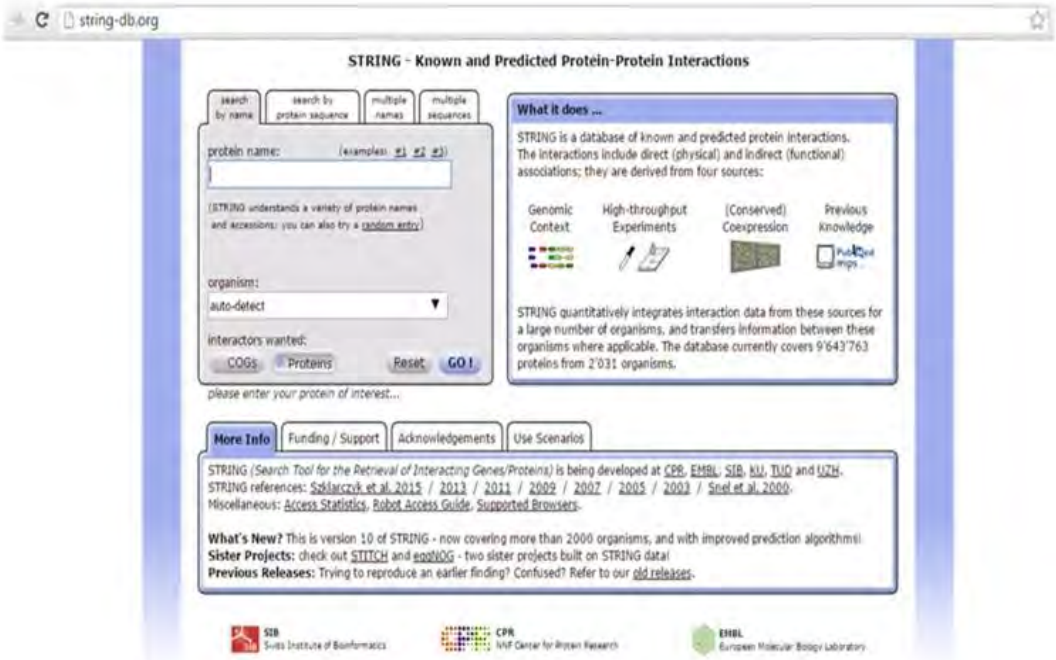


Figure 2.12. Screenshot of STRING query page (http://string-db.org)

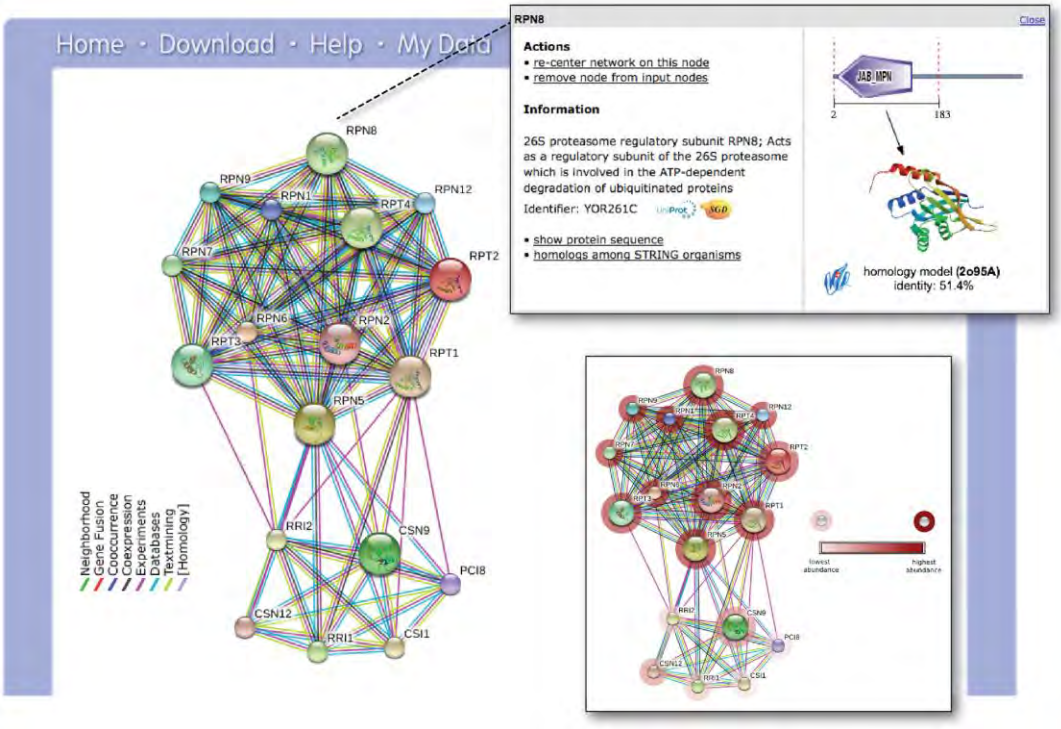


Figure 2.13. Example of STRING network visualization showing predicted neigbourhood, occurrence, fusion and coexpression of proteins [286].

2.8. Bioinformatics and Computational Analyses

Owing to the complexity and huge amounts of data generated in proteomics experiments, various computational analyses, algorithms and workflows are needed for systems-level identification of biologically significant differences in the proteome. Bioinformatics workflows permits systematic integration, management and elucidation of large amount of data generated from high throughput omics based experiments, in a seamless manner. Newly emerging software packages aim to a supplement basic statistics and overcome the bottleneck often faced by researcher in data analysis and experimental inference generation. Some of the computational workflows and software used for analyzing the proteomics data are described in the following sections.

2.8.1. MaxQuant Analysis

MaxQuant, which originates from Matthias Mann's Group [288], is a C++ based algorithm suite for statistical analysis of high throughput mass spectrometry-bases proteomics data. This software permits an accurate proteome-wide identification and quantification of proteins from the estimated spectral-based peptide masses generated from precursor scans [288]. This feature is enable by an integrated search engine know as Andromeda which assigns score-based identification to fragment ion, peptides and protein groups. In MaxQuant, proteins sharing the same identified peptides together are grouped into one and are called protein groups by the protein group assigning algorithm. All the proteins in one protein group have a similar or slightly less number of identified peptides [288, 289]. The Andromeda scoring system is based on the binomial distribution probability [289]. A binomial experiment requires repeated independent trials which can result in just two possible outcomes (i.e. success or failure); and the probability of success is the same for every trial. Peptide scoring functions are assigned in binomial distribution based probabilistic algorithms by comparing spectral peak characteristics in experimental to theoretical target databases. MaxQuant version 1.3.0.5 was used for the analyses and the steps followed are as listed below:

<u>Steps used for Maxquant analysis</u>

1. Load *raw Xcalibur files generated from the mass spectrometer into Maxquant.

2. Set experimental design template
3. Set experimental parameters using default settings with minor modifications (See **annexure III** for details)
4. Select number of "threads" for analysis
5. Run the experiment.
6. Generate result and Analyse.

2.8.2. RStudio

R is a free, open source, C++ -based programming language that is highly useful for large-scale statistical and graphical computing of high throughput omics data [290]. RStudio (www.rstudio.com) is a user friendly application based on the R interface and available as a desktop application. Various codes and scripts are already written and deposited as packages in the Comprehensive R Archive Network (CRAN) repository (https://cran.r-project.org/). Basic data in R can be numeric, logical, integer, complex, list, data-frame or character. RStudio software (*version 0.98.987.0*) was used to analyse and create visualization for our data. A summary of the steps used is as highlighted below:

Steps for Rstudio analysis:

1. Packages for desired analysis was downloaded from CRAN
2. Packages were installed.
3. Working directory was set on the computer.
4. Experimental dataset was read into RStudio.
5. R was then used to prepare data prior to analysis.
6. Final analysis was performed

2.8.3. Perseus

This is a user-friendly C++ based software platform for downstream computational analysis of shotgun proteomics output data, generated from the Maxquant software. Thus, this software is capable of completing the Maxquant proteomics analysis pipeline. Numerous independent

statistical tools are integrated into the Perseus workflow (available at http://www.biochem.mpg.de/5111810/perseus). The software is capable of processes like, normalization, data imputation, transformation, t-tests, ANOVA, clustering, correlation, enrichment, volcano plots, motif identification, scatter plots etc. System-wide analysis is possible because Perseus permits the integration of data from different omics sources. The filtering process in Perseus is user-defined. Usually, it is recommended that protein groups with less than 50% presence among replicates should be removed even though it is possible to adjust this percentage to meet various experimental objectives. Perseus (*version 1.4.0.20*) was used for analyses of the data generated from Maxquant analysis as described below:

Steps employed in Perseus analysis

1. Download and install Perseus software
2. Load Maxquant result into Perseus.
3. Select data frame.
4. Filter data.
5. Transform data.
6. Categorical row annotation
7. Filter based on valid values.
8. Downstream statistical analysis (correlation, hierarchical clustering etc.)

2.8.4. InfernoRDN

Previously referred to as DAnTE, InfernoRDN (available at https://omics.pnl.gov/software/infernordn) is a Windows-based GUI for R. It is used for downstream analysis and visualization of omics based data, with an emphasis on proteomics. Various diagnostic plots such as boxplots, Q-Q plots, histograms, principal component analysis, peptide-protein rollup plots, correlation plots, heat maps, hierarchical clustering and MA plots can be generated on this software. InfernoRDN (*version 1.1.5274*) was used to create additional visualization in our study.

2.8.5. Preprocessing and Normalization

Raw expression data require preliminary processing prior to proper downstream analysis and this is referred to as preprocessing. Noisy data is bound to obscure or bias statistical inference and data containing too many missing values may mask the actual biologic meaning of generated data, hence filtering steps need to be applied to make data more amenable to statistics. After data filtering, simple visualizations like boxplots, histograms and Q-Q plots can be used to assess the spread of data. In addition, skewedness (lepto-kurtosis or platy-kurtosis) can be assessed using simple histograms. It is important that the distribution of data points follow a normal Gaussian distribution to be able to apply parametric statistical tests to it; otherwise a non-parametric equivalent can be employed. Normality of data distribution can be tested using Shapiro-Wilk, Anderson–Darling or Kolmogorov–Smirnov test [291]. To increase signal-to-noise ratio, data normalization is essential and a few examples of normalization methods are; Lowess regression, logarithmic transformation, quantile normalization, peptide level ANOVA, etc. On the other hand, over-normalization can result in "overfitting" of data. In this study we used logarithmic transformation for normalization.

2.9. Statistical Analysis

Definitive downstream statistical analysis was performed en-suite with the computational workflows. A few important diagnostic statistical methods used for the proteomics data analyses are briefly described in the sections below.

2.9.1. Power Statistics

The ability to detect a statistically significant difference in a population if indeed the difference is present is known as the "power", given by the equation $(1 - \beta)$, where β is the false negative rate (Type II error). It can be demonstrated by applying confidence interval to an effect [292]. Power calculation helps in the determination of the sample size needed for the study, depending on the size of the effect we are interested in. The higher the power, the higher the confidence in the observed difference; however a minimum power of 0.8 is needed. The power calculating software G*Power (*version 3.1.9.2*) (http://www.gpower.hhu.de/en.html) was used for power and sample size calculation in this study [293, 294].

2.9.2. T-Tests with Bonferroni Correction

T-test is a parametric test for the comparison of two group means. Independent sample t-tests are important tools for identifying differential protein expression in proteomics data. Means of each protein expression level (dependent variable) is compared in between groups (independent variable) to determine statistical significance. Type I errors (due to huge number of measurement) can be reduced by applying multiple testing correction. This can be done by adjustment of the false discovery rate (FDR) as in Benjamini- Hochberg's method, or by Bonferroni's correction which maintains family wise error rate (FWER). Independent sample t-tests with Bonferroni correction was used in this study to identify differentially expressed proteins between groups with high level of stringency. T tests were performed on Perseus and RStudio and visualized with volcano plots (**Fig. 2.14**).

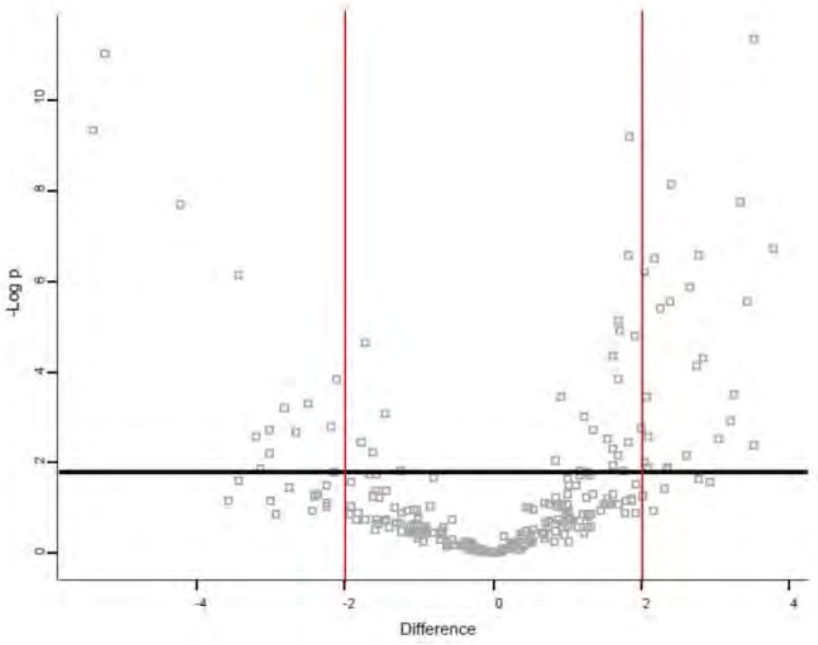


Figure 2.14. Example of volcano plot showing differentially expressed proteins (±2SD)

2.9.3. Analysis of Variance (ANOVA)

This is a parametric method used to test for statistically significant differences in two or more group means. This test was employed when comparing more than two group means, as using a t-test in such scenario would increase the chances of type I error. ANOVA was performed using Perseus software, and differentially expressed proteins between PCa, BPH and NC groups were identified.

2.9.4. Kruskal-Wallis Test

This is a non-parametric equivalent of ANOVA for identifying statistically significant differences between multiple groups (greater than two). This test was employed for identifying statistically significant differences in protein expression between ethnic groups in our PCa cohort, since the data was found not to be normally distributed.

2.9.5. Venn Diagram Analysis

Venn diagram is a mathematical mode of data representation; whereby circles are used to depict association or the lack thereof between different groups know as sets, drawn from a rectangular set known as the universal set. This tool is highly useful for preliminary assessment of unique and shared elements between study groups. A free online based Venn plotting software known as Venny (http://bioinfogp.cnb.csic.es/tools/venny/) was used to perform Venn diagram analyses in this study [295].

2.9.6. Hierarchical Clustering

Hierarchical cluster of analysis is a pattern detection method in proteomics which can be carried out using a bottom-up (agglomerative) or a top-down (divisive) approach. Hierarchical clustering data is typically depicted using a dendrogram generated from an iterative process of comparing similarity scores assigned to individual proteins in the dataset. Hierarchical clustering can be performed to identifying group differences, either by using training sets (supervised) or identifying random clustering patterns without the use of training set (unsupervised). Unsupervised hierarchical clustering was performed with Perseus on *z*-scored normalization of logarithimized data, using Euclidian distances; and on Cluster Treeview (*version 3.0*) using *k*-mean clustering.

2.9.7. Principal Component Analysis

Principal component analysis is a multivariate test in which potentially correlated variables are orthogonally transformed to linear uncorrelated sets of variables (eigenvectors of the covariance matrix) known as principal components. The first principal component always has the largest variance and subsequent principal components follow decreasing trend of variance. This analysis can be performed in one (1-D or loading plot), two (2-D) or three dimension (3-D) to identify presence or absence of clustering of variables within each group. Principal component analysis was performed using Perseus software on log-2 transformed data without *z*-scoring normalization.

CHAPTER 3: MASS SPECTOMETRY-BASED DISCOVERY OF NOVEL POTENTIAL BIOMARKERS OF PROSTATE CANCER

3.1. Background

PCa has the second highest incidence among cancers [296], and there are established disparities in PCa progression based on ethnicity [297, 298]. Even though factors for its aggressive course are largely unknown, fusion translocation of TMPRSS2-ERG and AKT hyperphosphorylation have been demonstrated as ominous prognostic factors in about 50% of all PCa patients [299]. New potential biomarkers of PCa have been described [300-303] in Caucasian populations and their significance in African populations remains unclear. Despite its invasiveness, biopsy is still the commonest approach to confirmation of the diagnosis of PCa [304]. In addition, tampering surgically with a malignant lesion may lead to further seeding of the cancer cells [305]. Prostate specific antigen (PSA) and Acid phosphatase (ACPP) are inherently inconsistent in sensitivity and specificity [304]; acid phosphatase is not specific for PCa [72] and PSA testing is plagued with inconsistencies [306, 307]. Even when combining TRUS, DRE, and PSA assessment, ca. 40% of all PCa cases would be diagnosed as localized when they are already extraprostatic [308]. Prior to the advent of mass spectrometry based proteomics methodologies, sequencing of proteins using Edman degradation was quite laborious and limited in capacity [309]. Differential proteomics aims to identify changes in specific proteomes that might correlate with disease status, disease progression, and response to cancer therapy. Usefully, proteomics techniques can be readily applied to urine specimens containing biomarkers that are directly deposited into the genitourinary tract [92]. Clinical proteomics has been used to elucidate the pathogenesis of disease; as well as to improve the diagnosis and monitoring of cancer and other infective diseases [85, 93, 310-314]. Furthermore, since urine represents a readily available and sterile specimen, diagnostic markers identified in urine could provide the basis of noninvasive assays for detection, early prevention and control of disease [308]. Ethnic-related biomarker disparity studies are lacking in African cohorts. Indeed, in spite of the extensive research on PCa to date, there is an urgent need for novel biomarker identification for prompt diagnosis and prognosis.

3.2. Methodology

3.2.1. Clinical Sample retrieval and preservation

In this study, first-stream urine was collected because it has been reported to contain more prostatic secretions in comparison to last-stream or mid-stream urine [301]. Even tough expressed prostatic secretion (EPS) generated by massaging the prostate has been reported to increase the coverage of biomarker pick-up [315-317], it was not used in this study since we were looking for the least invasive approach to diagnosing PCa. Besides, performing a second DRE after prior physical clinical examination would be unethical. Full ethical consent was obtained from all participants as described earlier in **Section 2.1** before sample collection. Patients suffering from other comorbidities like hypertension and Diabetes mellitus were excluded because the risk of developing essential arterial hypertension rises in older patients and has been reported to be higher in African populations [318]. All black African PCa patients in our cohort were confirmed to be normotensive.

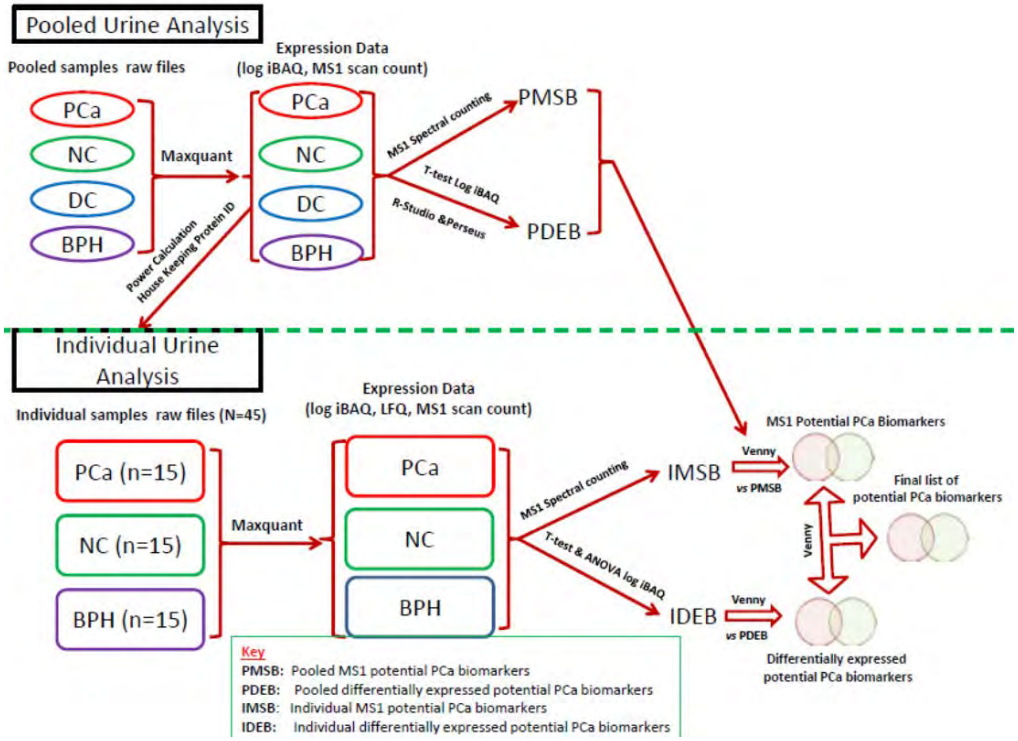


Figure 3.1. Two stage analysis design of mass-spectrometry based biomarker study [319]

The research was carried out with strict adherence to the ethical principles for medical research as prescribed by the Declaration of Helsinki (DoH) 2008 [197]. PCa patients were graded with the Gleason scale [302]. Demographic data such as age, race, as well as the (PSA) level were recorded **(Table 3.1).**

Sample	Number (n)	Individual/ pooled	Age	Race	PSA level	Gleason Score
PC1	1	Individual	61	M	7.8	3+3
PC2	1	Individual	74	M	1.7	3+3
PC3	1	Individual	71	M	0.4	3+4
PC4	1	Individual	61	B	3.9	4+5
PC5	1	Individual	74	M	31.0	3+4
PC6	1	Individual	61	M	27.4	3+3
PC7	1	Individual	67	M	185.0	4+5
PC8	1	Individual	80	B	49.0	3+4
PC9	1	Individual	70	M	113.0	4+5
PC10	1	Individual	54	W	9.6	5+4
PC11	1	Individual	68	M	100.0	4+5
PC12	1	Individual	80	M	30.1	5+5
PC13	1	Individual	69	W	5.9	3+4
PC14	1	Individual	64	W	9.1	2+3
PC15	1	Individual	81	B	10.5	2+2
PCa	9	Pooled	NA	NA	1.7-185.0	(3+3) to (5+5)
BPH	9	Pooled	NA	NA	0.7-48.4	NA
NC	9	Pooled	NA	NA	1.0-2.6	NA
DC	9	Pooled	NA	NA	NA	NA
BPH: Benign Prostatic Hyperplasia, PCa: Prostate Cancer, NC: Normal Healthy Controls, DC: Disease Control, PC: Individual Prostate Cancer, M: Mixed Ancestry, B: Black, W: White, NA: Not Applicable						

Table 3.1. Clinicopathologic data for individual and pooled samples showed variability in Gleason's score and serum PSA levels.

3.2.2. Experimental design and pre-analytic phase

The design of the mass spectrometry based aspect of this study is a two stage design which involved initial analysis of pooled samples and subsequent analysis of individual samples. Results were then cross validated between experimental stages as well as with other urinary

databases. Overlap between protein groups identified were documented and analysed for potential biomarkers of prostate cancer as shown in the workflow (**Figure 3.1** *vide supra*). Before definitive analysis on the mass spectrometer, several sample preparation and optimizations steps were followed to ensure good quality of sample for mass spectrometry analysis. At the initial stages of the experimental workflow, proteomic analysis was based on urine samples from 36 patients, which were divided into 4 pools (i.e. BPH, NC, PCa, & DC). Each group was made up of nine 5ml samples which were all thawed and pooled into a single 50 mL tube. Upstream pooling as done here reduces the level of statistical variance observed as a result of technical errors during the sample processing steps. These pooled samples were processed as described in **Section 2.2**. . Although dynamic range problems are more marked in blood serum, high abundance proteins in urine may mask the detection of low abundance and often biological significant analytes. Hence, high abundance proteins such as human serum albumin, uromodulin (a.k.a Tamm-horsfall protein), and immounoglobulins in the supernatant were filtered out using a 50KDa centrifugal molecular weight cut-off (Amicon® Ultra) filter unit. After protein quantification and buffer exchange, protein proteolysis was carried out using Filter Aided Sample Preparation (FASP) as previously described elsewhere [208]. All organic solvents used after tryptic digestion contained 0.1% Triflouroacetic acid (TFA), which is a very good ion-pairing agent. Peptide samples were then transferred to StageTips for offline desalting, solid phase extraction (SPE), and sample clean-up as described elsewhere [209]. Samples were then lyophilized on a Centrivap centrifugal vacuum concentrator and re-suspended to 1µg/µl in HPLC grade water (Sigma-Aldrich, St. Louis, MO, USA) containing 0.1% Formic acid (FA) and stored at -80^{0}C until ready to run on the mass spectrometer. Similar pre-analytic processing was performed prior to individual sample analysis. An overview of the workflow for pooled sample analysis is shown in **Figure 3.2**.

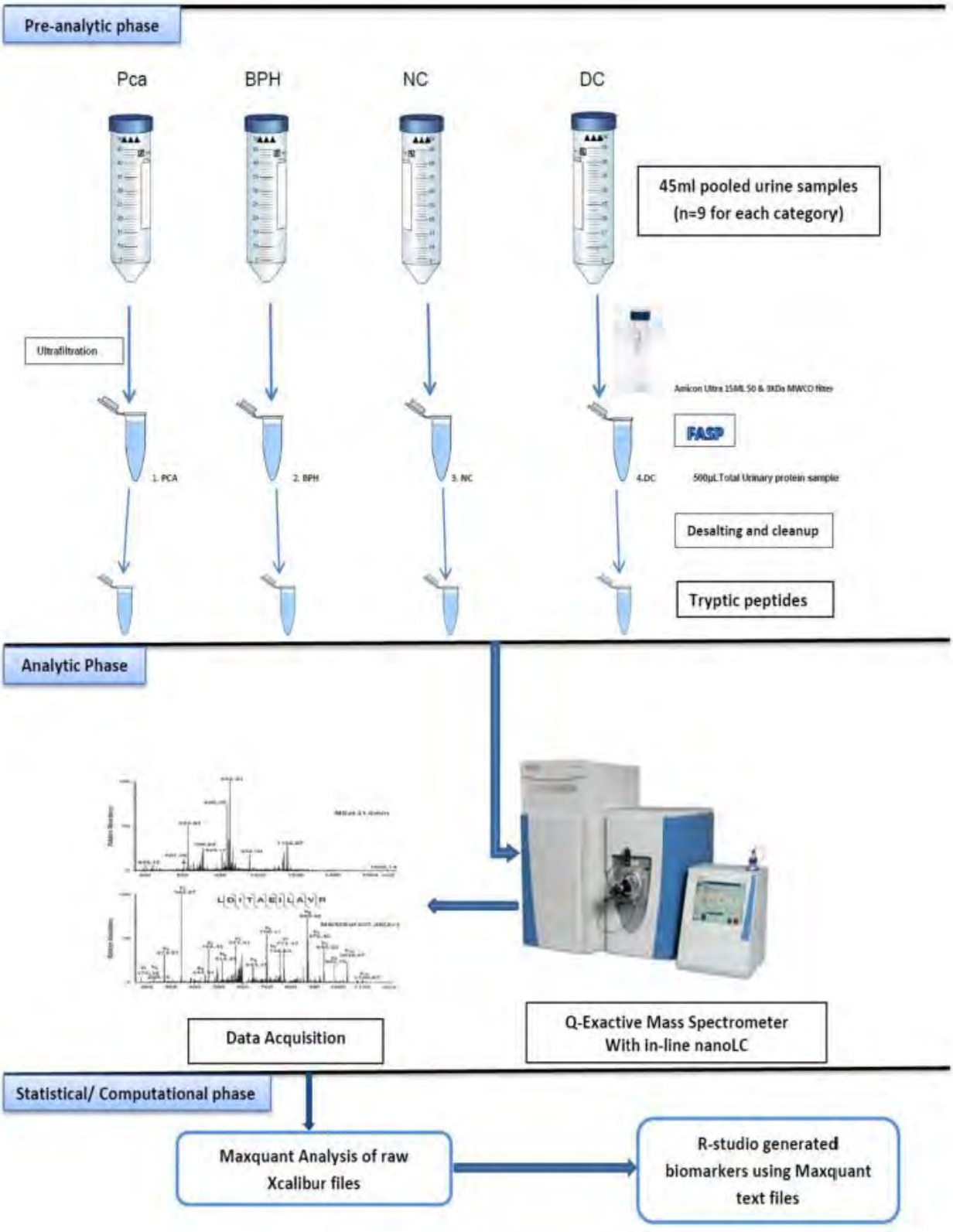


Figure 3.2. Pooled sample Label-free Urinary Proteomics workflow [319]

3.2.3. Ultra-High Performance Liquid Chromatography (UHPLC)

Ultra-high performance liquid chromatography (UHPLC) using a nano-flow system was done online on a Dionex UltiMate® 3500 RSNano UPLC systems (Thermo Fisher, San Jose, CA, USA) utilysing a C-18 reverse phase precolumn trap (100μm×5cm; 100 Å; 5cm; C-18) with an analytic column (75 μm×50cm; 100 Å; 5cm; C-18). Detailed description of the steps involved is contained in **Section 2.3.1**.

3.2.4. Mass Spectrometry Analysis

A bottom-up (shotgun) proteomics analysis was performed on a QExactive^{TM} Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher, San Jose, CA, USA) in-line with the Dionex UltiMate® 3500 RSNano UPLC systems (Thermo Fisher, San Jose, CA, USA). Flow injection analysis of samples introduced from the UHPLC system was done using the following system specifications as contained in **Section 2.3.3**.

3.2.5. Statistical and computational analysis of generated data

The Xcalibur raw* files generated from mass spectrometry experiments were quantitatively analysed using the Maxquant software (version 1.3.0.5) default settings with a few minor modifications. Carbamidomethylation of cysteine was specified as a fixed modification; and N-terminal acetylation and oxidation of methionine were selected as variable modification. Enzyme specification during the Andromeda search was specified as trypsin/P. Precursor mass tolerances and fragment ions mas tolerance were set at 20ppm and 0.5 Dalton respectively in initial scan and then set at 6ppm for the main search. Andromeda search engine which is integrated into Maxquant suite was used to perform tandem MS search, queries were run against target databases such as International Protein Index (IPI human version 3.87). Lower limit cutoff set for peptide length was 7 amino acids, and not more than two missed cleavage were permissible. Results generated from Maxquant were thereafter imported directly into Perseus software (version 1.4.0.20) for further computational analysis and visualization. False discovery rates (FDR) used for pooled and individual experiment analysis were 0.01 and 0.001 respectively.

3.3. Results

3.3.1. Pooled Sample Experimental Findings

We quantitatively analysed raw* files from "pooled" PCa, BPH, NC and DC samples with Maxquant software with integrated Andromeda search engine and found a total of 1102 protein groups. Of these, 217 (19.69%) were single peptide protein groups and 15 (1.36%) were only identified by modification site. There were 41 (3.72%) contaminants and 12 were identified as reverse hits. Total identified evidence at peptide level was 32,061 with peptide unique to a protein group being 5,069 (90.60%). Identified non-redundant peptides were 5,595 and reverse hits were 14. False discovery rates (FDR) were estimated to be 1.10% and 0.25% for protein groups and peptides respectively at FDR1 level; and 2.18% and 0.50% at FDR2 level (Table 2). FDRs for modified peptides, evidences and msms were 0.25%, 0.09% and 0.07% respectively at FDR1 level; and 0.50%, 0.17%, and 0.14% at FDR2 level respectively. A total of 596 modifications of peptides were observed which were mainly oxidation of methionine and N-terminal acetylation. There were no missed cleavages in 87.1%, one in 11.9% and 2 in 0.9% of the peptides identified. A summary of our pooled experimental findings is shown below (**Table 3.2**). The total ion chromatogram generated from the pooling experiment showed different profiles for the different groups PCa, BPH, DC and NC (**Figure 3.3**); and its even spread signifies a stable spray during measurement.

Pooled Experiment			
PG		**Peptides**	
Parameters	**Result**	**Parameters**	**Result**
Protein Groups	1102	**Identified Evidence**	32061
Single Peptide PG	217 (19.69%)	**Reverse Hits**	28
Only Identified by site	15 (1.36%)	**Identified Peptides (non-redundant)**	5595
Contaminants	41 (3.72%)	**Peptides Unique to a PG**	5069 (90.60%)
Reverse Hits	12	**Contaminants**	373 (6.68%)
FDR1	1.1	**Reverse Hits (non-Redundant)**	14

FDR2	2.18	**FDR1**	0.25
		FDR2	0.5
PG: Protein Groups, FDR: False Discovery Rate			

Table 3.2. Maxquant result summary of pooled experiment demonstrates that Protein groups compared with peptide unique to a protein group in pooled experiment was approximately 1:5 and level of contaminant was below 5% showing that the results generated were of good quality.

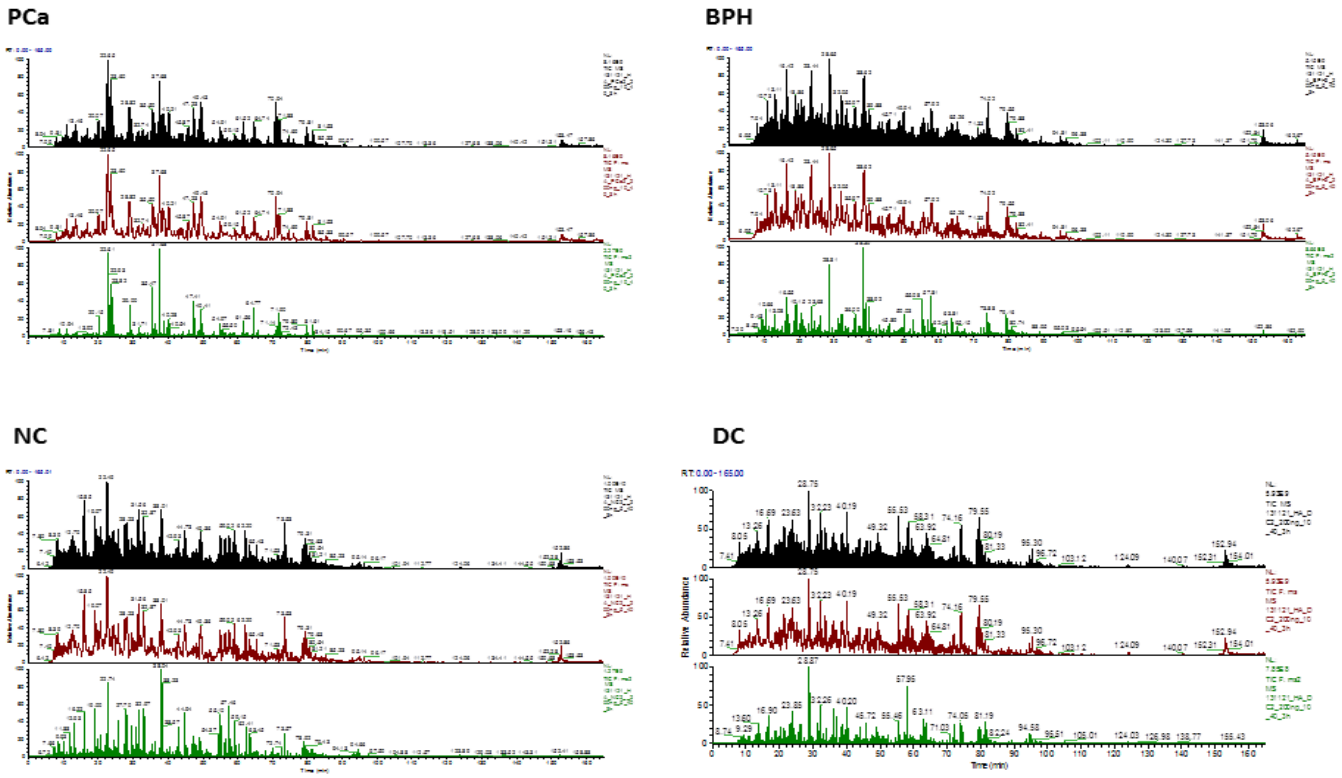


Figure 3.3. Total ion chromatogram showing differential profile between PCa, BPH, DC and NC both at MS1 and MS2 levels

3.3.2. Data Preprocessing for Statistical Computation

Output files generated from Maxquant for peptides, protein groups, and evidences in text (*.txt) format were processed from the "combined" folder of the Maxquant result. Spectral statistical analyses were performed on log-2 transformed expression data such as Intensity Based Absolute Quantification (iBAQ) values, raw Intensities, and Label free quantitation (LFQ). Depending on the total number of missing values per peptide, missing values in peptide expression were either

filtered out or imputed with values of intensity from the lower limit of the Gaussian distribution of peak intensities [320].

Computational analysis based on R-studio, involved the use of basic statistical packages (e.g. *gplots*, *Knitr*, *gdata*) which were download from the CRAN repository (www.cran.r-project.org/) and automatically installed; or manually with a *tar.gz* files. Following which the "working directory" is set and then R is used to read and assess information found in datasets. In this study, IBAQ values normalization was first performed by logarithmic transformation (log-2) prior to comparison. An initial filtering-out step was carried for proteins with over 50% missing values. After cleaning the data in this manner, data spread was then evaluated using box-whisker plots (**Figure 3.4**) and Q-Q plots (**Figure 3.5**).

The level of expression of 9 common housekeeping proteins from literature was further examined in our pooling experiment and observed that useful β-actin and β-2-microglobulin (B2M) where very dependable across all our experimental groups (**Figure 3.6**). This was also demonstrated by expression level in multiple samples as depicted in the protein profile plot shown below (**Figure 3.7**).

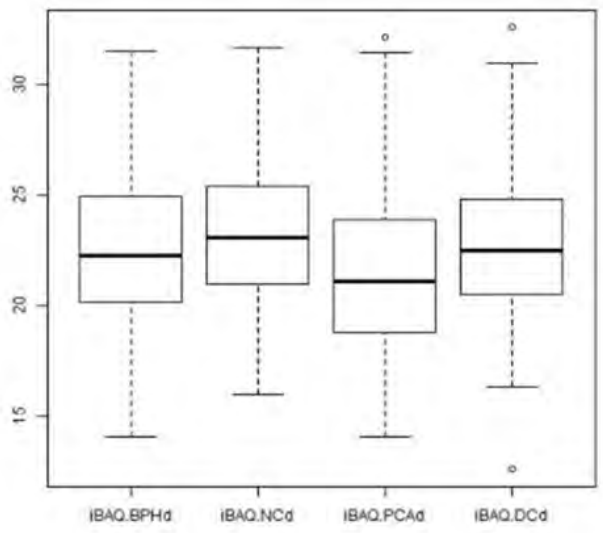


Figure 3.4. Box- whisker plot to assess the spread of data in each study group (BPH, NC, PCa &DC). On the x-axis are the different pooled groups and on the y-axis are log2 -transformed IBAQ values.

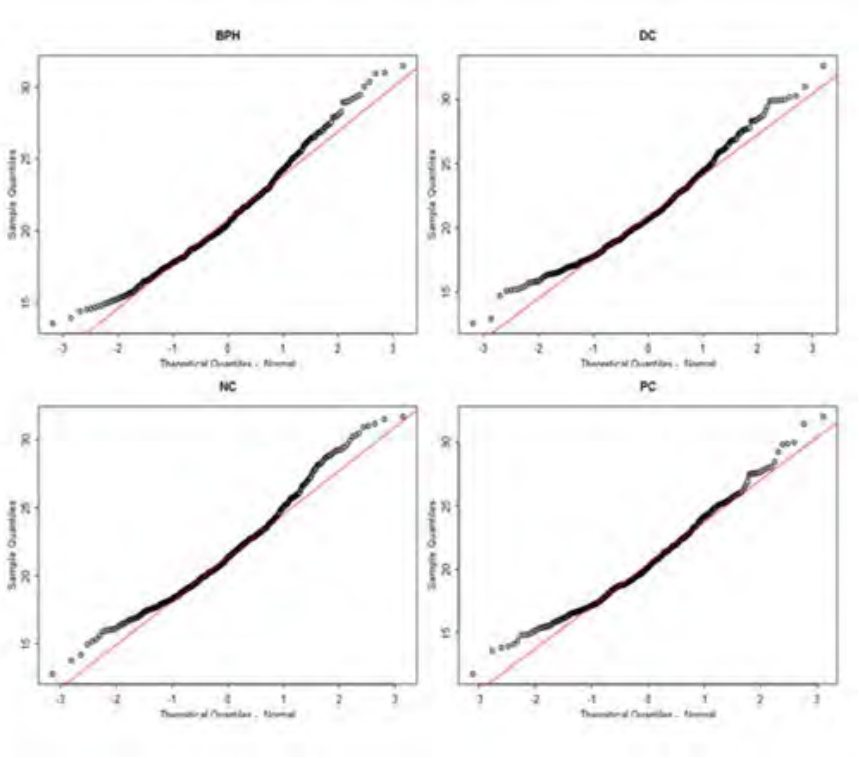


Figure 3.5. Q-Q plot used to assess data spread in PCa, BPH, DC and NC by comparing the experimental quantiles with theoretical quantiles. On the x-axis is the normal theoretical quantile while on the y-axis is the sample quantiles. The data points can be seen to be within acceptable distance of the reference line (red).

Uromodulin or Tamm-Horsefall protein, as expected when renal disease is not present, was found to be most abundant in individual and pooled normal healthy (NC) samples in comparison to PCa and BPH. In addition, we observed that despite uromodulin's dominance of the urinary proteome, more than 1,500 urinary unique proteins have been previously detected by mass-spectrometry based methods [321, 322]. Not least, quantitative disease-associated changes in the urinary proteome have been previously reported in literature, even when the disease site is remote from the genitourinary tract [313].

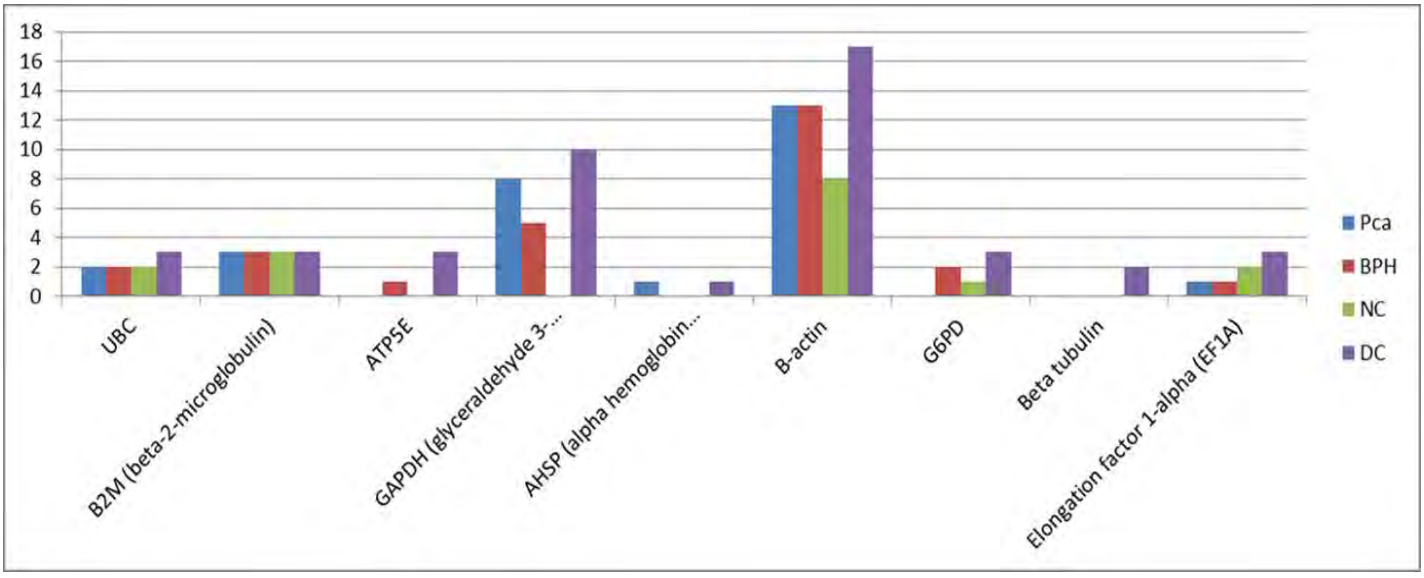


Figure 3.6. House-keeping protein profile revealed that β2-microglobulin (B2M) and β-actin were stably expressed across all 4 pooled groups. Nine common housekeeping proteins from literature were selected. On the x-axis are the housekeeping proteins and on the y-axis are the log2-transformed IBAQ intensities for the protein groups.

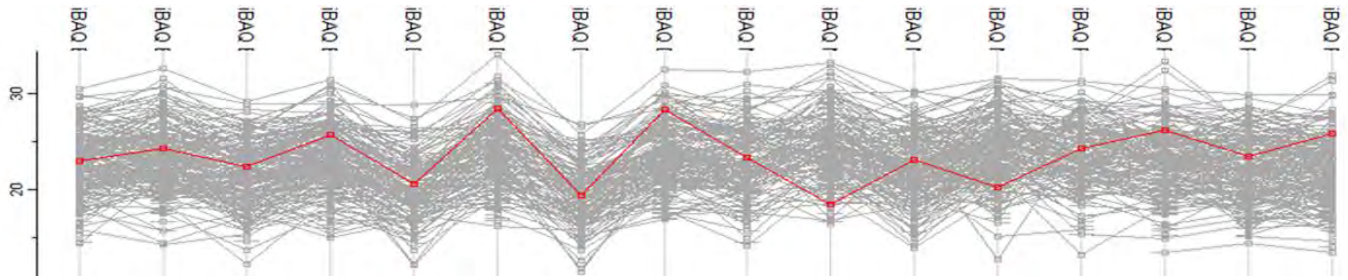


Figure 3.7. Protein profile shows β-actin to be stably expressed across duplicate pooled sample runs. On the x-axis are the pooled sample replicates and on the y-axis are log2-transformed IBAQ values.

3.3.3. Venn Diagram Analyses

Venny, a free online Venn diagram plotting software (link as described earlier) was used to plot a 4-way Venn diagram to assess the proteins unique to each, and the degree of overlap between them. Up to 63% of the identified peptides were found to be common to at least 3 out of the 4 pooled group which is in keeping with the reported proportion of the "core proteome" in literature [323]. Also from this plot 65, 19, 44 and 42 proteins groups were found to be unique NC, PCa, BPH and DC respectively and this unique protein sets represent an excellent niche for investigation of potential biomarkers (**Figure 3.8**).

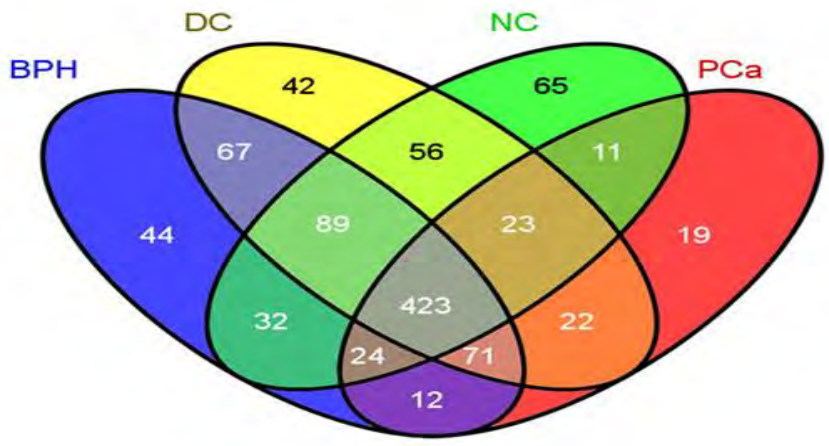


Figure 3.8. Venn diagram showing unique and overlapping protein sets in BPH, DC, NC, and PCa

3.3.4. Intensity- based Absolute Quantification (iBAQ) and MS1 Analysis

IBAQ-based MS1 quantification [324] was carried out on for precursor-ion mass analysis. Using this approach, total intensity of protein was normalized by length of peptide. Integration of the area under the MS1 peak for the top scoring peptides which were more abundant in the PCa pool as compared with DC, NC, and BPH pools were documented. Conversely, peptides found to be absent or less abundant were also identified by this method. A total of 20 protein groups were found to be potential biomarkers of PCa by this method (**Table 3.3**).

Candidate Biomarkers of PCa by MS1 spectral quantification

More Abundant in PCa (≥2-folds)	Less Abundant in PCa (≤2-folds)
sp|P01023|A2MG_HUMAN Alpha-2-macroglobulin	tr|E7EPZ9|E7EPZ9_HUMAN Tenascin-X
sp|P01024|CO3_HUMAN Complement C3	sp|Q99715|COCA1_HUMAN Collagen alpha-1(XII) chain
sp|P04114|APOB_HUMAN Apolipoprotein B-100	sp|P12109|CO6A1_HUMAN Collagen alpha-1(VI) chain
sp|P00918|CAH2_HUMAN Carbonic anhydrase 2	sp|P24821|TENA_HUMAN Tenascin
sp|P08603|CFAH_HUMAN Complement factor H	sp|P12111|CO6A3_HUMAN Collagen alpha-3(VI) chain
sp|P02649|APOE_HUMAN Apolipoprotein E	sp|P14543|NID1_HUMAN Nidogen-1
tr|Q32Q12|Q32Q12_HUMAN Nucleoside diphosphate kinase	
sp|P02775|CXCL7_HUMAN Platelet basic protein	

OS=Homo sapiens GN=PPBP PE=1 SV=3	
sp|P24666|PPAC_HUMAN Low molecular weight phosphotyrosine protein phosphatase	
sp|P30041|PRDX6_HUMAN Peroxiredoxin-6	
sp|P19827|ITIH1_HUMAN Inter-alpha-trypsin inhibitor heavy chain H1	
sp|P37840|SYUA_HUMAN Alpha-synuclein	
tr|G3V2F7|G3V2F7_HUMAN HCG2044781 (Ubiquitin-conjugating enzyme E2 variant 1)	
sp|P04003|C4BPA_HUMAN C4b-binding protein alpha chain	

Table.3.3. List of more and less abundant protein groups by MS1 spectral quantification

Even from this initial binary MS1 spectral quantification result, it was observed that that protein groups that were found to be less abundant included collagens and other extracellular matrix molecules such as Tenascin and Nidogen-1; which made biologic sense because downregulation of extracellular matrix molecules is required for the mobility of cancer cells from the surface layer through the basement membrane into the lamina propria [325, 326].

Furthermore, we carried out an independent sample t-test with Bonferroni's correction for multiple testing was performed using iBAQ values that were log-transformed (**Figure 3.9**). This was done to find differential peptide iBAQ values in order to identify differentially expressed peptides between PCa and NC. The dependent variables were normalized iBAQ values for proteins, while the independent variables were PCa and NC. Using this method, 17 proteins were

found to be differentially expressed (±2SD) as shown by **Table 3.4**.

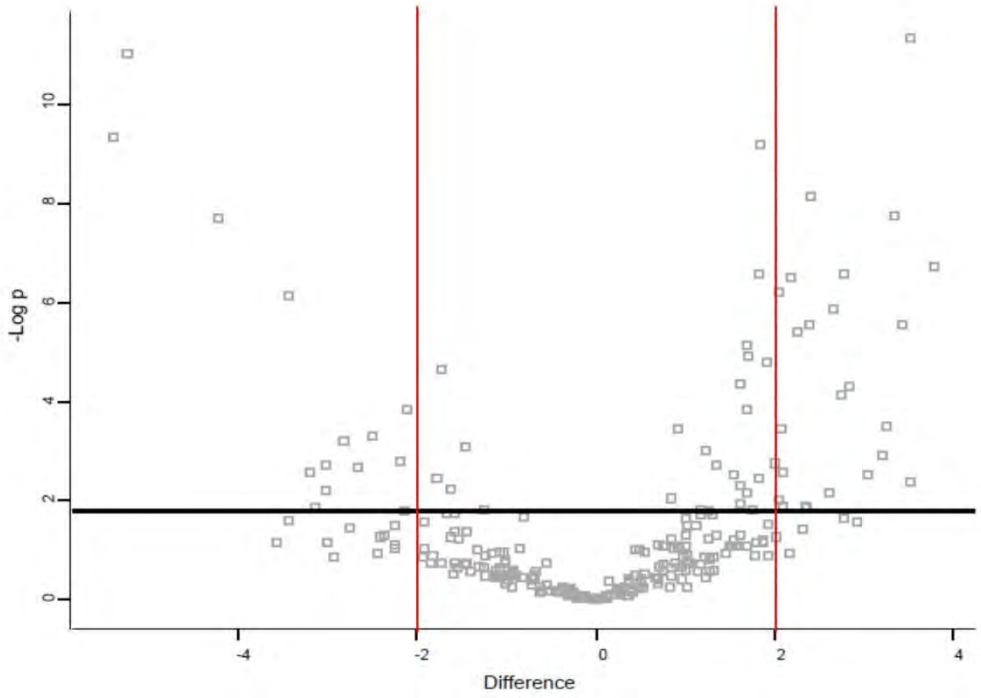


Figure 3.9. Volcano plot of differentially expressed (±2SD) protein groups between PCa and NC

Candidate Biomarkers of Pca by IBAQ (±2SD)
P02656 (APOC3_HUMAN) Apolipoprotein C-III variant 1
P00738 (HPT_HUMAN) Haptoglobin
P00915 (CAH1_HUMAN) Carbonic anhydrase I
P01009 (A1AT_HUMAN) Alpha-1-antitrypsin
P01876 (IGHA1_HUMAN) Ig alpha-1 chain C region
P02647 (APOA1_HUMAN) Apolipoprotein A1
P02675 (FIBB_HUMAN) Fibrinogen beta chain
P02750 (A2GL_HUMAN) Leucine-rich alpha-2-glycoprotein
P06702 (S10A9_HUMAN) Protein S100-A9
P07737 (PROF1_HUMAN) Profilin-1 (Epididymis tissue protein Li 184a)
P19961 (AMY2B_HUMAN) Alpha-amylase 2B
P30043 (BLVRB_HUMAN) Flavin reductase (NADPH)
P30740 (ILEU_HUMAN) Leukocyte elastase inhibitor
P61626 (LYSC_HUMAN) Lysozyme C
P62937 (PPIA_HUMAN) Peptidyl-prolyl cis-trans isomerase A (Cyclophilin A)

P68871 (HBB_HUMAN) Hemoglobin subunit beta
P69905 (HBA_HUMAN) Hemoglobin subunit alpha

Table 3.4. List of 17 differentially expressed protein groups by t-test

Using hierarchical clustering of the protein groups *versus* pooled PCa, NC, BPH and DC groups, a similar molecular signature was found for DC and BPH groups. However, there was a distinct signature for NC which was completely different for PCa. The BPH and DC signature looks a bit more like the PCa signature as compared with the NC as shown in the heat map presented below (**Figure 3.10**).

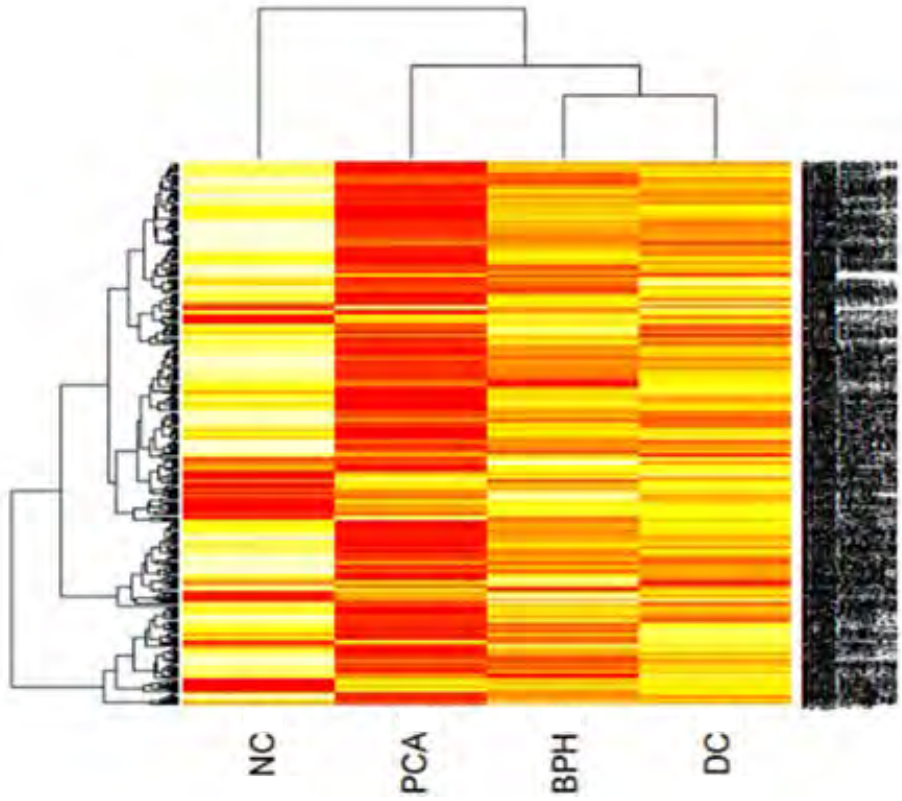


Figure 3.10. Hierarchical clustering of PCa, BPH, NC and DC

Even though it is widely known that prostatitis and some other pathologies of the genitourinary tract which formed the DC group in this study can be responsible for false positive in PSA results when trying to diagnose prostate cancer; our hierarchical clustering result interestingly differentiated them quite clearly, possibly due to the high mass accuracy and resolution of the mass spectrometry instrument used.

3.3.5. Power Calculation and Sample Size Determination

Power calculation is highly essential in the determination of the extent to which differences between the different study-groups could be identified, if indeed they existed. Based on the statistics generated from the preliminary pooled sample experiments, we calculated the required total number of individual samples (N) needed to attain a statistical power (1-β) of 0.8, and found it to be 43 individual samples (**Figure 3.11**). Individual sample analyses was therefore performed on 45 samples comprised of three subgroups (n=15) which were PCa, NC and BPH. DC was eliminated because it had similar molecular signature with BPH and technically speaking, BPH is also a form of disease control.

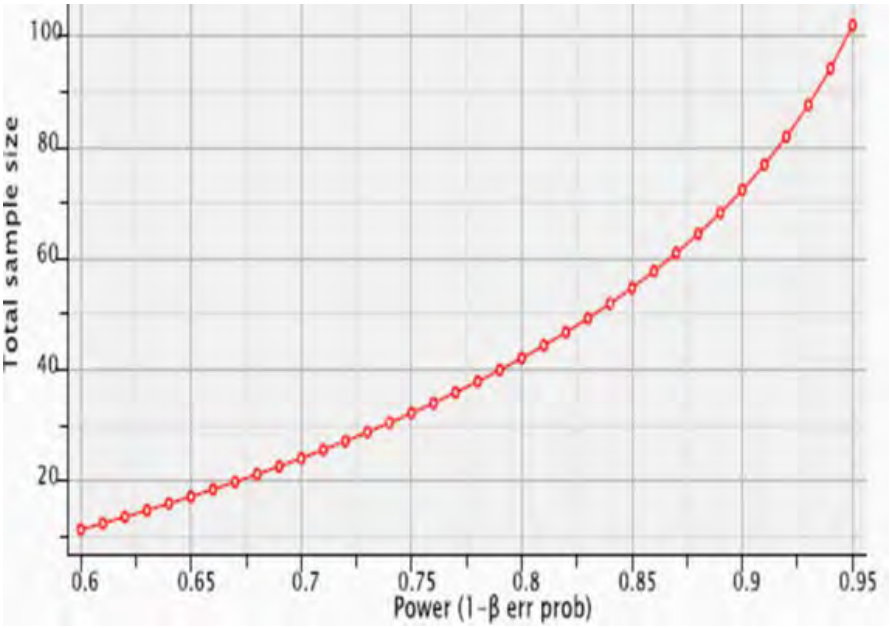


Figure 3.11. Power plot showing that 43 samples were required to attain a statistical power of 0.8. This is a power curve generated automatically from G*Power using statistical parameters generated from the initial pooling experiment

3.3.6. Individual Analysis of PCa, BPH and NC Samples

Using 45 individual samples (PCa (n)=15, BPH (n)=15, NC (n)=15), quantitative analysis was performed with Maxquant software with integrated Andromeda search engine. A total of 1,545 protein groups were identified in all experiments (N=45). Of these, 206 (13.33%) were single peptide protein groups and 20 (1.29%) were only identified by modification site. There were 41

(2.65%) contaminants and 19 were identified as reverse hits. Total identified evidence at peptide level was 122,257 with peptide unique to a protein group being 9,191 (91.99%). Identified non-redundant peptides were 9,991 and reverse hits were 29. False discovery rates (FDR) were estimated to be 1.25% and 0.29% for protein groups and peptides respectively at FDR1 level; and 2.46% and 0.58% at FDR2 level. FDRs for modified peptides, evidences and msms were 0.28%, 0.09% and 0.06% respectively at FDR1 level; and 0.57%, 0.19%, and 0.13% at FDR2 level respectively (**Table 3.5**). A total of 809 modifications of peptides were observed which were mainly oxidation of methionine and N-terminal acetylation. There were no missed cleavages in 81.6%, one in 16.6% and 2 in 1.8% of the peptides identified. False discovery rates (FDR) for proteins, peptides, site and **A**ll **I**on **F**ragmentation (AIF) threshold were at 0.01%. Maximum peptide **P**osterior **E**rror **P**robability (PEP) was at 1%.

Individual Experiment			
PG		**Peptides**	
Parameters	**Result**	**Parameters**	**Result**
Protein Groups	1545	**Identified Evidence**	122257
Single Peptide PG	206 (13.33%)	**Reverse Hits**	114
Only Identified by site	20 (1.29%)	**Identified Peptides (non-redundant)**	9991
Contaminants	41 (2.65%)	**Peptides Unique to a PG**	9191 (91.99%)
Reverse Hits	19	**Contaminants**	465 (4.65%)
FDR1	1.25	**Reverse Hits (non-Redundant)**	29
FDR2	2.46	**FDR1**	0.29
		FDR2	0.58
PG: Protein Groups, FDR: False Discovery Rate			

Table 3.5. Summary of individual sample mass spectrometry runs.

Protein profile plots demonstrated that β-2-microglobulin (**Figure 3.12**) and β–actin (**Figure 3.13**) were stably expressed across multiple samples in the individual sample mass spectrometry runs.

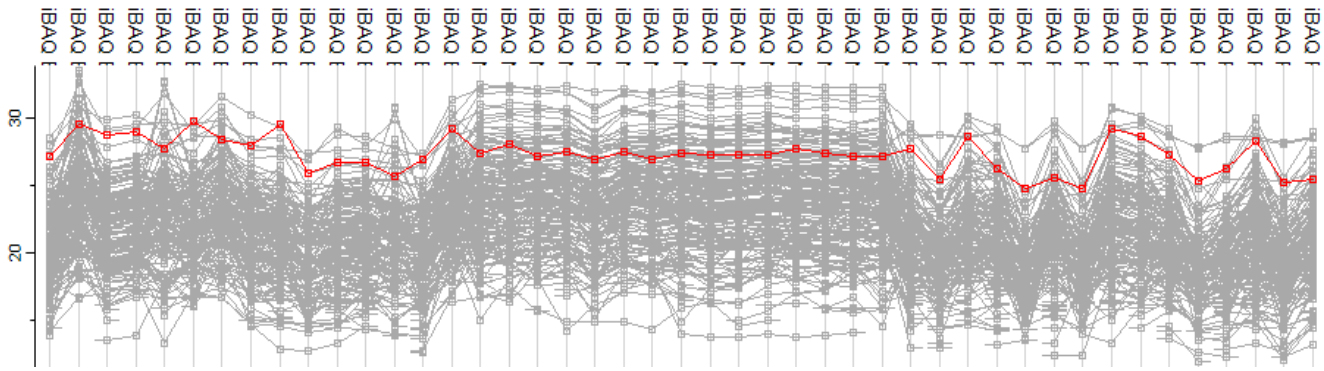


Figure 3.12. Protein profile plot demonstrated β-2-microglobulin to be stably expressed across multiple individual samples

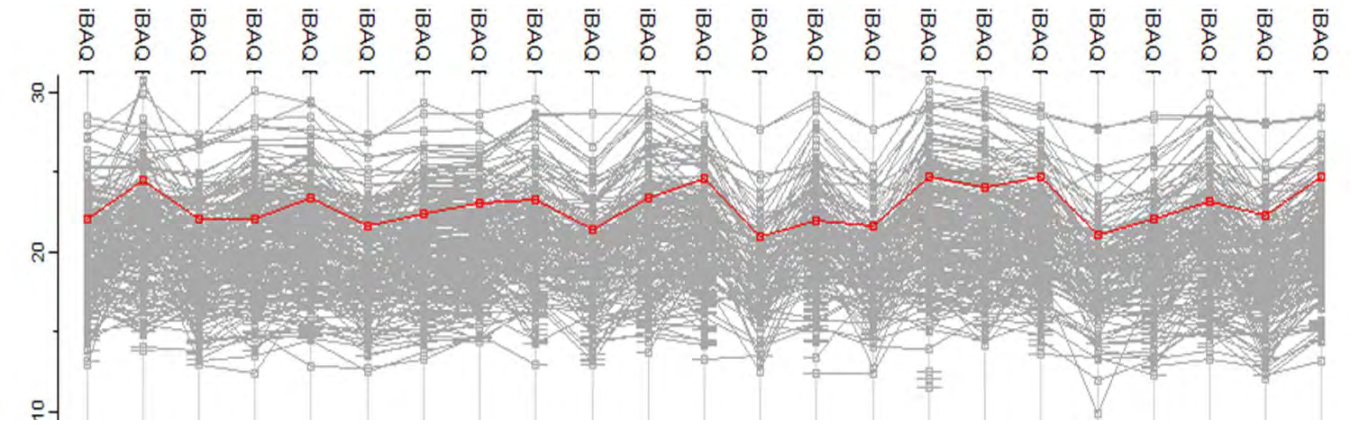


Figure 3.13. Protein profile plot demonstrates β–actin to be stably expressed across multiple individual samples

For quality control, three technical replicate samples were randomly selected from the 45 individual samples and repeatedly injected and then analyzed to assess the accuracy of the mass spectrometer. It was observed from this quality control check that the coefficient of variation difference ranged from 0.0009 to 0.0026. This high degree of correlation between these replicate samples was also observed by scatterplot (**Figure 3.14**), multivariate principal component analysis (**Figure 3.15**), and unsupervised hierarchical clustering (**Figure 3.16**).

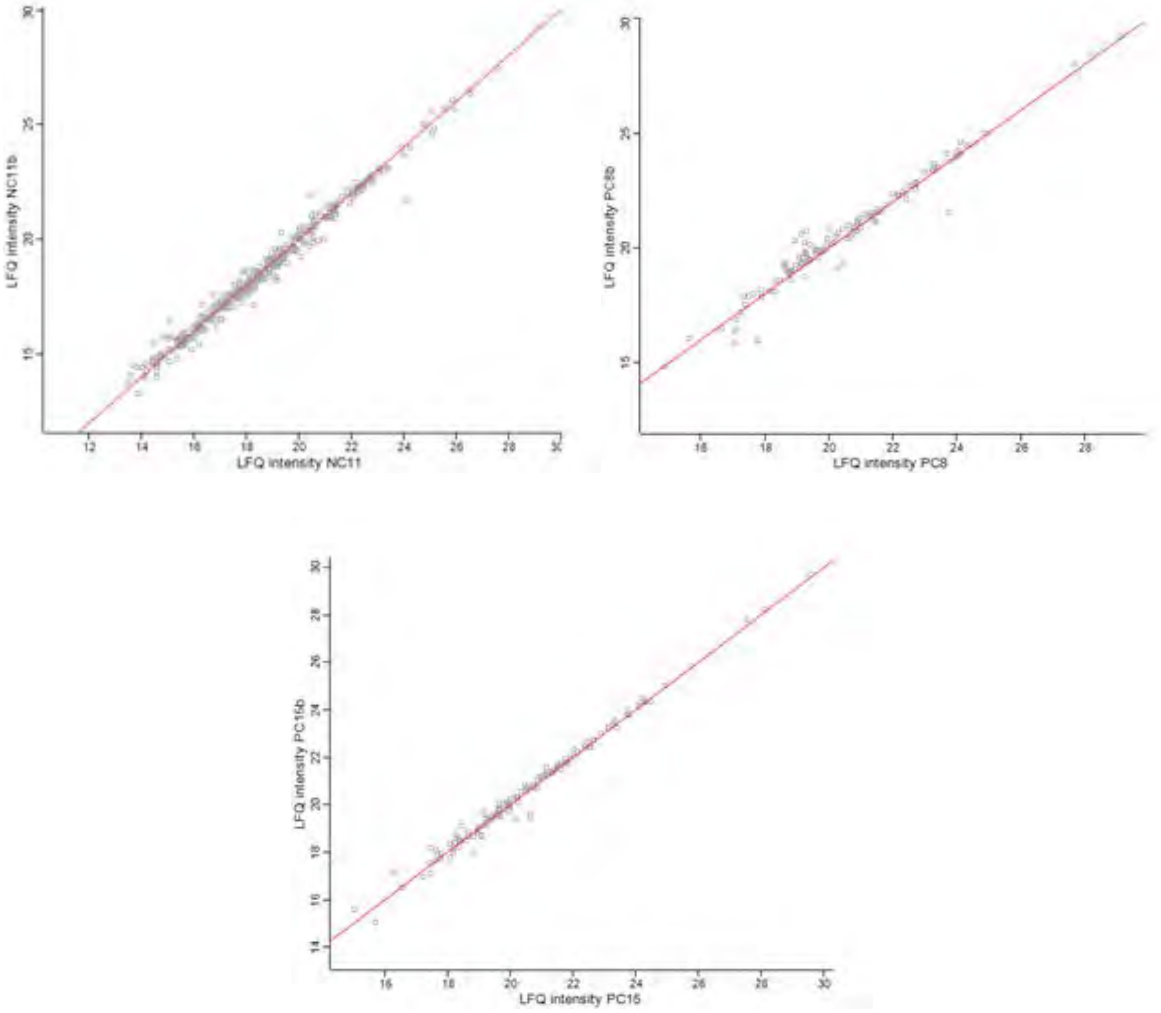


Figure 3.14. High correlation between repeat injection of randomly selected samples (NC11, PC8 and PC15)

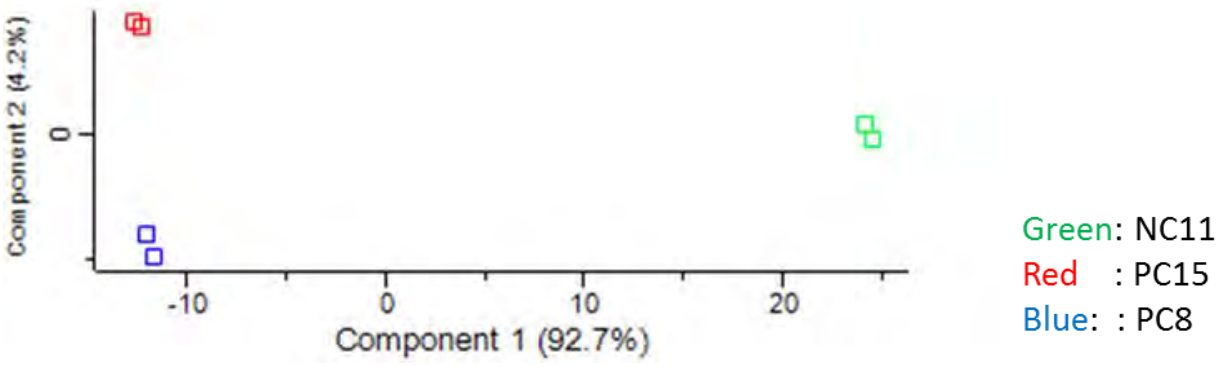


Figure 3.15. High level of correlation between NC11, PC15 and PC8 demonstrated by principal component analysis

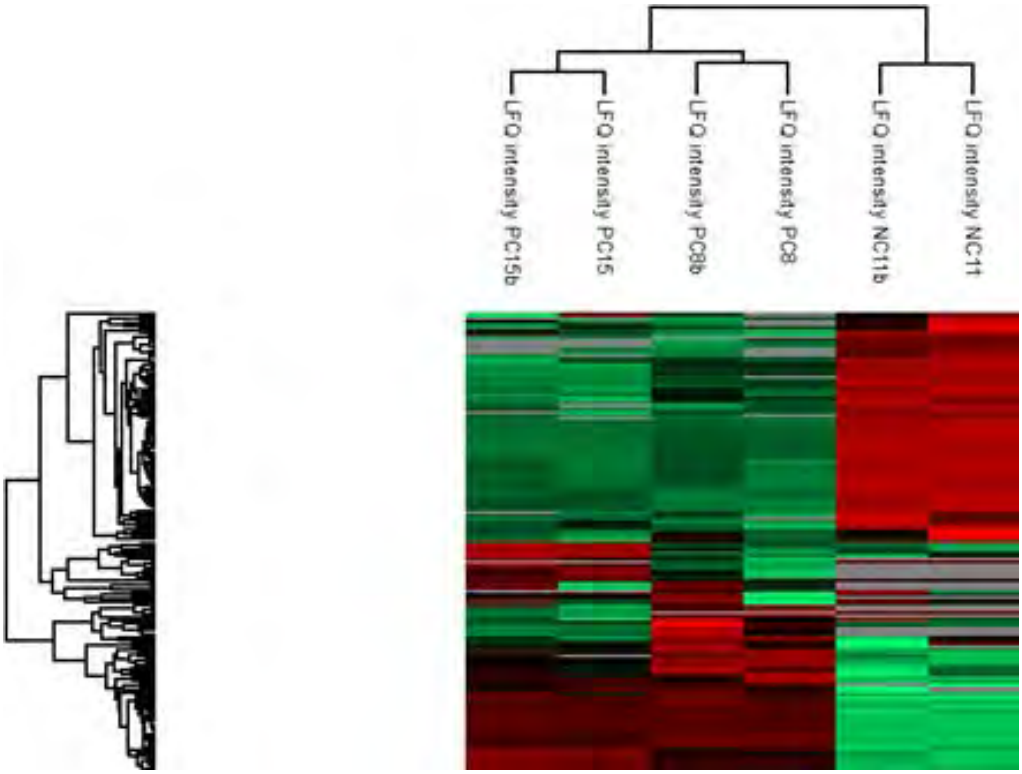


Figure 3.16. Correlation between repeat injection technical replicate sample (NC11, PC15 and PC8) demonstrated by Hierarchical clustering

3.3.7. Urinary Proteome Covered for South African PCa Patients

When the "pooled" and "Individual" proteomics experiments were combined, a total of 1788 proteins were identified of which 832 protein groups were from PCa samples (**Annexure I**). This identified catalog of proteome coverage discovered in these experiments included reverse Identifications (IDs), isoforms, and a few contaminants of potential human origin.

3.3.8. Potential Prostate Cancer Biomarker Characterization

Individual analysis of patient sample was performed to improve the proteome coverage and to increase the number of potential PCa biomarkers to be tested by **P**arallel **R**eaction **M**onitoring (PRMs). The relative abundance of MS1 precursor ion spectral count normalized by the iBAQ method was compared between PCa, BPH and NC in a binary mode. Using this method, 48 protein groups were found to be more abundant in PCa and 32 were less abundant as compared with NC and PCa. This phase of the individual sample analysis was carried out with stringent

visual assessment of performance of peptides across the individual samples. Using label free quantification (LFQ) values and iBAQ, differentially expressed proteins were also assessed using the Perseus Software (version 1.4.0.20). Logarithmically normalized LFQ and iBAQ values were used to perform two independent t-tests with Bonferroni's correction for "PCa *vs* BPH" and "PCa *vs* NC" and found differentially expressed proteins between these groups. For further validation, a one-way analysis of variance (ANOVA) was performed for BPH, PCA, and NC for identification of peptide-level variance between these groups. In addition, Venn diagrams were used to identify overlap between differentially expressed peptides identified by independent sample t-test and ANOVA. A total of 157 differentially expressed peptides were found to overlap between independent sample t-test and ANOVA. On further stringency increase, by lowering the false discovery rate (FDR) from 0.01 to 0.001 using the Perseus software, and eliminating elements already found by MS1 spectral quantification; a total of 45 protein groups were found unique to differential expression analysis and 28 were found unique to the MS1 spectral peak integration analysis (**Table 3.6**). Summing up, 73 protein groups were found to be potential biomarkers of prostate cancer by both MS1 spectral quantification and by differential expression analysis.

S/N	Potential Biomarker ID	Protein Name	MS1 /LFQ	PEP Scores
1	P00915	Carbonic anhydrase 1	LFQ	1.86E-138
2	P69905	Hemoglobin subunit alpha	LFQ	0
3	B0YIW2	Apolipoprotein C-III	LFQ	4.32E-64
4	P00738	Haptoglobin	LFQ	0
5	P06702	Protein S100-A9	LFQ	0
6	P30043	Flavin reductase (NADPH)	LFQ	1.54E-17
7	P68871	Hemoglobin subunit beta	LFQ	0
8	Q6GTX8	Leukocyte-associated immunoglobulin-like receptor 1	LFQ	2.64E-304

9	B1AVU8	Saposin-D	LFQ	3.02E-171
10	P0CG48	Polyubiquitin-C	LFQ	0
11	P01133	Pro-epidermal growth factor;Epidermal growth factor	LFQ	0
12	Q07654-2	Trefoil factor 3	LFQ	2.52E-61
13	E9PR17	CD59 glycoprotein	LFQ	0
14	F8W7I2	Prostaglandin-H2 D-isomerase	LFQ	0
15	H9KV70	Neutrophil gelatinase-associated lipocalin	LFQ	0
16	O00187	Mannan-binding lectin serine protease 2 A chain	LFQ	0
17	P04155	Trefoil factor 1	LFQ	9.93E-101
18	P04406	Glyceraldehyde-3-phosphate dehydrogenase	LFQ	2.54E-07
19	P04430	Ig kappa chain V-I region BAN	LFQ	3.46E-120
20	P04745	Pancreatic alpha-amylase	LFQ	0
21	P05451	Lithostathine-1-alpha	LFQ	9.99E-239
22	M0R1F0	Prostate-specific antigen	LFQ	3.13E-101
23	P07998	Ribonuclease pancreatic	LFQ	0
24	P08123	Collagen alpha-2(I) chain	LFQ	7.71E-23
25	P10153	Non-secretory ribonuclease	LFQ	0
26	P10451	Osteopontin	LFQ	0
27	P11684	Uteroglobin	LFQ	2.44E-68
28	P15309-2	Prostatic acid phosphatase;PAPf39	LFQ	0
29	P16401	Histone H1.5	LFQ	3.93E-22

30	P17900	Ganglioside GM2 activator	LFQ	3.71E-88
31	P35555	Fibrillin-1	LFQ	0
32	P35754	Glutaredoxin-1	LFQ	4.25E-97
33	P36957	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	LFQ	6.86E-52
34	P60022	Beta-defensin 1	LFQ	6.80E-123
35	P61626	Lysozyme C	LFQ	0
36	G3V3D1	Epididymal secretory protein E1	LFQ	8.38E-225
37	P08571	Monocyte differentiation antigen CD14	LFQ	0
38	P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	LFQ	0
39	Q03403	Trefoil factor 2	LFQ	8.14E-160
40	Q14508	WAP four-disulfide core domain protein 2	LFQ	0
41	Q15828	Cystatin-M	LFQ	1.18E-281
42	Q96PD5-2	N-acetylmuramoyl-L-alanine amidase	LFQ	4.42E-138
43	Q9BZG9	Ly-6/neurotoxin-like protein 1	LFQ	7.06E-298
44	Q9H299;Q5T123	SH3 domain-binding glutamic acid-rich-like protein 3	LFQ	0
45	P02652	Apolipoprotein A-II;Truncated apolipoprotein A-II	LFQ	0
46	P07225	Vitamin K-dependent protein S	MS1	0
47	P00739-2	Haptoglobin-related protein	MS1	1.27E-292
48	P00918	Carbonic anhydrase 2	MS1	4.36E-10

49	P14543	Nidogen-1	MS1	7.43E-118
50	P01023	Alpha-2-macroglobulin	MS1	0
51	P08603	Complement factor H	MS1	0
52	P04003	C4b-binding protein alpha chain	MS1	3.36E-192
53	P19823	Inter-alpha-trypsin inhibitor heavy chain H2	MS1	0
54	F5H4W9	Serum paraoxonase/arylesterase 1	MS1	3.96E-95
55	H0YAC1	Plasma kallikrein	MS1	7.42E-27
56	P07357	Complement component C8 alpha chain	MS1	6.00E-22
57	P02788	Lactotransferrin	MS1	0
58	P04114	Apolipoprotein B-100;Apolipoprotein B-48	MS1	0
59	P19827	Inter-alpha-trypsin inhibitor heavy chain H1	MS1	0
60	P12814	Alpha-actinin-1	MS1	0
61	P13796	Plastin-2	MS1	0
62	Q99715	Collagen alpha-1(XII) chain	MS1	4.23E-156
63	P12109	Collagen alpha-1(VI) chain	MS1	0
64	P22105	Tenascin	MS1	5.66E-149
65	P12111	Collagen alpha-3(VI) chain	MS1	5.06E-177
66	P55287	Cadherin-11	MS1	6.06E-150
67	Q8WZ75	Roundabout homolog 4	MS1	7.90E-124
68	Q9UBR2	Cathepsin Z	MS1	7.84E-50
69	P54802	Alpha-N-acetylglucosaminidase	MS1	1.59E-90

70	P15169	Carboxypeptidase N catalytic chain	MS1	1.16E-09
71	P01880-2	Ig delta chain C region	MS1	3.40E-60
72	Q06033	Inter-alpha-trypsin inhibitor heavy chain H3	MS1	1.12E-32
73	P20742	Pregnancy zone protein	MS1	1.05E-240

LFQ: Label Free Quantification; MS1: Precursor ion spectral peak quantification; PEP: Posterior Error Probability; S/N: Serial Number

Table 3.6. List of discovered potential biomarkers of prostate cancer

Unsupervised hierarchical clustering was performed for "protein groups *vs* PCa and NC samples", and this showed distinct molecular signatures between them **(Figure 3.17).**

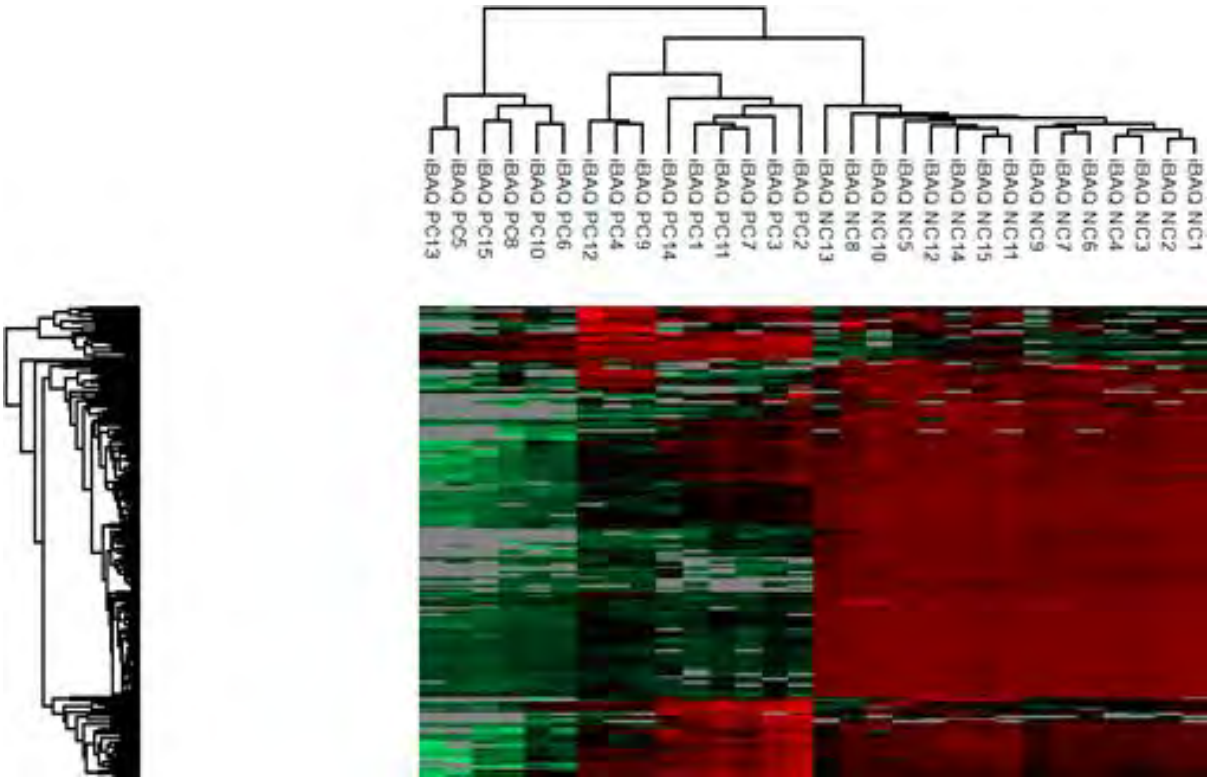


Figure 3.17. Heatmap shows different molecular signature between NC and PCa.

When BPH samples were included to the hierarchical clustering, it was observed that 4 of the BPH samples clustered closer to NC, while three of the PCa samples clustered closer to BPH.

These small overlaps warrant further research. Generally, NC, BPH and PCa samples clustered distinctly (**Figure 3.18**).

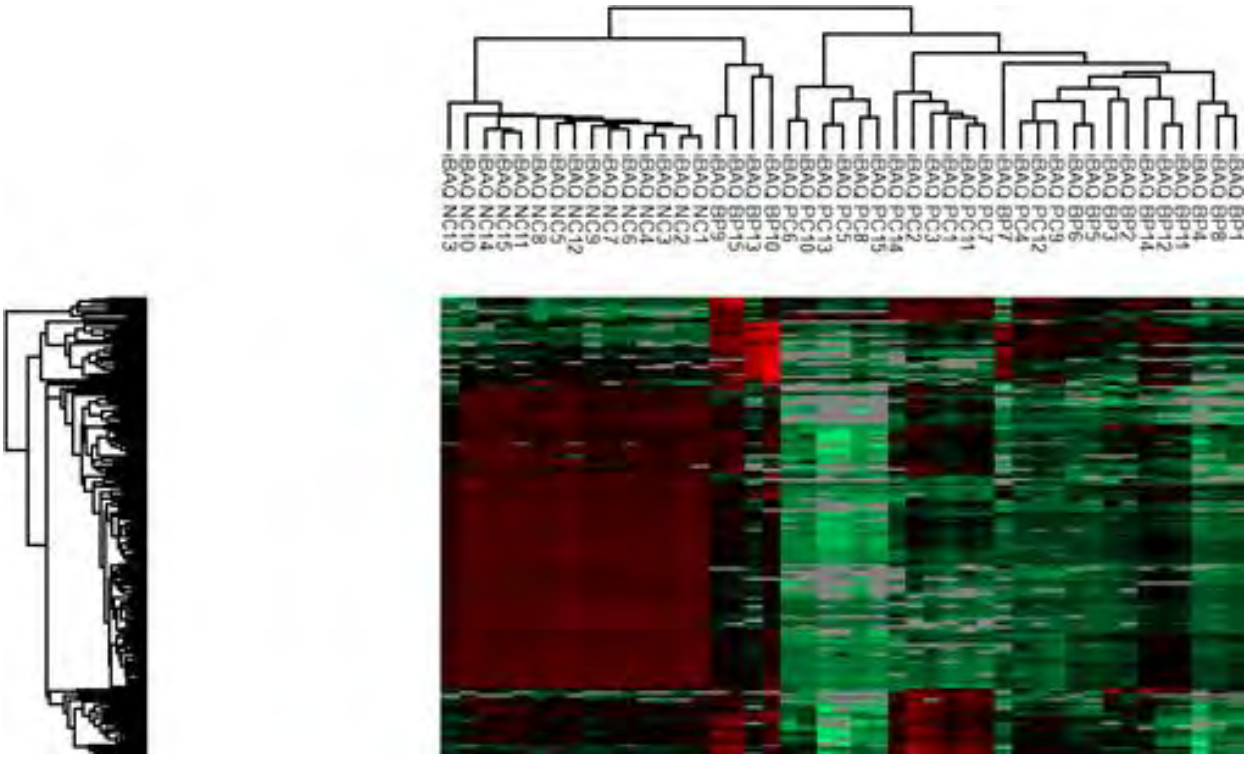


Figure 3. 18. Heatmap shows distinct molecular signature between PCa, NC and BPH, with slight overlap between a few samples

Multivariate testing using principal component analysis for PCa, NC and BPH samples was also performed, and this revealed a similar trend with the hierarchical clustering when a 2-D (*principal components1&2*) or a 3-D (*Principal components 1,2&3*) was performed (**Figure 3.19**). It was observed that NC samples clustered more tightly compared with PCa or BPH. By these proteomics analyses, there seems to be at least 2 sub-groups of BPH and PCa and possibly multiple overlaps in peptide signature between BPH and PCa (**Figure 3.20**).

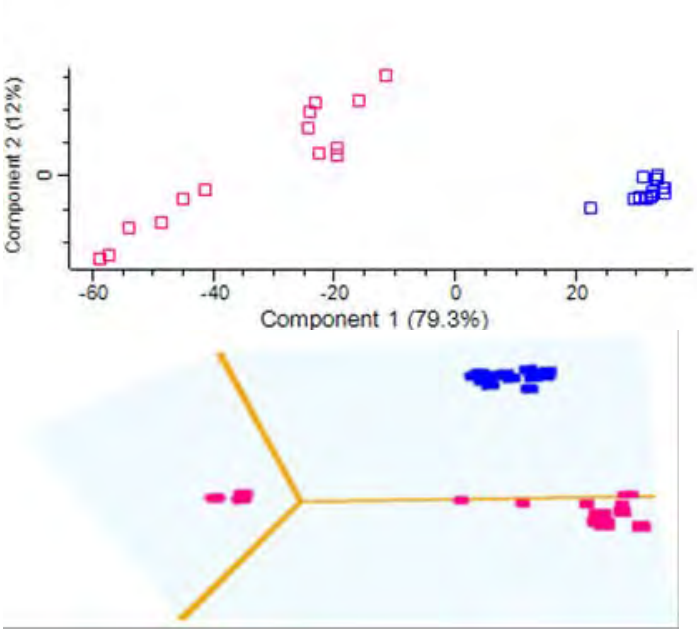


Figure 3.19. 2-D and 3-D principal component analyses shows separate clustering of peptides in NC (blue) and PCa (pink), even though there seems to be 2 distinct subgroups of PCa

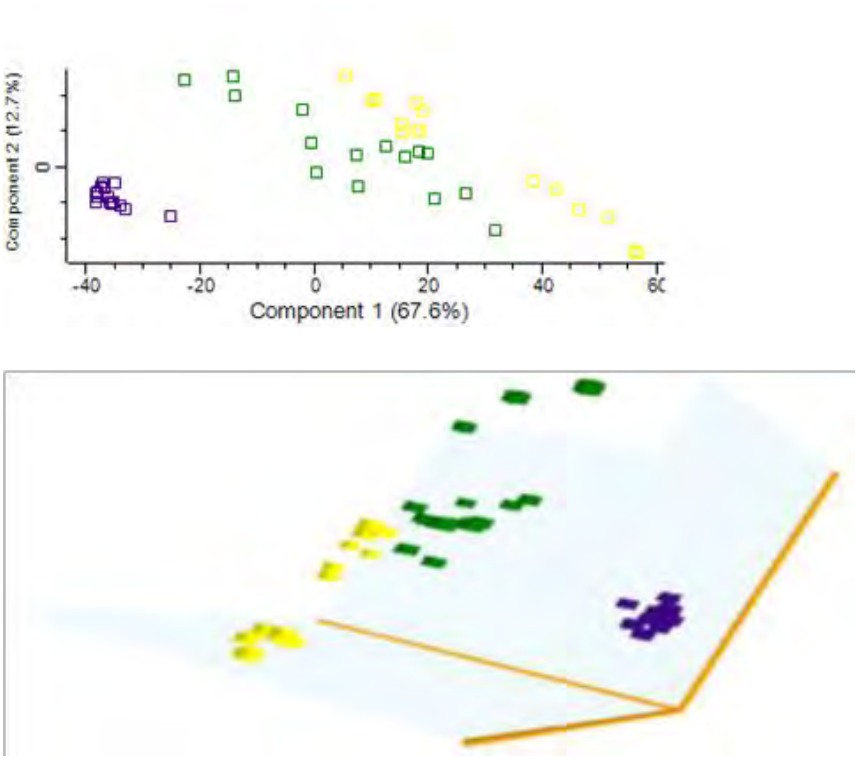


Figure 3.20. 2-D and 3-D principal component analyses show separate clustering of NC (purple), PCa (yellow) and BPH (green); however some overlap can be observed between PCa and BPH.

3.3.9. Racial Trends in Potential Prostate Cancer Biomarkers

The 15 PCa samples in our cohort was carefully analysed for racial trends in potential biomarkers identified. In this PCa cohort, nine samples were of Mixed Ancestry or "Coloured" origin (PCa_M), three were blacks Africans (PCa_A) and the last three were Caucasians (PCa_A). Even though Q-Q plots (**Figure 3.21**), histograms (**Figure 3.22**), Boxplots (**Figure 3.23**) and Shapiro-Wilk test (p value > 0.05) indicated that these samples were normally distributed, we could not identify any racial biomarker between these three racial groups with statistical significance both by ANOVA and non-parametric Kruskal-Wallis test; this is possibly because our sample size is underpowered to decipher the true between-group differences.

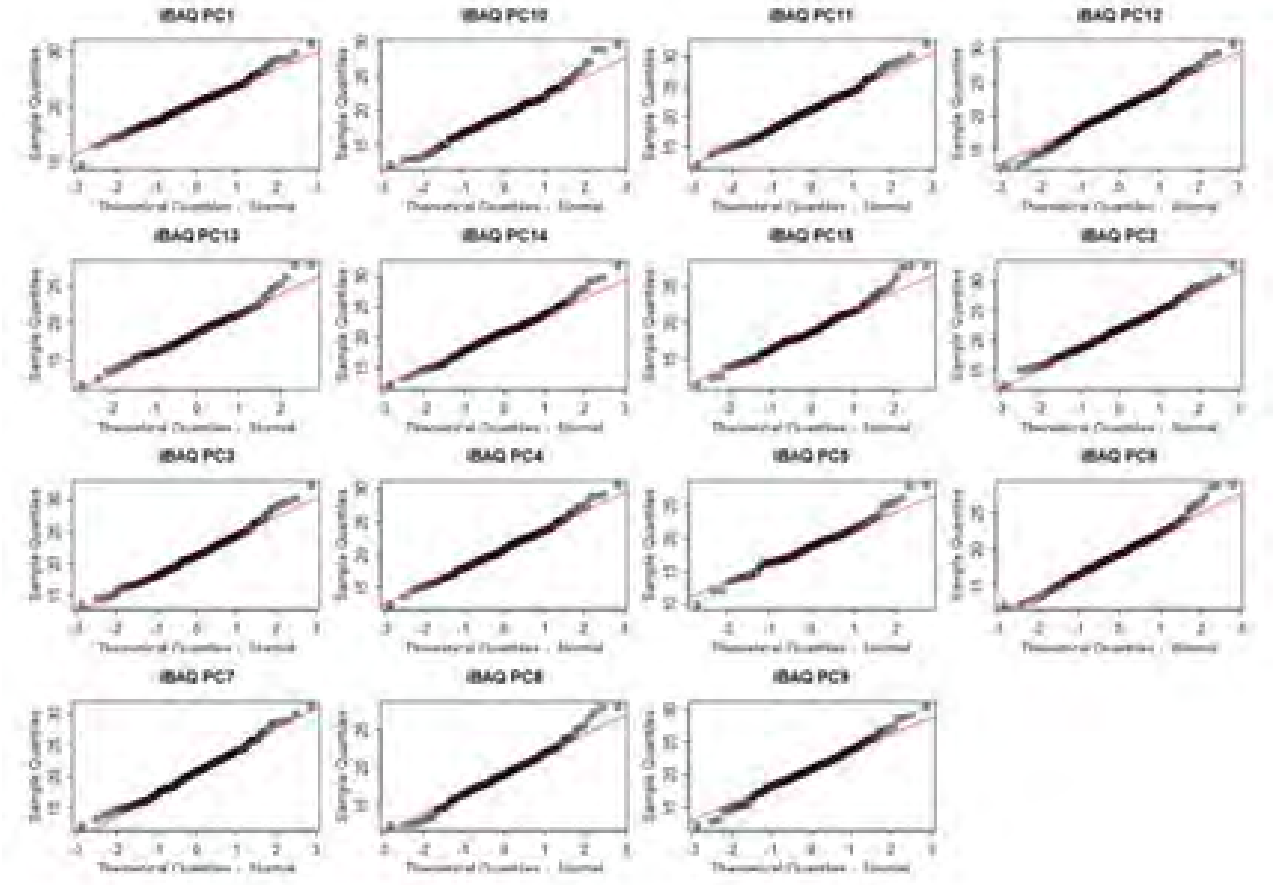


Figure 3.21. Q-Q plots of all 15 PCa samples show normal distribution of all samples in relation to the reference line

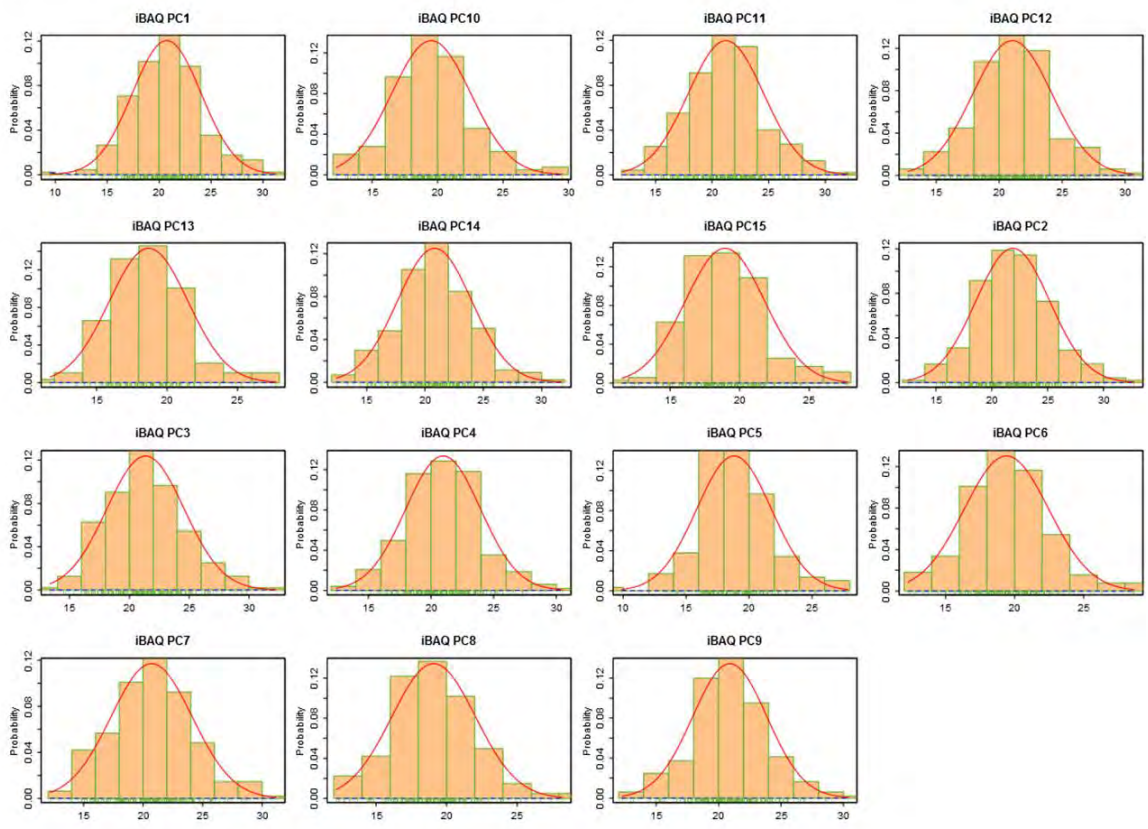


Figure 3.22. Histograms of all 15 PCa samples demonstrating a Gaussian distribution of all samples. A range of log2-transfromed IBAQ intensities were plotted on the x-axis, while the probability is plotted on the y-axis.

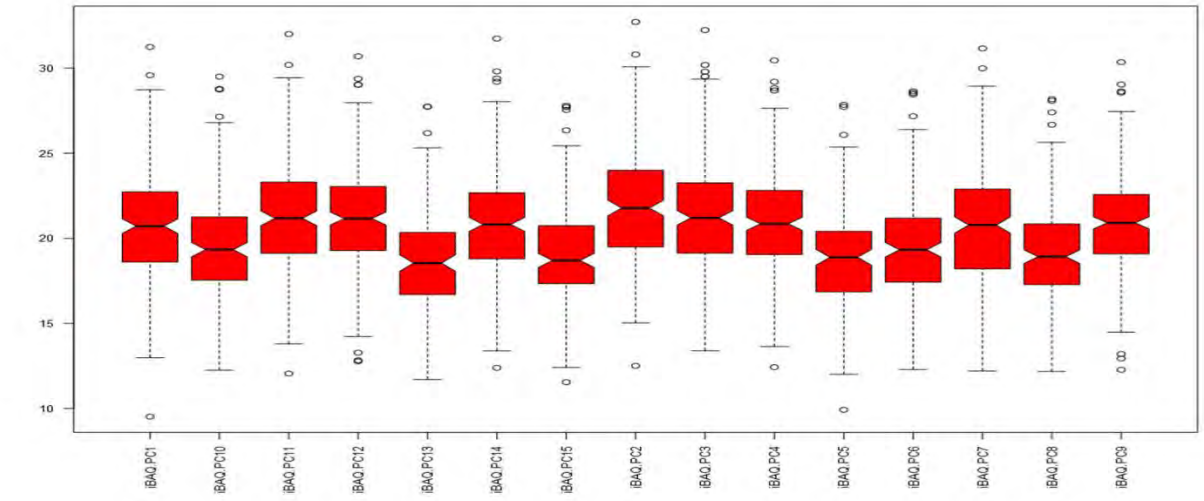


Figure 3.23. Boxplot showing normal distribution of all 15 PCa samples in the cohort

The “numeric-venn” function in Perseus Software was further used to identify protein groups that are unique to preprocessed and filtered protein groups. Even though this method is a binary approach without considering relative expression, we identified that MAN1A1 (Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA) and SLAIN1 (SLAIN motif-containing protein 1) were found to be unique to PCa_A; and GP1BA (Platelet glycoprotein Ib Alpha chain; Glycocalicin), PSMF1 (proteasome inhibitor P131 subunit), MYOC (Myocilin) , FABP6 (Gastrotropin), and Ig Heavy chain V-III region ZAP were found to be unique to the PCa_M. There were no protein groups found unique to the PCa_C group by this approach (**Table 3.7**).

PCa_A	PCa_M	PCa_C
MAN1A1	GP1BA	
SLAIN1	PSMF1	
	MYOC	
	FABP6	
	Ig Heavy chain V-III region ZAP	

Table 3.7. List of proteins unique to racial groups by numeric Venn analysis

Further analysis was carried out on the 45 and 28 potential biomarkers identified by iBAQ/LFQ analysis and MS1 spectral analysis respectively. For each of these 2 sets of potential prostate cancer biomarkers, 4-way Venn analysis with protein groups in PCa_A, PCa_M and PCa_C was performed to identify potential ethnic-based biomarker trend. Thirty-seven (37) out of 45 differentially expressed potential biomarkers were shared by the 3 racial groups. HPT (Haptoglobin) was shared by PCa_C, and PCa_M. AMY1 (Alpha-amylase 1), PTGDS (Prostaglandin-H2 D-isomerase), REG1A (Lithostathine-1-alpha), OSTP (Osteopontin), and SAP3 (Ganglioside GM2 activator) were found unique to PCa_C. ODO2 (Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex) was only found in PCa_A (**Figure 3.24**).

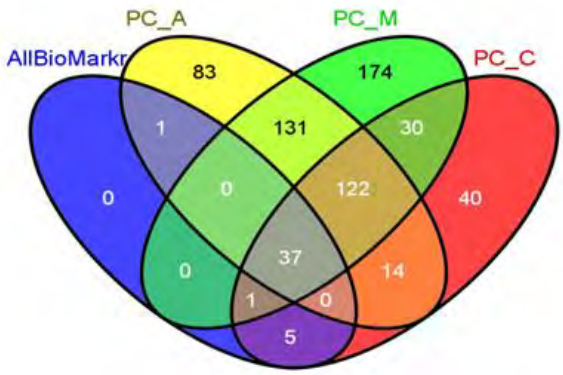


Figure 3.24. Venn diagram showing overlap between differentially expressed potential biomarkers and the 3 racial groups

Nine (9) out of the 28 potential biomarkers by MS1 spectral analysis were shared by the 3 racial groups. PON1 (Serum paraoxonase/arylesterase 1) and C4BPA (C4b-binding protein alpha chain) were found to be unique to PCa_A. KLKB1 (Plasma kallikrein heavy chain), ITIH3 (Inter-alpha-trypsin inhibitor heavy chain H3), ITIH2 (Inter-alpha-trypsin inhibitor heavy chain H2), and COL6A1 (Collagen alpha-1(VI) chain) were observed to be common to PCa_C and PCa_M. PLSL (Plastin-2), ACTN1 (Alpha-actinin-1), COL6A3 (Collagen alpha-3(VI) chain), and TENX (Tenascin-X) were common to PCa_A and PCa_M. HPTR (Isoform 2 of Haptoglobin-related protein), PZP (Pregnancy zone protein), PROS (Vitamin K-dependent protein S), IGHD (Isoform 2 of Ig delta chain C region), and CBPN (Carboxypeptidase N catalytic chain) were unique to PCa_M (**Figure 3.25**).

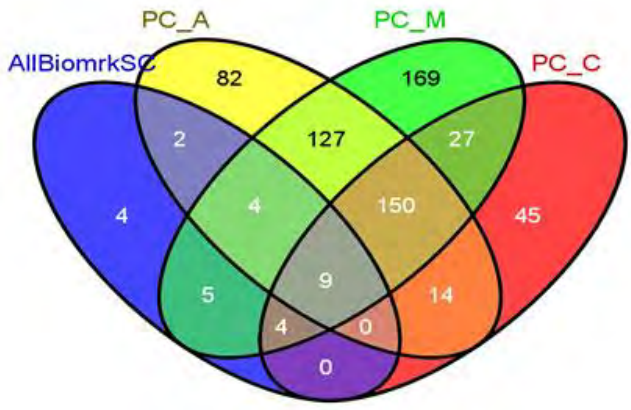


Figure 3.25. Venn diagram showing overlap between 28 potential biomarker by MS1 spectral analysis and three racial groups

3.4. Discussion

It is well reported that the urinary proteome is far less complex as compared with plasma, serum, cell or tissue proteomes [327, 328]. As performed initially here, pooling reduces the level of biologic variability among individuals and minimizes the experimental runs needed, as well as reduces the total cost of experiments [329]. Furthermore, it also permits for method development and optimization; and avoidance of sample wastage [330]. However, it is unable to provide adequate basis for good biological inferences concerning the types of samples being compared. This strategy can generate useful comparative information about pooled samples themselves, albeit unable to provide adequate information about the individuals in the population from which the pools were drawn [329]. Biologically independent replicate sub-pooling can improve the information derived from pooling experiments [329]. Considering that proteome composition within and between individuals varies with disease state, diet, time of the day, and possible environment or ethnicity [331-334]; it is necessary to statistically normalize proteomics data to increase the signal to noise ratio. Various normalization methods have been reported [320, 335-337], but great caution must be exercised to avoid over-fitting of data [335].

As observed in this study, up to 63% of the peptides were shared by at least 3 out of the 4 pooled groups. Previous studies have reported the measurable "core proteome" in proteomics experiment to be similar [323, 338] . The advent of high resolution, accurate mass (HRAM) mass spectrometers with mass accuracy less than 10ppm has drastically improved quantification of integrated MS1 peak intensities, even in a label-free experimental setup [339-341]. Heretofore, MS2 spectral counting methods were more appealing to proteomics researchers due to the inherent inaccuracies of MS1 signals quantification on lower resolution instruments [342, 343]. It is now known that MS2 spectral counting is highly data dependent and true abundance level of peptides can be easily saturated [341]. MS2 signal saturation is a consequence of ion packeting; particularly when ions are present at high concentrations [344]. This affects the accurate quantitation of proteins due to the fact that a huge number of ions are reaching the detector at the same time, particularly ions from high abundance proteins [344]. (A figure illustrating the essential differences between MS1 and MS2 quantitation is presented in **Annexure VI**) The

statistical power of 0.8 as we have used in this study is the minimum acceptable level for proteomics experiments in keeping with literature [345],.

Disease related and normal human urinary proteomes have be characterized by various studies [116, 321, 323, 346-353]; detecting between 124 to 1823 proteins. Two landmark researches recently described a comprehensive draft of the human proteome using several fetal and adult human tissues, cell lines, and body fluids. Between 17, 294 - 18,097 of the over 20,000 annotated genes in the human genome were identified in these studies [323, 352]. These aggregations of data from several studies have expedited the development of the proteome databases. However there remains a dearth of proteome catalog constituents from African-based studies, particularly a urine-based proteome. Hence, we established a draft physiologic urinary proteome of a Southern African human cohort, in addition to PCa specific urinary proteome.

Previous studies have employed smaller cohort of patients in comparison this study [317, 354]. Validation of proteins generated from mass spectrometry based experiments with real-time qualitative PCR (RTqPCR) as performed by some studies [354] could potentially interfere with downstream statistical inferences because transcriptomics does not always correlate accurately with proteomics and *vice versa* due to factors like posttranslational modifications and translational control [355].

From the individual sample analysis, distinct clustering of PCa, BPH and NC with minor overlaps in a few BPH and PCa samples may indicate underlying disparity between molecular and histopathologic classification. It was particularly interesting to note that a subset of BPH patients were similar to NC patients, and some PCa patients looked more like BPH patients, based on their urinary proteomes. This further buttresses the fact that molecular signatures may be more reliable for classification than phenotypic characterization. In addition, this disparity may point to the fact that some prostate cancers may originate from a prior BPH state and may have a different prognostic indication. However, these observed differences in PCa urinary sub-proteomes require further research.

Many of the candidate biomarkers discovered in this study have been previously described in the literature as playing a role in cancer, including prostate cancer **(Table 3.8).** For instance, Pregnancy Zone Protein (PZP), which is a homologous inhibitor of alpha-2 macroglobulin, is able to cleave prostate specific antigen at Lys145-lys146 to form “nicked PSA”. This cleavage

does not inactivate PSA but limits recognition of PSA epitopes by commercial ELISA kits. Also, Trefoil Factor 3 (TFF3) overexpression has been reported to increase ERG-mediated cellular invasion and tumour aggressiveness in castration-resistant PCa [356].

Protein Names	References in Other Diseases	References in Cancer	References in Prostate Cancer
Vitamin K-dependent protein S	El Asmar MS et al, Oman Med J. May 2014; 29(3): 172–177	Viegas CS et al, Biomed Res Int. 2014; 2014: 340216	Samykutty A et al, Evid Based Complement Alternat Med. 2013; 2013: 287358
Haptoglobin-related protein	Ko DH et al, Biomed Res Int. 2013; 2013: 390630	Kuhajda FP et al, Proc Natl Acad Sci U S A. Feb 1989; 86(4): 1188–1192	Mavondo GA et al, Clin Chim Acta. 2012 Jan 18;413(1-2):334-6
Carbonic anhydrase 2	Sly WS et al, Proc Natl Acad Sci U S A. May 1983; 80(9): 2752–2756	Nógrádi A, Am J Pathol. Jul 1998; 153(1): 1–4	Pértega-Gomes N et al, BMC Cancer. 2014; 14: 352
Nidogen-1	Xu X et al, J Virol. Nov 2010; 84(22): 11866–11875	Ulazzi L et al, Mol Cancer. 2007; 6: 17	Dedhar S et al, J Biol Chem. 1992 Sep 15;267(26):18908-14
Alpha-2-macroglobulin	Housley J, J Clin Pathol. Jan 1968; 21(1): 27–31	Lin Z et al, Electrophoresis. 2014 Aug;35(15):2108-15	Misra UK et al, PLoS One. 2012; 7(12): e51735
Complement factor H	de Córdoba SR et al, Clin Exp Immunol. 2008 Jan;151(1):1-13	Junnikkala S et al, Br J Cancer. 2002 Nov 4;87(10):1119-27	Not yet described
C4b-binding protein alpha chain	Bergamaschini L et al, Clin Exp Immunol. 1999 May;116(2):220-4	Liu YS et al, J Mol Cell Biol. 2012 Oct;4(5):344-7	Not yet described
Inter-alpha-trypsin inhibitor heavy chain H2	García-Ramírez M et al, Diabetologia. 2007 Jun;50(6):1294-303	Hamm A et al, BMC Cancer. 2008 Jan 28;8:25	Hamm A et al, BMC Cancer. 2008 Jan 28;8:25
Serum paraoxonase/arylesterase 1	Bindu CM et al, Indian J Clin Biochem. 2011 Jul;26(3):230-4	Elkiran ET et al, BMC Cancer. 2007 Mar 15;7:48	Not yet described
Plasma kallikrein;Plasma kallikrein heavy chain;Plasma kallikrein light chain	Colman RW et al, J Clin Invest. Feb 1978; 61(2): 287–296	Chee J et al, Biol Chem. 2008 Sep;389(9):1225-33	Ferreira JG et al, J Biol Chem. 2013 May 10;288(19):13641-54
Complement component C8 alpha chain	Tedesco F et al, J Clin Invest. Sep 1990; 86(3): 884–888	Nasim FU et al, Biomark Cancer. 2012 Dec 10;4:19-34	Not yet described

Lactotransferrin;Kaliocin-1;Lactoferroxin-A;Lactoferroxin-B;Lactoferroxin-C	Lamb CA et al, Br J Surg. 2009 Jun;96(6):663-74	Deng M et al, Oncogene 32, 4273-4283, Sept 2013	Shaheduzzaman S et al, Cancer Biol Ther. 2007 Jul;6(7):1088-95
Apolipoprotein B-100;Apolipoprotein B-48	Harper CR et al, Mayo Clin Proc. 2010 May;85(5):440-5	Gong Y et al, PLoS One. 2013 Apr 18;8(4):e61456	Not yet described
Inter-alpha-trypsin inhibitor heavy chain H1	Eiberg H et al, Hum Hered. 1993 Jul-Aug;43(4):250-6	Paris S et al, Int J Cancer. 2002 Feb 10;97(5):615-20	Not yet described
Alpha-actinin-1	Guéguen P et al, PLoS One. 2013 Sep 17;8(9):e74728	Downey C et al, Am J Surg. 2011 Nov;202(5):520-3	Not yet described
Plastin-2	Rosa N et al, Biomed Res Int. 2014;2014:569632	Shinomiya H et al, Int J Cell Biol. 2012;2012:213492	Zheng J et al, Am J Pathol. 1999 Jul;155(1):115-22
Collagen alpha-1(XII) chain	Prockop DJ, J Clin Invest. Mar 1985; 75(3): 783–787	Ikeda K et al, Am J Pathol. 2006 Mar;168(3):856-65	Not yet described
Collagen alpha-1(VI) chain	Prockop DJ, J Clin Invest. Mar 1985; 75(3): 783–787	Ikeda K et al, Am J Pathol. 2006 Mar;168(3):856-65	Not yet described
Tenascin	Riley GP et al, Am J Pathol. Sep 1996; 149(3): 933–943	Juuti A et al, J Clin Pathol. 2004 Nov;57(11):1151-5	Gravina GL et al, Int J Oncol. 2013 Jun;42(6):2116-22
Collagen alpha-3(VI) chain	Prockop DJ, J Clin Invest. Mar 1985; 75(3): 783–787	Ikeda K et al, Am J Pathol. 2006 Mar;168(3):856-65	Not yet described
Cadherin-11	Schneider DJ et al, FASEB J. 2012 Feb;26(2):503-12	Carmona FJ et al, J Pathol. 2012 Oct;228(2):230-40	Chu K et al, Mol Cancer Res. 2008 Aug;6(8):1259-67
Roundabout homolog 4	Nafar M et al, Int J Nephrol. 2014;2014:574261	Dickinson RE et al, PLoS One. 2011;6(11):e27792	Choi YJ et al, Pathol Oncol Res. 2014 Jul;20(3):517-9
Cathepsin Z	Cheng XW et al, Circulation. 2012 Mar 27;125(12):1551-62	Tan GJ et al, World J Biol Chem. 2013 Nov 26;4(4):91-101	Podgorski I et al, Biochem Soc Trans. 2007 Aug;35(Pt 4):701-3

Alpha-N-acetylglucosaminidase;Alpha-N-acetylglucosaminidase 82 kDa form;Alpha-N-acetylglucosaminidase 77 kDa form	Schmidtchen A et al, Am J Hum Genet. 1998 Jan;62(1):64-9	Zhao Z et al, Mamm Genome. 1996 Sep;7(9):686-90	Kraft P et al, PLoS Genet. 2005 Nov;1(5):e68
Carboxypeptidase N catalytic chain	Rangan SK et al, Biochemistry. 2003 Dec 9;42(48):14328-34	Bhatia J et al, Int J Cancer. 2000 Feb 15;85(4):571-7	Mesters JR et al, EMBO J. 2006 Mar 22;25(6):1375-84
Ig delta chain C region	Bertoldi C et al, J Clin Periodontol. 2013 Jun;40(6):573-82	Ueda C et al, Cancer Genet Cytogenet. 2001 Oct 1;130(1):42-50	Yasuda N et al, Am J Hematol. 1994 Sep;47(1):65-6
Inter-alpha-trypsin inhibitor heavy chain H3	Ebana Y et al, J Hum Genet. 2007;52(3):220-9	Chong PK et al, J Proteome Res. 2010 Jul 2;9(7):3671-9	Not yet described
Pregnancy zone protein	Ijsselstijn L et al, J Proteome Res. 2011 Nov 4;10(11):4902-10	Damber MG et al, Arch Gynakol. 1976 Sep 17;221(2):97-101	Not yet described
Carbonic anhydrase 1	Frost SC et al, Subcellular Biochemistry Vol 75 2014	Pastorekova S et al, Curr Pharm Des 2008 2008;14(7):685-98.	Takakura M et al, ISRN Oncol. 2012;2012:768190
Hemoglobin subunit alpha	Yu C et al, J Hepatol. 2012 Jan;56(1):241-7	Kim JH et al, Int J Oncol 2013 10.3892/ijo.2013.1803	Not yet described
Apolipoprotein C-III	Hu ZJ et al, Exp Ther Med. 2014 Sep;8(3):951-956	Honda K et al, PLoS One. 2012;7(10):e46908	Zhang XM, Zhonghua Nan Ke Xue. 2010 Aug;16(8):721-5
Haptoglobin	Williams R et al, Gut. Dec 1961; 2(4): 297–303	Zhao C et al, Neoplasia. Jan 2007; 9(1): 1–7	Fujimura T et al, Int J Cancer. 2008 Jan 1;122(1):39-49
Protein S100-A9	Vogl T et al, Int J Mol Sci. 2012;13(3):2893-917	Gebhardt C, Biochem Pharmacol. 2006 Nov 30;72(11)	Hermani A et al, Clin Cancer Res. 2005 Jul 15;11(14):5146-52
Flavin reductase (NADPH)	Hustad S et al, Am J Clin Nutr. 2004 Oct;80(4):1050-7	Wojcieszyńska D, Int J Mol Sci. 2012 Dec 7;13(12)	McDonald CA, Biochemistry. 2013 Sep 3;52(35)
Hemoglobin subunit beta	Yu C et al, J Hepatol. 2012	Kim JH et al, Int J Oncol 2013	Not yet described

	Jan;56(1):241-7	10.3892/ijo.2013.1803	
Leukocyte-associated immunoglobulin-like receptor 1	Zhang Y et al, Clinics vol.68 no.4 São Paulo Apr. 2013	Xue JN et al, Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi. 2008 Apr;24(4):373-4, 377	Not yet described
Proactivator polypeptide;Saposin-A;Saposin-B-Val;Saposin-B;Saposin-C;Saposin-D	Morimoto S et al, Proc Natl Acad Sci U S A. May 1990; 87(9): 3493–3497	Chu Z et al, PLoS One. 2013; 8(10): e75507	Lee TJ et al, Mol Cancer. 2004; 3: 31
Polyubiquitin-C;Ubiquitin;Polyubiquitin-B;Ubiquitin;	Contu VR et al, J Neurochem. 2014 May 15. doi: 10.1111/jnc.12762. [Epub ahead of print]	Hurst-Kennedy J et al,Biochem Res Int. 2012; 2012: 123706	Shi D et al, Cancer Biol Ther. Oct 15, 2010; 10(8): 737–747
Pro-epidermal growth factor;Epidermal growth factor	Makki N et al, Int J Mol Sci. Oct 2013; 14(10): 20597–20613	Wykosky J et al, Chin J Cancer. Jan 2011; 30(1): 5–12	Maddy SQ et al, Br J Cancer. Jul 1989; 60(1): 41–44
Trefoil factor 3	Aamann L et al, World J Gastroenterol. Mar 28, 2014; 20(12): 3223–3230	Huang Z et al, BMC Gastroenterol. 2014; 14: 74	Rickman DS, Neoplasia. Dec 2010; 12(12): 1031–1040
CD59 glycoprotein	Asimakopoulos JV et al, Med Sci Monit. 2014; 20: 123–139	Koretz K et al, Br J Cancer. Nov 1993; 68(5): 926–931	Liu AY et al, American Journal of Pathology, Vol. 160, No. 1, January 2002
Prostaglandin-H2 D-isomerase	Martins-De-Souza D, World J Biol Psychiatry. 2010 Aug;11(5):719-28	Tippin BL et al, Prostaglandins Other Lipid Mediat. 2012 Jan;97(1-2):22-8	Haj-Ahmad TA et al, J Cancer. 2014; 5(2): 103–114
Neutrophil gelatinase-associated lipocalin	Devaraian, Scand J Clin Lab Invest Suppl. 2008;241:89-94	Krysan et al, Am J Transl Res 2013;5(5):481-496	Mahadevan NR et al, BMC Cancer 2011, 11:229
Mannan-binding lectin serine protease 2 A chain	Pągowska-Klimek I et al, Biomed Res Int. 2014; 2014: 616817	Ytting H et al, Scand J Immunol. 2011 Feb;73(2):122-7	Not yet described
Trefoil factor 1	Aamann L et al, World J Gastroenterol. Mar 28, 2014; 20(12):	Markićević M et al, Int J Med Sci. 2014; 11(7): 663–673	Bougen NM, Cancer Letters 2013, 332(1):19-29

	3223–3230		
Glyceraldehyde-3-phosphate dehydrogenase	Berry MD, J Psychiatry Neurosci. Sep 2004; 29(5): 337–345	Wang D et al, PLoS One. 2013 Apr 19;8(4):e61262	Harada N et al, J Biol Chem. 2007 Aug 3;282(31):22651-61
Ig kappa chain V-I region BAN	Ozaki S et al, Clin Immunol Immunopathol. 1994 May;71(2):183-9	Solomon A et al, Am J Pathol. Oct 1990; 137(4): 855–862	Not yet described
Pancreatic alpha-amylase	Kojecký Z et al, Mater Med Pol. 1981 Jan-Mar;13(1):18-22	Higashiyama et al, J Clin Pathol. Feb 1991; 44(2): 144–146	Cheng HL et al, Proteomics Clin Appl. 2011 Apr;5(3-4):121-32
Lithostathine-1-alpha	Grégoire C et al, EMBO J. Jul 2, 2001; 20(13): 3313–3321	Sato Y et al, Ann Surg Oncol. 2013 Sep;20(9):3044-51	Hayashi T et al, Oncol Rep. 2009 Jan;21(1):95-100
Prostate-specific antigen	Romppanen J et al, Br J Cancer. Mar 1999; 79(9-10): 1583–1587	Carder PJ et al, J Clin Pathol. Jan 2005; 58(1): 69–71	Scosyrev E et al, Cancer. 2012 Dec 1;118(23):5768-76
Ribonuclease pancreatic	Yang D et al, J Immunol. Nov 15, 2004; 173(10): 6134–6142	Huang W et al, PLoS One. 2014; 9(5): e96490	Chakrabarti A et al, J Interferon Cytokine Res. Jan 2011; 31(1): 49–57
Collagen alpha-2(I) chain	Holden P et al, Am J Hum Genet. Jul 1999; 65(1): 31–38	Fenhalls et al, Br J Cancer. Dec 1999; 81(7): 1142–1149	Hall CL et al, Cancer Res. 2006 Sep 1;66(17):8648-54
Non-secretory ribonuclease	Sakakibara R et al, J Biochem. 1992 Mar;111(3):325-30	Fernández-Salas E et al, Eur J Biochem. 2000 Mar;267(5):1484-94	Not yet described
Osteopontin	Kahles F et al, Mol Metab. Jul 2014; 3(4): 384–393	Rud AK et al, BMC Cancer. 2013; 13: 540	Thoms JW et al, Br J Cancer. 2012 Aug 21;107(5):840-6
Uteroglobin	Antico G et al, Mediators Inflamm. 2014; 2014: 876395	Bignotti E et al, J Transl Med. 2013; 11: 162	Patierno SR et al, Clin Prostate Cancer. 2002 Sep;1(2):118-24
Prostatic acid phosphatase;PAPf39	Bull H et al, Mol Pathol. Apr 2002;	Graddis TJ et al, Int J Clin Exp	Quintero IB et al, PLoS One. 2013

	55(2): 65–72	Pathol. Mar 31, 2011; 4(3): 295–306	Sep 10;8(9):e73072
Histone H1.5	Li JY et al, PLoS Genet. Aug 2012; 8(8): e1002879	Harshman SW et al, J Proteome Res. 2014 May 2;13(5):2453-67	Dryhurst D et al, Cancer Metastasis Rev. 2014; 33: 429–439
Ganglioside GM2 activator;Ganglioside GM2 activator isoform short	Kodama T et al, PLoS One. 2011;6(12):e29074	Hettmer S et al, Br J Cancer. Jul 19, 2004; 91(2): 389–397	Van Slambrouck et al, Int J Oncol. Oct 2009; 35(4): 693–699
Fibrillin-1	Kim KL et al, Biomol Ther (Seoul). Mar 2014; 22(2): 143–148	Harsha HC et al, PLoS Med. 2009 Apr 7;6(4):e1000046	Wallace TA et al, Cancer Res. 2008 Feb 1;68(3):927-36
Glutaredoxin-1	Kuipers I et al, Eur Respir J. Feb 2013; 41(2): 10.1183/09031936.00115212.	Mollbrink A et al, Int J Immunopathol Pharmacol. 2014 Apr-Jun;27(2):169-83	Not yet described
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	Dumont M et al, Free Radic Biol Med. 2009 Oct 1;47(7):1019-27	Rezaul K et al, Genes Cancer. Mar 2010; 1(3): 251–271	Not yet described
Beta-defensin 1	Andresen E et al, PLoS One. 2011;6(7):e21898	Han Q et al, PLoS One. 2014; 9(3): e91867	Kim HJ, BJU Int. 2011 Jan;107(1):144-9
Lysozyme C	Gotes J et al, J Appl Physiol (1985). 2012 Feb;112(4):638-50	Serra C et al, Breast Cancer Res. 2002; 4(6): R16	Sfanos KS et al, Proc Natl Acad Sci U S A. Mar 3, 2009; 106(9): 3443–3448
Epididymal secretory protein E1	Liou HL et al, J Biol Chem. 2006 Dec 1;281(48):36710-23	Yang Z et al, Onco Targets Ther. 2013; 6: 957–966	Not yet described
Monocyte differentiation antigen CD14; urinary form,membrane-bound form	Yang J, Biomark Res. 2014; 2: 1	Li K et al, World J Gastroenterol. Mar 14, 2014; 20(10): 2688–2694	Idorn M et al, Cancer Immunol Immunother. 2014 Aug 2. [Epub ahead of print] PMID: 25085000
Basement membrane-specific heparan sulfate proteoglycan core protein;Endorepellin;LG3	Arikawa-Hirasawa E et al, Am J Hum Genet. May 2002; 70(5): 1368–1375	Ahsan MS et al, J Oral Pathol Med. 2011 Aug;40(7):552-9	Datta MW et al, Mol Cancer. 2006; 5: 9

peptide			
Trefoil factor 2	Aamann L et al, World J Gastroenterol. Mar 28, 2014; 20(12): 3223–3230	Jiang P et al, Oncol Lett. May 2014; 7(5): 1525–1531	Vestergaard EM et al, Int J Cancer. 2010 Oct 15;127(8):1857-65
WAP four-disulfide core domain protein 2	Liu W et al, Int J Tuberc Lung Dis. 2013 Oct;17(10):1346-53	Hellström I et al, Cancer Res. 2003 Jul 1;63(13):3695-700	Watson JE et al, Prostate. 2004 Oct 1;61(2):192-9
Cystatin-M	Shlipak MG et al, N Engl J Med. 2013 Sep 5;369(10):932-43	Sotiropoulou G et al, J Biol Chem. 1997 Jan 10;272(2):903-10	Pulukuri SM et al, Oncogene. Aug 6, 2009; 28(31): 2829–2838
N-acetylmuramoyl-L-alanine amidase	Yang J et al, J Bacteriol. Jun 2013; 195(12): 2887–2897	Hoijer MA et al, Blood. 1997 Aug 1;90(3):1246-54	Not yet described
Ly-6/neurotoxin-like protein 1	Tsuji H, Genomics. 2003 Jan;81(1):26-33	Choi SH et al, Int J Oncol. 2009 Sep;35(3):601-7	Hruska M et al, J Neurosci. 2009 Nov 25;29(47):14847-54
SH3 domain-binding glutamic acid-rich-like protein 3	Lichtenfels R et al, PLoS One. 2012;7(7):e41345	Makkinje A et al, J Biol Chem. Aug 10, 2012; 287(33): 27703–27714	Zhang H et al, Cancer Genomics Proteomics. 2005 Apr;2(2):97-114
Apolipoprotein A-II;Truncated apolipoprotein A-II	Corsetti JP et al, PLoS One. 2012;7(6):e39110	Ahn JY et al, PLoS One. 2014; 9(4): e93794	Dwivedi S et al, Indian J Clin Biochem. Apr 2013; 28(2): 107–109

Table 3.8. Literature evidence of discovered potential biomarkers in disease, cancer and prostate cancer

Subcellular characterization of these potential biomarkers shows that most were predicted as secreted, cell membrane, and extracellular proteins. Many inflammatory molecules were found to be upregulated in PCa as well. This was not strange as inflammation is an emerging hallmarks of cancer [25]. Also, cluster of differentiation (CD) family molecules and other proteins derived from leukocytes such as neutrophil gelatinase-associated lipocalin 2 (LCN2) have been reported in human urinary proteomes previously [321, 351, 353, 357]. In an NF-kB dependent manner, the renal tubular enzyme LCN2 has been previously reported to be upregulated in PCa cells [358]. Individualized medicine is quite challenging in the Low and middle income countries (LMICs) who carry up to 80% of the global cancer burden, but receives only about 5% of the world's spending on cancer [132]. Even though several prostate cancer genetic biomarkers are known [359, 360], the journey to finding newer proteomic biomarkers has been a difficult one [361-364]. One big challenge is how to improve the reliability of PSA in order to avoid false positives and overtreatment [365]. Novel potential diagnostic biomarkers discovered in this study may complement the efficacy of PSA or its parameters. Several factors including epigenetic modification [366] of the androgen receptors, epithelial-mesenchymal transition [367] and cancer stem cells [368] may play fundamental roles in aggressive progression of PCa. High throughput technologies have resulted in the emergence of novel PCa biomarkers like Circulating Tumour Cell (CTC) telomerase activity [369].

This study requires large scale, multiplatform, and multicenter clinical validation. However, it demonstrated the efficacy of urinary proteomics in the identification of novel potential biomarkers of prostate cancer. It also established a draft physiologic and prostate cancer specific urinary proteome for a South African population. Although racial trends within the PCa cohort was not established by differential expression, ethnic-related biomarker variations identified by the methods described in this study warrants further study.

CHAPTER 4: TARGETED PROTEOMICS PREVALIDATION AND IN-SILICO VERIFICATION OF CANDIDATE PROSTATE CANCER BIOMARKERS

4.1. Background

PCa is responsible for the highest number of cancer deaths in African men [370]. Despite this, it receives inadequate public health attention due to the concurrent high burden of Malaria, Tuberculosis, and HIV/AIDS. PSA has resulted in overdiagnosis and overtreatment due to its poor negative predictive value and specificity [371]. PCa disparity in Africa needs to be addressed. Even though some promising molecular biomarkers of PCa such as the TMPRSS2-ERG fusion gene, E-Cadherin, PCA3, PTEN and EZH2 have been discovered [372], these have not been well research or validated in African populations.

The process of validating multiple potential biomarkers generated from high throughput discovery shotgun proteomics can be very tasking. This validation gap prevents a glut of promising discovery phase potential biomarkers from making it to clinical utility [373-376]. Given the *status quo*, the use of preexisting experimental databases as an adjunct to actual high throughput systems biology experiments, offers a promising option to improve the prevalidation of potential biomarkers of interest and expedite the biomarker pipeline processes. Cancer proteomics research databases can be grouped into five classes [377], *viz* tumour antigen, gene/protein expression, protein interaction/ pathway, gene mutation and SNP, and cancer-associated genes databases. Several proteomics databases have emerged for translation of systems biology data into useable diagnostic and therapeutic tools; such as the Max-Planck Unified database (MAPU) [378], CancerResource [379], PepSeeker [380], Global Proteome Machine Database (GPMDB) [381], Yale Protein Expression Database (YPED) [382], Genome Annotating Proteomic Pipeline (GAPP) [383], NeXtProt [384], Tranche [385], and Proteomic Identifications database (PRIDE) [381]. In addition, various reports have emphasized the importance of mining proteomics data repositories for clinical research, functional genomics, and understanding of post-translational modifications [386-391].

The Peptide Atlas Project (http://www.peptideatlas.org) [392, 393] is a proteomic database that permits high throughput, in-silico identification of tandem mass spectrometry based peptides, which are algorithmically mapped to the eukaryotic cell genome.. This database allows the robust in-depth interrogation of information deposited in a centralized repository of targeted proteomics assays [393, 394]. The PeptideAtlas database possesses capabilities for large scale quantitation and compilation of proteotypic peptides (PTPs) and proteins across multiple experiments [394]. PeptideAtlas has recently initiated a new repository known as the PeptideAtlas SRM Experiment Library (PASSEL) which collects single reaction monitoring (SRM) experiments and allows researchers access and submit targeted SRM proteomics datasets [236]. This database is also useful for optimizing transitions (which are parent-fragment ion pairs) for each peptide using SRMCollider software in conjunction with a new SRM database known as the SRMAtlas [395]. Furthermore, new data independent acquisition (DIA) mass spectrometry methods such as SWATH that combine high throughput data collection with consistent reproducibility can be more accurately reproduced by using this approach. SRMAtlas has established many integrated data builds, for instance N-Glycoproteome recapitulation across different datasets for different cancer types have been demonstrated by SRMAtlas [396]. Certainly, databases such as SRMAtlas, PASSEL and PeptideAtlas are effective web-based resources for the verification of potential biomarkers [397].

Another valuable online resource is the Human Protein Atlas (http://www.proteinatlas.org/), which creates an atlas of ca. 400,000 high quality images of localization characteristics and expression level of over 700 human proteins in 20 and 48 different cancerous and normal human tissues respectively, using antibody-centered proteomics [398]. Information included in the database includes western blots, immunohistochemistry and fluorescent digital imaging of tissues. Rigorous recombinant affinity purification with careful antibody design of protein epitope signature tags (PrESTs) was carried out [398, 399]. Antibodies used in this database are only those with reduced sequence homology to any other proteins of human origin and also increased tissue specificity. Stringent quality assurance was followed for assigning bioinformatics validation scores to proteins when comparing literature- to experiment-based information [398]. This database offers a promising approach to biologic screening and verification of potential biomarkers to assess tissue specificity and to provide a preliminary in-silico differential expression profile of biomarkers between normal and cancerous tissues.

A triple quadrupole (QQQ) mass spectrometer is traditionally described as the workhorse of targeted Single Reaction Monitoring (SRM). In a QQQ, the first and third quadrupoles are mass filters, and the second quadrupole is a collision cell [233]. This setup allows for accurate identification and quantitation of transitions and it is good for targeting and absolute quantification of analytes. However assays on this instrument requires intensive optimization for each transition and there is limited capacity for multiplexing. In addition, expensive synthetic peptides are often required as standards for quantitative assays. However, Parallel Reaction Monitoring (PRM) in principle has overcome some of the setbacks experienced with SRM assays, albeit PRMs are better for survey mode of targeted proteomics rather than quantitation. Using a high resolution, high mass accuracy instrument such as the hybrid Quadrupole-Orbitrap mass spectrometer (QQ-Orb), PRM can be performed for multiplexed analysis of peptide transitions. This instrument has a modus operandi that is somewhat similar to the QQQ in some respects, and allows for a simultaneous highly multiplexed quantitation and identification of multiple transitions. The basic difference between the QQQ and the QQ-Orb is that the third quadrupole is replaced by an Orbitrap [238, 400]. With remarkable improvement in instrument resolution and mass accuracy, PRM has been demonstrated to provide comparable performance metrics with SRM in terms of dynamic range, precision, and linearity [401].

This study demonstrates intermediate in-silico verification and potential biomarker prevalidation of potential PCa biomarkers using parallel reaction monitoring. The approach used here could potentially improve the potential biomarker validation pipeline and improve the number of clinically useful biomarkers.

4.2. Methodology

4.2.1. Sample Cohort and Experimental Design

From **Chapter 3**, a total of 73 potential biomarkers of PCa were discovered, as well as nine biomarkers that demonstrated racial trends [319]. To prevalidate these 82 potential urinary biomarkers of prostate cancer, in-silico verification and PRM prevalidation was carried out on pooled samples of PCa (n=15) and NC (n=15), derived from the same cohort (**Table 3.1**). Every PCa patient that participated in this study had localized primary disease (≤TNM Stage III) without metastasis. Before PRM, in-silico approaches were employed to streamline the list of 82 discovered potential biomarkers for top ranking candidate biomarkers. This was followed by retrospective confirmation of top ranking biomarkers in the discovery shotgun urinary proteomics database we generated as well as 14 other publicly available proteomics databases (**Figure 4.1**).

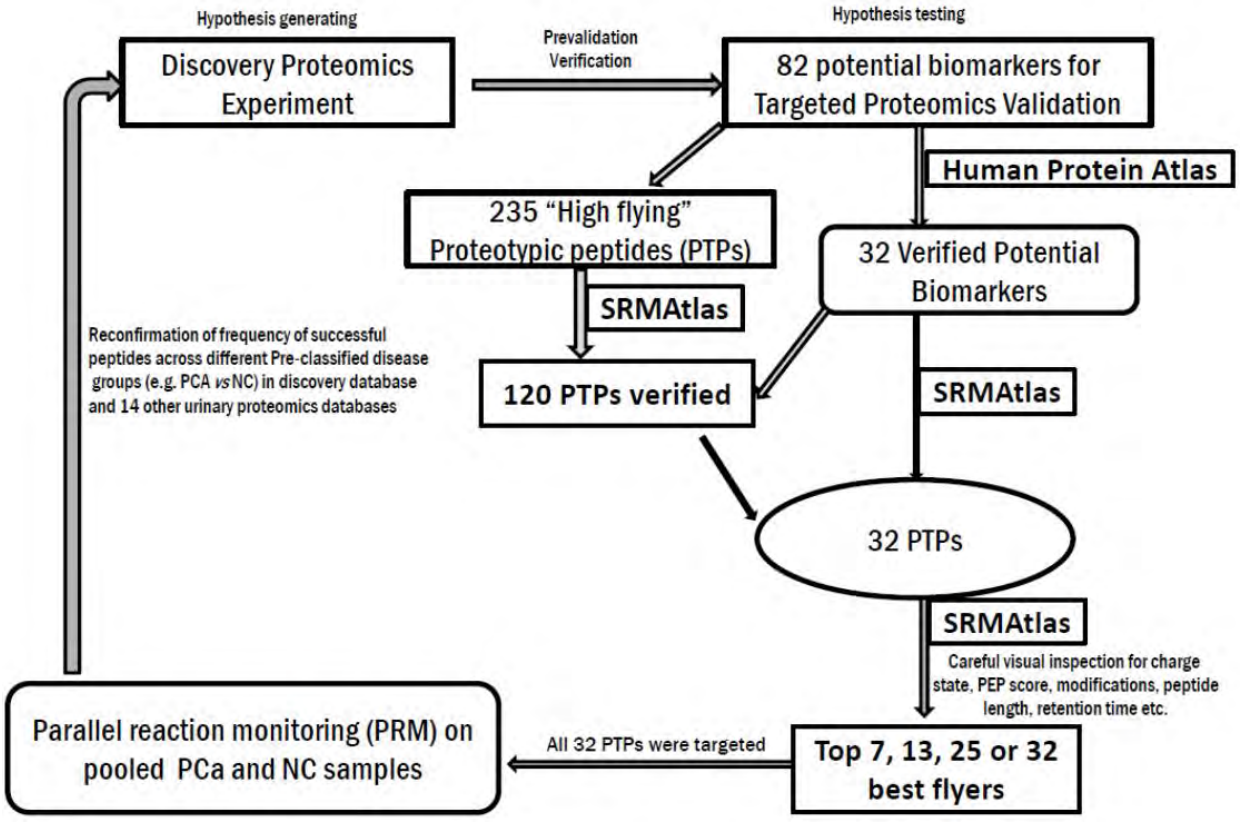


Figure 4.1. Overview of In-silico verification and PRM prevalidation of potential prostate cancer biomarkers [402]

4.2.2. Biologic Screening with Human Protein Atlas

Initial biological screening of the potential biomarkers was done by assessing the performance of target biomarkers in the Human Protein Atlas database. A detail of immunohistochemical procedure used for the Human Protein Atlas has been well described elsewhere [398]. In summary, specialized tissue microarrays (TMAs) optimized with serial antibody dilutions were used. Recipient TMA blocks had both negative and positive controls arrayed on it. Citrate buffer (pH 6.0) and autoclave method was used for antigen retrieval. Diaminobenzidine (DAB) was used to develop the immunohistochemical slides and they were counterstained with hematoxylin. Slides were scanned with an automated scanner; and then annotated and scored by certified pathologists. Immunohistochemical staining of cancerous and normal prostate tissue was assessed by searching either the name or the gene symbol of the protein of interest. Target mRNA expression level in normal and cancerous prostate tissue, as well as human cell lines were assessed; even though it is widely known that mRNA may not correlate with protein expression. Subcellular localization of target proteins is documented in this database as well. Considering that a large proportion of the discovered potential biomarkers were predicted to be of extracellular matrix and membrane protein origin, stromal staining was assessed in determining presence or absence of biomarkers. Differential staining of the identified potential biomarkers between normal and cancerous prostate tissue was used as a surrogate marker for performance and prostate tissue specificity. Expression of potential biomarkers in normal and cancerous prostate tissue or cell lines both at the mRNA or protein level was used as stringency criteria to streamline candidates for further targeted proteomics. In-silico verification with Human Protein Atlas and SRMAtlas were used to increase the stringency for potential PCa biomarker prior to targeted PRM prevalidation

4.2.3. Peptide Selection

To establish that target peptides are unique and specific for a particular protein in targeted proteomics, a set of proteotypic peptides (PTP) need to be selected. Successful selection of such

peptides is heavily dependent on their fragmentation pattern during tandem mass spectrometry. Proteotypic peptides were selected as described in **Section 2.3.4.**

4.2.4. In-Silico Verification of PTPs

In silico verification of selected peptides was performed using the SRMAtlas database. This database is a compilation of high quality SRM experiments for quantitation and identification of proteins. Over 170,000 peptides from the human proteome have been identified by this database with ca. 99.9% coverage for human proteins (http://www.srmatlas.org/). The presence of targeted proteomics evidence of previous peptide assay in this database increases confidence in the potential using such peptides as PTPs in this study. Selected proteotypic peptides for our potential biomarkers were verified in the SRM database using the setting described in **Section 2.3.5**.

4.2.5. Biomarker Prevalidation with Parallel Reaction Monitoring (PRM)

After in-silico biologic screening of potential biomarkers with the Human Protein Atlas database and target PTP verification using the SRMAtlas; top ranking biomarkers were prevalidated with parallel reaction monitoring on a QExactiveTM Mass spectrometer in line with Dionex UltiMate® 3500 RSnano UPLC system (Thermo Fisher, San Jose, CA, USA). This initial targeted PRM experiment is aimed at evaluating the selected "best flying" PTPs prior to definitive targeted proteomics. Steps involved in PRM are described in **Section 2.3.7**, and results generated were analysed using Skyline which is an open source analysis tool for targeted proteomics [239]. Because this was a preliminary survey type of targeted proteomics experiment, and observed mean retention time range for each selected PTP varied remarkably (**Figure 4.2**), the experiments were not scheduled. Singly injected samples were used as it was observed in previous experiments that there was very little variation in transitions detected in multiply injected peptides from the same urine samples (**Figure 4.3**). Besides, the high resolution and high mass accuracy of the instrument used contributed to the accuracy of the findings.

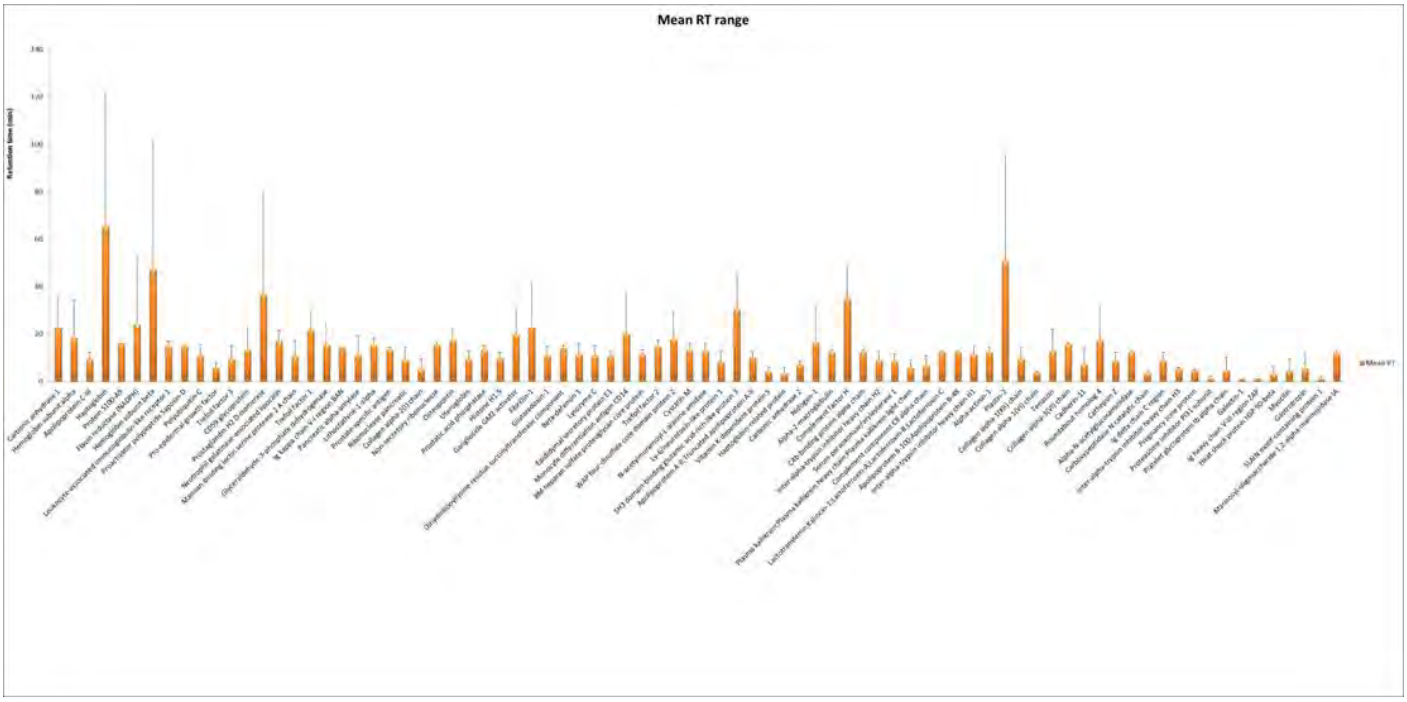


Figure 4.2. Bar graph demonstrating variability in mean retention time range in selected proteotypic peptides

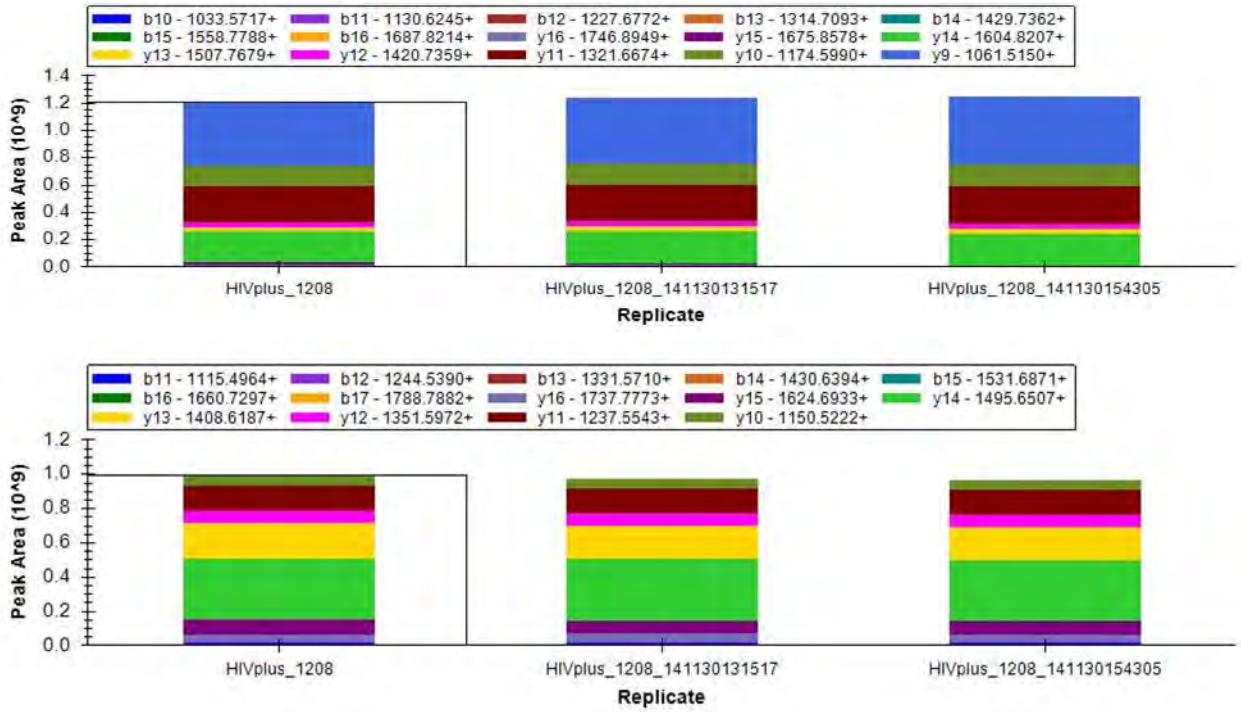


Figure 4.3. Stacked bar graph showing low variations in peak area of transitions detected in multiple peptide injection. Each sub-section of one bar in this stacked bar graph is the area under the curve (AUC) of a single transition which can be easily compared between different groups.

4.2.6. Ultra-High Performance Liquid Chromatography

The same liquid chromatography gradients and column dimensions used in shotgun discovery proteomics experiments were used for PRM. Pooled urinary peptides from 15 prostate cancer and 15 normal healthy control participants at 50ng/µl (in HPLC grade water containing 0.1% (v/v) Formic acid) were subjected to Nanoflow UHPLC in-line on a Dionex UltiMate® 3500 RSnano UPLC system (Thermo Fisher, San Jose, CA, USA) equipped with a 75µm×50cm; 5 µm; 100Å C-18 resin packed analytic column and a 100µm×5cm; 5µm; 100Å C-18 resin packed pre-column. Gradient chromatography was carried as described in **Section 2.3.1.**

4.2.7. Quadrupole-Orbitrap Analyses

A QExactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher, San Jose, CA, USA) was used for parallel reaction monitoring. This instrument combines accurate mass Orbitrap detection with high resolution quadrupole precursor ion selection. A spray current of 0.1µA, an ionization voltage of 1.86kV, an S-lens RF level of 50.0, and a capillary temperature of 320^0C was used for the nano-electrospray ionization source. An unscheduled *4-plex* tandem mass spectrometry targeted methodology was used at a mass spectra acquisition resolution of 35000. All Ion Fragmentation (AIF) full scan mode was used with positive polarity at a scan range of 79 to 1945 *m/z*, charge exclusion z = 2, and maximum time per peptide of 30 milliseconds. A total scan time of 140 milliseconds and required cycle time was 2 second. A fixed automatic gain control (AGC) value of $5 \times E^6$ was used. In-source High-energy Collision Dissociation (HCD) with Normalized Collision Energy (NCE) set at 27eV was used for peptide fragmentation. Acquisition method comprises of targeted PRM scan events directed at the +2 and +3 PTP precursor charge states. For each precursor, an isolation window of 2m/z units was used at a starting *m/z* of 80 and followed by automatic derivation of final *m/z* from the charge state and PTP precursor *m/z*. Upon generation of MS2 data from PRM experiments, raw data files were imported into Skyline to determine precursors with good signals and high concordance of charge states in terms of retention time and intensity. With the isolation list, this process is repeated using a different order to ascertain that multiplexing interferences did not bias generated

results. Further quantitative analyses were performed to compare biomarker transitions between pooled normal and PCa samples.

4.2.8. Database Verification of Candidate Biomarkers

Candidate biomarkers that were observed to have good performance by human protein atlas, SRMAtlas and parallel reaction monitoring were identified. These biomarkers were further re-confirmed using the previous shotgun urinary proteomics database. Their differential expression in PCa, BPH and normal healthy controls were assessed in this database. In addition, their performance was also confirmed in 14 other publicly available urinary proteomics databases to assess their ability to differentiate between PCa, BPH and normal healthy states.

4.2.9. Functional Network Association Analysis

Top ranking potential biomarkers by this approach in addition to well established PCa pathways like PTEN and ETS was performed using GENEMANIA (http://www.genemania.org/). Predicted interaction between these top ranking biomarkers with pre-exisiting PCa pathogenesis pathways may offer an insight into the pathogenesis of PCa in this cohort of patients.

4.3. Results

4.3.1. Screening of Candidate Biomarkers

Even though three proteotypic peptides were targeted for each biomarker, a few biomarkers had less than 3 identified peptides. Hence, a total of 235 proteotypic peptides were selected (**Table 4.1**). These 235 selected proteotypic peptides were characterized with posterior error probability (PEP) scores, peptide length (PL), retention time ranges, charge states, and modification status. They were further characterized with results from biologic screening with Human Protein Atlas (HPA) and SRMAtlas verification. Using the features described in **Table 4.1**, the list of 82 potential biomarkers were tailored down to a list of 32 top ranking potential biomarkers found most suitable for targeted proteomics analysis (**Table 4.2**). Higher priority was given to peptides with 2+ and 3+ charges in comparison to 1+ or >3+ charges. Medium sized peptides were preferred to longer ones. Peptides with a narrower range of retention time were preferred over those with wider retention time window. We found 115 (48.9%) of our 235 PTPs were not previously reported in the SRMAtlas database while 120 (51.1%) were already reported in this database.

	Potential Biomarker ID	**Potential biomarkers**	PEP Scores	PL	Peptides	C	m/z	mass	Modification	SRM Atlas	HPA
1	P00915;E5RHP7;E5RH81;E5RFE7	Carbonic anhydrase 1	1.86E-138	12	ADGLAVIGVLMK	2	593.8468	1185.679	Unmodified	yes	low expression in PCa
				9	GGPFSDSYR	2	493.22233	984.43011	Unmodified	no	
				14	ESISVSSEQLAQFR	2	790.89938	1579.7842	Unmodified	no	
2	P69905	Hemoglobin subunit alpha	0	12	FLASVSTVLTSK	2	626.861	1251.7075	Unmodified	Yes	low expression in PCa
				15	VGAHAGEYGAEALER	2	765.37079	1528.727	Unmodified	Yes	
				11	VLSPADKTNVK	2	586.3377	1170.6608	Unmodified	no	
3	B0YIW2;P02656	Apolipoprotein C-III	4.32E-64	16	DALSSVQESQVAQQAR	2	858.9292	1715.8438	Unmodified	Yes	Not yet published, no mRNA data in tissue
				11	GWVTDGFSSLK	2	598.80095	1195.5873	Unmodified	yes	
				16	DALSSVQESQVAQQAR	3	572.95522	1715.8438	Unmodified	yes	
4	P00738;J3QR68;Q0VAC5;J3QLC9	Haptoglobin	0	12	DIAPTLTLYVGK	2	645.86883	1289.7231	Unmodified	no	low expression in PCa
				10	LPECEAVCGK	2	581.76519	1161.5158	Unmodified	no	
				12	SCAVAEYGVYVK	2	673.32647	1344.6384	Unmodified	no	
5	P06702	Protein S100-A9	0	15	VIEHIMEDLDTNADK	2	871.91691	1741.8193	Unmodified	yes	Moderate to low expression in PCa
				15	VIEHIMEDLDTNADK	3	581.6137	1741.8193	Unmodified	yes	
				13	LGHPDTLNQGEFK	2	728.36497	1454.7154	Unmodified	yes	
6	P30043	Flavin reductase (NADPH)	1.54E-17	10	LQAVTDDHIR	2	584.30947	1166.6044	Unmodified	no	High to medium expression in PCa
				15	TVAGQDAVIVLLGTR	2	756.94085	1511.8671	Unmodified	yes	
				17	PAHVVVGDVLQAADVDK	3	578.31246	1731.9155	Unmodified	no	
7	P68871;F8W6P5	Hemoglobin subunit beta	0	12	EFTPPVQAAYQK	2	689.85371	1377.6929	Unmodified	no	Moderate to low expression in PCa
				13	GTFATLSELHCDK	2	739.85084	1477.6871	Unmodified	yes	
				16	VLGAFSDGLAHLDNLK	2	835.44904	1668.8835	Unmodified	no	
8	Q6GTX8;D3YTC8;Q6GTX8-2	Leukocyte-associated	2.64E-	9	GPVGVQTFR	2	480.76671	959.51886	Unmodified	yes	Not detected in PCa

		immunoglobulin-like receptor 1	304	13	IDSVSEGNAGPYR	2	682.82568	1363.6368	Unmodified	yes	
				13	WSEQSDYLELLVK	2	805.40905	1608.8035	Unmodified	yes	
9	B1AVU8;P07602;P07602-3	Proactivator polypeptide;Saposin-A;Saposin-B-Val;Saposin-B;Saposin-C;Saposin-D	3.02E-171	11	GCSFLPDPYQK	2	656.30554	1310.5965	Unmodified	yes	High expression in PCa
				10	LGPGMADICK	2	531.25954	1060.5045	Unmodified	yes	
				13	SDVYCEVCEFLVK	2	824.37329	1646.732	Unmodified	no	
10	P0CG48;F5H041;F5H7K6;Q96C32	Polyubiquitin-C;Ubiquitin;Polyubiquitin-B;Ubiquitin;	0	9	EGIPPDQQR	2	520.26199	1038.5094	Unmodified	yes	Very high to medium expression in PCa
				13	IQDKEGIPPDQQR	2	762.39426	1522.774	Unmodified	no	
				16	TITLEVEPSDTIENVK	2	894.46729	1786.92	Unmodified	yes	
11	P01133;E7EVD2;P01133-2	Pro-epidermal growth factor;Epidermal growth factor	0	11	AEDDTWEPEQK	2	674.2886	1346.5626	Unmodified	no	Not detected in PCa
				14	CISEGEDATCQCLK	2	835.84457	1669.6746	Unmodified	no	
				11	IESSSLQGLGR	2	573.8093	1145.6041	Unmodified	no	
12	E9PBB5;Q07654	Trefoil factor 3	2.52E-61	9	IPGVPWCFK	2	552.28896	1102.5634	Unmodified	no	Moderate to low expression in PCa
				18	IPGVPWCFKPLQEAECTF	2	1090.0212	2178.0278	Unmodified	yes	
				11	VDCGYPHVTPK	2	636.80571	1271.5969	Unmodified	yes	
13	E9PR17;P13987;E9PNW4	CD59 glycoprotein	0	8	AGLQVYNK	2	446.74798	891.48142	Unmodified	yes	Not detected in PCa
				10	ENELTYYCCK	2	690.28395	1378.5533	Unmodified	no	
				16	TAVNCSSDFDACLITK	2	901.40839	1800.8022	Unmodified	yes	
14	H0Y5A1;P41222;F8W7I2;Q5SQ11	Prostaglandin-H2 D-isomerase	0	17	AQGFTEDTIVFLPQTDK	2	955.48073	1908.9469	Unmodified	yes	High to medium expression in PCa
				7	GPGEDFR	2	389.17993	776.34532	Unmodified	no	
				16	TMLLQPAGSLGSYSYR	2	872.44035	1742.8662	Unmodified	yes	
15	H9KV70;P80188;P80188-2	Neutrophil gelatinase-associated lipocalin	0	9	MYATIYELK	2	566.29137	1130.5682	Unmodified	yes	low expression in PCa
				11	SYPGLTSYLVR	2	628.3377	1254.6608	Unmodified	no	
				15	VPLQQNFQDNQFQGK	2	895.94465	1789.8747	Unmodified	yes	

16	O00187;O00187-2	Mannan-binding lectin serine protease 2 A chain	0	10	WTLTAPPGYR	2	581.3062	1160.5978	Unmodified	yes	low expression in PCa
				15	DTFYSLGSSLDITFR	2	861.42268	1720.8308	Unmodified	no	
				16	LASPGFPGEYANDQER	2	875.90519	1749.7958	Unmodified	yes	
17	P04155	Trefoil factor 1	9.93E-101	9	GCCFDDTVR	2	565.22369	1128.4328	Unmodified	yes	Not detected in PCa
				18	ERQNCGFPGVTPSQCANK	2	1025.4651	2048.9156	Unmodified	no	
				16	QNCGFPGVTPSQCANK	2	882.89325	1763.7719	Unmodified	yes	
18	O14556	Glyceraldehyde-3-phosphate dehydrogenase	2.54E-07	11	AGIALNDNFVK	2	581.31676	1160.619	Unmodified	Yes	Medium expression in PCa
				7	LTGMAFR	2	398.21272	794.41089	Unmodified	Yes	
				7	LTGMAFR	2	406.21018	810.40581	Oxidation (M)	Yes	
19	P04430	Ig kappa chain V-I region BAN	3.46E-120	18	DIQLTQSPSSLSASVGDR	2	930.96852	1859.9225	Unmodified	Yes	No data available
				18	DIQLTQSPSSLSASVGDR	3	620.98144	1859.9225	Unmodified	Yes	
20	P04746	Pancreatic alpha-amylase	0	11	LTGLLDLALEK	2	593.3581	1184.7016	Unmodified	no	Not detected in PCa
				13	NWGEGWGFVPSDR	2	753.84166	1505.6688	Unmodified	no	
				21	GFGGVQVSPPNENVAIYNPFR	2	1131.5687	2261.1229	Unmodified	no	
21	P05451	Lithostathine-1-alpha	9.99E-239	11	ISCPEGTNAYR	2	634.29042	1266.5663	Unmodified	yes	Not detected in PCa
				10	SYCYYFNEDR	2	708.77988	1415.5452	Unmodified	yes	
				17	ESGTDDFNVWIGLHDPK	2	965.4525	1928.8905	Unmodified	yes	
22	P07288;G3XAE3;G3V0H4	Prostate-specific antigen	3.13E-101	12	LSEPAELTDAVK	2	636.83773	1271.6609	Unmodified	yes	Very high to medium expression in PCa
				9	IVGGWECEK	2	539.25532	1076.4961	Unmodified	no	
				7	SVILLGR	2	379.25016	756.48577	Unmodified	no	
23	P07998;G3V357	Ribonuclease pancreatic	0	21	PVNTFVHEPLVDVQNVCFQEK	3	833.74941	2498.2264	Unmodified	no	Not detected in PCa

				24	HIIVACEGSPYVPVHFDASVEDST	2	1315.1156	2628.2166	Unmodified	yes	
				7	YPNCAYR	2	472.20817	942.40179	Unmodified	yes	
24	P08123	Collagen alpha-2(I) chain	7.71E-23	15	GEAGAAGPAGPAGPR	2	618.31	1234.6054	Unmodified	yes	No data on cancer
				11	GPAGPSGPAGK	2	448.23524	894.45593	Unmodified	no	
				18	GETGPSGPVGPAGAVGPR	2	781.89971	1561.7849	Unmodified	yes	
25	P10153	Non-secretory ribonuclease	0	17	YAQTPANMFYIVACDNR	2	1017.464	2032.9135	Unmodified	yes	Medium expression in PCa
				17	YAQTPANMFYIVACDNR	3	678.64512	2032.9135	Unmodified	yes	
				14	DPPQYPVVPVHLDR	2	816.43065	1630.8467	Unmodified	yes	
26	P10451;P10451-4;P10451-2;D6R9C5	Osteopontin	0	17	AIPVAQDLNAPSDWDSR	2	927.95267	1853.8908	Unmodified	yes	low expression in PCa
				9	GDSVVYGLR	2	483.25617	964.49779	Unmodified	no	
				13	ISHELDSASSEVN	2	694.32043	1386.6263	Unmodified	yes	
27	P11684;E9PN95	Uteroglobin	2.44E-68	10	LVDTLPQKPR	2	583.84823	1165.6819	Unmodified	yes	low expression in PCa
				10	LVDTLPQKPR	3	389.56791	1165.6819	Unmodified	yes	
				27	VIETLLMDTPSSYEAAMELFSPDQDMR	3	1030.4766	3088.4079	Unmodified	yes	
28	P15309-2;P15309;Q5FBY0;E9PFE6	Prostatic acid phosphatase;PAPf39	0	8	ATQIPSYK	2	454.24782	906.48108	Unmodified	no	High expression in PCa
				10	FQELESETLK	2	612.31134	1222.6081	Unmodified	no	
				12	SPIDTFPTDPIK	2	665.84809	1329.6816	Unmodified	yes	
29	P16401	Histone H1.5	3.93E-22	11	ALAAGGYDVEK	2	547.27985	1092.5451	Unmodified	yes	medium to low expression in PCa
				9	NGLSLAALK	2	443.77146	885.52837	Unmodified	yes	
				12	ATGPPVSELITK	2	606.84535	1211.6762	Unmodified	yes	
30	P17900;H0YBY3	Ganglioside GM2 activator;Ganglioside GM2 activator isoform	3.71E-88	10	IESVLSSSGK	2	503.7744	1005.5342	Unmodified	yes	High to medium expression in PCa
				7	VDLVLEK	2	408.24729	814.48002	Unmodified	yes	
				8	EGTYSLPK	2	447.732	893.44945	Unmodified	no	

		short									
31	P35555	Fibrillin-1	0	13	YEDEECTLPIAGR	2	776.85104	1551.6875	Unmodified	no	Not detected in PCa
				14	CLCPEGFSLSSSGR	2	778.84523	1555.6759	Unmodified	no	
				9	AGYQSTLTR	2	498.75908	995.50361	Unmodified	yes	
32	P35754	Glutaredoxin-1	4.25E-97	11	AQEILSQLPIK	2	620.369	1238.7234	Unmodified	yes	High to medium expression in PCa
				21	DCIGGCSDLVSLQQSGELLTR	2	1154.549	2307.0835	Unmodified	yes	
				12	RAQEILSQLPIK	2	698.41955	1394.8245	Unmodified	no	
33	P36957;Q86SW4;B7Z5W8;G3V3F0;G3V5M3;H0YJF9	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	6.86E-52	14	TPAFAESVTEGDVR	2	739.85972	1477.7049	Unmodified	yes	Medium expression in PCa
				11	VEGGTPLFTLR	2	595.33241	1188.6503	Unmodified	no	
34	P60022	Beta-defensin 1	6.80E-123	7	IQGTCYR	2	449.21599	896.41744	Unmodified	no	No data on cancer
				23	SDHYNCVSSGGQCLYSACPIFTK	2	1326.0698	2650.125	Unmodified	yes	
				23	SDHYNCVSSGGQCLYSACPIFTK	3	884.38229	2650.125	Unmodified	Yes	
35	P61626;F8VV32	Lysozyme C	0	8	WESGYNTR	2	506.72778	1011.441	Unmodified	no	Not detected in PCa
				7	YWCNDGK	2	471.69235	941.37015	Unmodified	Yes	
				12	STDYGIFQINSR	2	700.84387	1399.6732	Unmodified	Yes	
36	P61916;G3V3D1;E7EMS2	Epididymal secretory protein E1	8.38E-225	9	SGINCPIQK	2	508.76331	1015.5121	Unmodified	Yes	High to medium expression in PCa
				9	DKTYSYLNK	2	566.28767	1130.5608	Unmodified	no	
				16	EVNVSPCPTQPCQLSK	2	922.43749	1842.8604	Unmodified	yes	
37	P08571;D6RFL4	Monocyte differentiation antigen CD14;Monocyte differentiation antigen	0	9	ATVNPSAPR	2	456.74852	911.48248	Unmodified	no	Not detected in PCa
				8	ELTLEDLK	2	480.76604	959.51753	Unmodified	no	
				10	FPAIQNLALR	2	571.83767	1141.6608	Unmodified	no	

		CD14, urinary form;Monocyte differentiation antigen CD14, membrane-bound form									
38	P98160	Basement membrane-specific heparan sulfate proteoglycan core protein;Endorepellin;LG 3 peptide	0	11	SPGPNVAVNAK	2	527.28801	1052.5615	Unmodified	no	Medium expression in PCa
				12	TPSGLYLGTCER	2	677.327	1352.6395	Unmodified	no	
				10	FSSGITGCVK	2	528.26314	1054.5117	Unmodified	no	
39	Q03403	Trefoil factor 2	8.14E-160	13	QESDQCVMEVSDR	2	791.82724	1581.6399	Unmodified	yes	Not detected in PCa
				15	NCGYPGISPEECASR	2	848.85633	1695.6981	Unmodified	yes	
				15	NCGYPGISPEECASR	3	566.23998	1695.6981	Unmodified	yes	
40	Q14508;Q14508-3;Q14508-2	WAP four-disulfide core domain protein 2	0	16	CCSAGCATFCSLPNDK	2	924.36261	1846.7107	Unmodified	no	Not detected in PCa
				14	DQCQVDSQCPGQMK	2	840.84236	1679.6702	Unmodified	yes	
				18	EGSCPQVNINFPQLGLCR	2	1045.0013	2087.9881	Unmodified	yes	
41	Q15828	Cystatin-M	1.18E-281	10	AQSQLVAGIK	2	507.80075	1013.5869	Unmodified	yes	low expression in PCa
				11	DLSPDDPQVQK	2	621.30405	1240.5935	Unmodified	yes	
				13	YFLTMEMGSTDCR	2	805.83602	1609.6575	Unmodified	no	
42	Q96PD5-2;Q96PD5	N-acetylmuramoyl-L-alanine amidase	4.42E-138	13	TDCPGDALFDLLR	2	746.85867	1491.7028	Unmodified	yes	medium to low expression in PCa
				9	DTLPSCAVR	2	509.75294	1017.4913	Unmodified	no	
				13	GCPDVQASLPDAK	2	679.32446	1356.6344	Unmodified	no	
43	Q9BZG9;Q9BZG9-3;H0YB12	Ly-6/neurotoxin-like protein 1	7.06E-298	13	VLSNTEDLPLVTK	2	714.90086	1427.7872	Unmodified	no	low expression in PCa
				11	DSTHCVTTATR	2	624.7855	1247.5564	Unmodified	no	
				11	DSTHCVTTATR	3	416.85943	1247.5564	Unmodified	no	
44	Q9H299;Q5T123	SH3 domain-binding glutamic acid-rich-like protein 3	0	10	VYSTSVTGSR	2	528.76965	1055.5247	Unmodified	yes	Not detected in PCa
				15	IQYQLVDISQDNALR	2	888.46796	1774.9214	Unmodified	yes	
				15	IQYQLVDISQDNALR	3	592.64773	1774.9214	Unmodified	yes	

45	P02652	Apolipoprotein A-II;Truncated apolipoprotein A-II	0	9	SPELQAEAK	2	486.75346	971.49238	Unmodified	yes	Not detected in PCa
				8	EQLTPLIK	2	471.28694	940.55933	Unmodified	yes	
				10	SKEQLTPLIK	2	578.85044	1155.6863	Unmodified	no	
46	P07225;G5E9F8	Vitamin K-dependent protein S	0	8	HCLVTVEK	2	493.2604	984.50625	Unmodified	yes	Medium to low expression in PCa
				15	SQDILLSVENTVIYR	2	875.47271	1748.9309	Unmodified	no	
				8	VYFAGFPR	2	478.75307	955.49159	Unmodified	no	
47	P00739-2;P00739	Haptoglobin-related protein	1.27E-292	7	NYAEVGR	2	404.70103	807.38752	Unmodified	no	medium to low expression in PCa
				14	VGYVSGWGQSDNFK	2	772.36243	1542.7103	Unmodified	yes	
				10	VTSIQHWVQK	2	613.33803	1224.6615	Unmodified	yes	
48	P00918	Carbonic anhydrase 2	4.36E-10	9	VVDVLDSIK	2	494.28968	986.56481	Unmodified	no	Not detected in PCa
				10	SADFTNFDPR	2	585.26473	1168.5149	Unmodified	yes	
				12	QSPVDIDTHTAK	2	656.3306	1310.6466	Unmodified	yes	
49	P14543;P14543-2	Nidogen-1	7.43E-118	10	EDLSPSITQR	2	573.29348	1144.5724	Unmodified	no	low expression in PCa
				12	SNGAYNIFANDR	2	671.31274	1340.6109	Unmodified	no	
				15	GGADTYSVPSVLSPR	2	753.38336	1504.7522	Unmodified	no	
50	P01023	Alpha-2-macroglobulin	0	15	AAQVTIQSSGTFSSK	2	756.38864	1510.7627	Unmodified	no	low expression in PCa
				14	IAQWQSFQLEGGLK	2	802.9252	1603.8358	Unmodified	no	
				11	LPPNVVEESAR	2	605.82495	1209.6354	Unmodified	no	
51	P08603	Complement factor H	0	14	CFEGFGIDGPAIAK	2	741.3583	1480.7021	Unmodified	no	Not detected in PCa
				18	DTSCVNPPTVQNAYIVSR	2	1010.9915	2019.9684	Unmodified	no	
				10	TGDEITYQCR	2	621.7746	1241.5347	Unmodified	no	
52	P04003	C4b-binding protein alpha chain	3.36E-192	11	EDVYVVGTVLR	2	625.34298	1248.6714	Unmodified	Yes	Not detected in PCa
				14	FSAICQGDGTWSPR	2	791.35936	1580.7042	Unmodified	no	
				9	GYILVGQAK	2	474.77928	947.54402	Unmodified	no	
53	P19823;Q5T985	Inter-alpha-trypsin inhibitor heavy chain H2	0	9	SLAPTAAAK	2	415.24254	828.47052	Unmodified	no	Not detected in PCa
				15	IQPSGGTNINEALLR	2	791.93101	1581.8475	Unmodified	no	
				11	FYNQVSTPLLR	2	669.36424	1336.7139	Unmodified	yes	
54	P27169;F5H4W9	Serum paraoxonase/arylesterase	3.96E-95	19	VVAEGFDFANGINISPDGK	2	975.4838	1948.9531	Unmodified	no	medium to low expression in PCa

		1		16	IFFYDSENPPASEVLR	2	942.46234	1882.9101	Unmodified	yes	
				14	EVQPVELPNCNLVK	2	819.92962	1637.8447	Unmodified	yes	
55	H0YAC1;P03952;E9PBC5	Plasma kallikrein;Plasma kallikrein heavy chain;Plasma kallikrein light chain	7.42E-27	11	GVNVCQETCTK	2	648.28956	1294.5646	Unmodified	no	High to medium expression in PCa
				14	IAYGTQGSSGYSLR	2	730.36243	1458.7103	Unmodified	yes	
				13	LVGITSWGEGCAR	2	703.34827	1404.682	Unmodified	no	
56	P07357	Complement component C8 alpha chain	6.00E-22	17	AIDEDCSQYEPIPGSQK	2	968.93329	1935.852	Unmodified	yes	Data not avalilable for cancer
				10	HTSLGPLEAK	2	526.79038	1051.5662	Unmodified	yes	
				17	LGSLGAACEQTQTEGAK	2	860.91216	1719.8098	Unmodified	yes	
57	P02788;E7EQB2;P02788-2;E7ER44	Lactotransferrin;Kaliocin-1;Lactoferroxin-A;Lactoferroxin-B;Lactoferroxin-C	0	10	DGAGDVAFIR	2	510.75908	1019.5036	Unmodified	no	medium expression in PCa
				13	GGSFQLNELQGLK	2	695.86989	1389.7252	Unmodified	no	
				12	LADFALLCLDGK	2	668.35249	1334.6904	Unmodified	yes	
58	P04114	Apolipoprotein B-100;Apolipoprotein B-48	0	11	TEVIPPLIENR	2	640.86408	1279.7136	Unmodified	yes	Not detected in PCa
				13	TLADLTLLDSPIK	2	700.40577	1398.797	Unmodified	no	
				12	GFEPTLEALFGK	2	654.84535	1307.6762	Unmodified	yes	
59	P19827;F5H7E1;F5H165	Inter-alpha-trypsin inhibitor heavy chain H1	0	16	GFSLDEATNLNGGLLR	2	838.93375	1675.8529	Unmodified	no	Medium expression in PCa
				10	EVAFDLEIPK	2	580.81352	1159.6125	Unmodified	no	
				13	QAVDTAVDGVFIR	2	695.86989	1389.7252	Unmodified	yes	
60	P12814;P12814-3;P12814-2	Alpha-actinin-1	0	10	DLLLDPAWEK	2	600.31897	1198.6234	Unmodified	no	High to medium expression in PCa
				13	ELPPDQAEYCIAR	2	781.3694	1560.7242	Unmodified	no	
				14	FAIQDISVEETSAK	2	769.39085	1536.7672	Unmodified	no	
61	P13796	Plastin-2	0	9	NEALIALLR	2	506.81112	1011.6077	Unmodified	no	low expression in PCa
				8	ISFDEFIK	2	499.7633	997.51205	Unmodified	no	
				10	QFVTATDVVR	2	568.30894	1134.6033	Unmodified	no	
62	Q99715;D6RGG3;Q99715-4	Collagen alpha-1(XII) chain	4.23E-156	8	GPPGPPGR	2	367.70084	733.38712	Unmodified	no	High to medium expression in PCa
				10	ITEVTSEGFR	2	569.79058	1137.5666	Unmodified	no	

				12	ITVDPTTDGPTK	2	622.82207	1243.6296	Unmodified	no	
63	P12109	Collagen alpha-1(VI) chain	0	9	IALVITDGR	2	479.29002	956.56548	Unmodified	no	Not detected in PCa
				13	ENYAELLEDAFLK	2	777.88795	1553.7613	Unmodified	no	
				16	VAVVQYSGTGQQRPER	3	592.31134	1773.9122	Unmodified	yes	
64	P24821;P24821-4;J3QSU6	Tenascin	5.66E-149	16	CECDDGFTGADCGELK	2	917.33985	1832.6651	Unmodified	no	medium to low expression in PCa
				9	IQALNGPLR	2	491.29563	980.57671	Unmodified	no	
				10	LDAPSQIEVK	2	550.30332	1098.5921	Unmodified	no	
65	P12111;E9PCV6;E7ENL6	Collagen alpha-3(VI) chain	5.06E-177	12	FWYGGCGGNENK	2	694.78804	1387.5615	Unmodified	no	Not detected in PCa
				8	WYYDPNTK	2	543.74818	1085.4818	Unmodified	no	
				13	EVQVFEITENSAK	2	747.37775	1492.7409	Unmodified	no	
66	P55287;P55287-2;H3BUU9	Cadherin-11	6.06E-150	12	FFTINPEDGFIK	2	714.36391	1426.7133	Unmodified	Yes	Not detected in PCa
				10	VLDVNDNAPK	2	542.7853	1083.556	Unmodified	Yes	
				10	EEYHVVIQAK	2	608.32205	1214.6295	Unmodified	no	
67	Q8WZ75;Q8WZ75-2;Q8WZ75-3	Roundabout homolog 4	7.90E-124	10	GPDSNVLLLR	2	542.31148	1082.6084	Unmodified	Yes	No data on cancer
				7	LSVAVLR	2	379.25016	756.48577	Unmodified	no	
				14	MSCQASGQPPPTIR	2	765.36359	1528.7126	Unmodified	no	
68	Q9UBR2	Cathepsin Z	7.84E-50	12	NVDGVNYASITR	2	654.83077	1307.647	Unmodified	Yes	High to medium expression in PCa
				10	VGDYGSLSGR	2	505.74871	1009.4829	Unmodified	Yes	
				11	FNQCGTCNEFK	2	702.78719	1403.5598	Unmodified	no	
69	P54802	Alpha-N-acetylglucosaminidase;Alpha-N-acetylglucosaminidase 82 kDa form;Alpha-N-acetylglucosaminidase 77 kDa form	1.59E-90	15	PGLDTYSLGGGGAAR	2	696.34932	1390.6841	Unmodified	no	High to medium expression in PCa
				13	NVFQLEQAFVLSK	2	761.91684	1521.8191	Unmodified	Yes	
				16	LLGPGPAADFSVSVER	2	807.92794	1613.8413	Unmodified	Yes	
70	P15169;CON__Q2KJ83;B1AP58	Carboxypeptidase N catalytic chain	1.16E-09	15	EALIQFLEQVHQGIK	2	876.98579	1751.957	Unmodified	no	medium to low expression in PCa
				15	EALIQFLEQVHQGIK	3	584.99295	1751.957	Unmodified	no	

				9	IVQLIQDTR	2	543.31931	1084.6241	Unmodified	Yes	
71	P01880-2;P01880	Ig delta chain C region	3.40E-60	12	APDVFPIISGCR	2	666.34245	1330.6704	Unmodified	Yes	No data available
				22	VPAPPSPQPATYTCVVSHEDSR	3	799.04975	2394.1274	Unmodified	Yes	
				13	VPTGGVEEGLLER	2	678.3619	1354.7092	Unmodified	Yes	
72	Q06033;Q06033-2;E7ET33	Inter-alpha-trypsin inhibitor heavy chain H3	1.12E-32	15	EHLVQATPENLQEAR	2	867.94211	1733.8697	Unmodified	Yes	High to medium expression in PCa
				10	EVSFDVELPK	2	581.80315	1161.5918	Unmodified	Yes	
				22	STSIVIMLTDGDANVGESRPEK	3	773.72141	2318.1424	Unmodified	Yes	
73	P20742;P20742-2	Pregnancy zone protein	1.05E-240	18	GSFALSFPVESDVAPIAR	2	931.98598	1861.9574	Unmodified	no	High to medium expression in PCa
				13	TLLVEAEGIEQEK	2	729.88795	1457.7613	Unmodified	no	
				10	SLFTDLVAEK	2	561.8057	1121.5968	Unmodified	Yes	
74	Q92530;Q5QPM7;E7ER20;Q5QPM9	Proteasome inhibitor PI31 subunit	1.81E-19	17	AGLEVLFASAAPAITCR	2	894.96928	1787.924	Acetyl (Protein N-term)	no	High to medium expression in PCa
75	E7ES66;P07359	Platelet glycoprotein Ib alpha chain;Glycocalicin	7.48E-10	11	GLGELQELYLK	2	631.85318	1261.6918	Unmodified	Yes	Not detected in PCa
				13	WLQDNAENVYVWK	2	832.907	1663.7995	Unmodified	Yes	
				8	LTQLNLDR	2	486.77727	971.53999	Unmodified	Yes	
76	P09382;F8WEI7	Galectin-1	2.88E-10	8	SFVLNLGK	2	439.26073	876.5069	Unmodified	Yes	Not detected in PCa
				18	ACGLVASNLNLKPGECLR	2	1007.5141	2013.0136	Acetyl (Protein N-term)	no	
				18	ACGLVASNLNLKPGECLR	3	672.0118	2013.0136	Acetyl (Protein N-term)	no	
77	P01778	Ig heavy chain V-III region ZAP	0.000436	19	EVQLVESGGALVQPGGSGR	2	920.4816	1838.9486	Unmodified	NDA	No data available

78	P08238	Heat shock protein HSP 90-beta	1.73E-18	15	NPDDITQEEYGEFYK	2	924.40215	1846.7897	Unmodified	Yes	High to medium expression in PCa
				13	TLTLVDTGIGMTK	2	675.37088	1348.7272	Unmodified	no	
				13	TLTLVDTGIGMTK	2	683.36834	1364.7221	Oxidation (M)	no	
79	Q99972	Myocilin	3.82E-06	21	SGEGDTGCGELVWV GEPLTLR	2	1116.5337	2231.0528	Unmodified	Yes	medium expression in PCa
				7	TLTIPFK	2	410.25237	818.49019	Unmodified	no	
				17	LNPENLELEQTWET NIR	2	1050.0238	2098.0331	Unmodified	no	
80	P51161-2;P51161	Gastrotropin	4.58E-05	11	LLGISSDVIEK	2	587.3399	1172.6653	Unmodified	Yes	Not detected in PCa
				14	LVEVSTIGGVTYER	2	761.90921	1521.8039	Unmodified	no	
81	Q8ND83;Q8ND83-4;Q8ND83-2	SLAIN motif-containing protein 1	0.000581	7	LQEESLR	2	437.73507	873.4556	Unmodified	NDA	High to medium expression in PCa
82	P33908;Q6P052	Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	0.000116	18	EEGAPGDPEAALED NLAR	2	927.42923	1852.8439	Unmodified	Yes	High to medium expression in PCa

Table 4.1. Characterization of 235 proteotypic peptides from the list of 82 potential biomarkers of prostate cancer

Protein Name	Gene Name	PL	Sequence	Presence in PCa	Presence in NC	Level 1 stringency	Level 2 stringency	Level 3 Stringency	PRM pooled prevalidation
Flavin reductase (NADPH)	BLVRB	15	TVAGQDAVIVLLGTR	Very high (++)	high (+)	Yes	no	no	
Prostatic acid phosphatase	ACPP	12	SPIDTFPTDPIK	high (+)	high (+)	no	no	no	
N-acetylmuramoyl-L-alanine amidase	PGLYRP2	13	TDCPGDALFDLLR	Extrememly high (+++)	very low (--)	Yes	Yes	Yes	Good
Epididymal secretory protein E1	NPC2	16	EVNVSPCPTQPCQLSK	high (+)	very high (++)	Yes	no	no	
Glutaredoxin-1	GLRX	21	DCIGGCSDLVSLQQSGELLTR	Very high (++)	high (+)	Yes	no	no	
CD59 glycoprotein	CD59	16	TAVNCSSDFDACLITK	low (-)	high (+)	Yes	Yes	no	
Trefoil factor 3	TFF3	11	VDCGYPHVTPK	high (+)	low (-)	Yes	Yes	no	
Non-secretory ribonuclease	RNASE2	17	YAQTPANMFYIVACDNR	high (+)	low (-)	Yes	Yes	no	
Prostate-specific antigen	KLK3	12	LSEPAELTDAVK	high (+)	low (-)	Yes	Yes	no	
Proactivator polypeptide;Saposin	PSAP	11	GCSFLPDPYQK	high (+)	high (+)	no	no	no	
Vitamin K-dependent protein S	PROS1	8	HCLVTVEK	Very high (++)	low (-)	Yes	Yes	Yes	Good
Haptoglobin-related protein	HPR	10	VTSIQHWVQK	Very high (++)	low (-)	Yes	Yes	Yes	Good
Nidogen-1	NID1	10	EDLSPSITQR	high (+)	very high (++)	Yes	no	no	
Serum paraoxonase/arylesterase 1	PON1	16	IFFYDSENPPASEVLR	Very high (++)	high (+)	Yes	no	no	
Plasma kallikrein	KLKB1	14	IAYGTQGSSGYSLR	Very high (++)	high (+)	Yes	no	no	
Lactotransferrin	LTF	12	LADFALLCLDGK	Very high (++)	high (+)	Yes	no	no	
Inter-alpha-trypsin inhibitor heavy chain H1	ITIH1	13	QAVDTAVDGVFIR	high (+)	high (+)	no	no	no	

Carboxypeptidase N catalytic chain	CPN1	9	IVQLIQDTR	Extrememly high (+++)	high (+)	Yes	Yes	Yes	Good
Inter-alpha-trypsin inhibitor heavy chain H3	ITIH3	22	STSIVIMLTDGDANVGESRPEK	Very high (++)	very high (++)	no	no	no	
Pregnancy zone protein	PZP	10	SLFTDLVAEK	Extrememly high (+++)	high (+)	Yes	Yes	Yes	Good
Alpha-actinin-1	ACTN1	14	FAIQDISVEETSAK	Extrememly high (+++)	low (-)	Yes	Yes	Yes	Good
Plastin-2	LCP1	9	NEALIALLR	high (+)	low (-)	Yes	Yes	no	
Collagen alpha-1(XII) chain	COL12A1	12	ITVDPTTDGPTK	Very high (++)	high (+)	Yes	no	no	
Tenascin	TNC	16	CECDDGFTGADCGELK	Very high (++)	high (+)	Yes	no	no	
Cadherin-11	CDH11	12	FFTINPEDGFIK	low (-)	high (+)	Yes	Yes	no	
Cathepsin Z	CTSZ	12	NVDGVNYASITR	Very high (++)	high (+)	Yes	no	no	
Alpha-N-acetylglucosaminidase	NAGLU	16	LLGPGPAADFSVSVER	Very high (++)	Extremely high (+++)	Yes	no	no	
Proteasome inhibitor PI31 subunit	PSMF1	17	AGLEVLFASAAPAITCR	Very high (++)	very high (++)	no	no	no	
Heat shock protein HSP 90-beta	HSP90AB1	15	NPDDITQEEYGEFYK	Very high (++)	very high (++)	no	no	no	
Myocilin	MYOC	21	SGEGDTGCGELVWVGEPLTLR	Very high (++)	low (-)	Yes	Yes	Yes	Good
SLAIN motif-containing protein 1	SLAIN1	7	LQEESLR	Very high (++)	Extremely high (+++)	Yes	no	no	
Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	MAN1A1	18	EEGAPGDPEAALEDNLAR	Very high (++)	very high (++)	no	no	no	
				Total number of Biomarkers		**25**	**13**	**7**	
Level 1 Stringency: any difference in expression									
Level 2 stringency: Favour low (-) / high (+) and "Extrememly"									

Level 3 stringency: Greater than 2 unit of difference + Favour "Extremely"

PL= Protein Length

Table 4.2. List of top ranking 32 biomarker PTPs with increasing level of stringency based on various parameters [402]

4.3.2. In-Silico Screening with Human Protein Atlas

Initial screening of all 82 potential biomarkers with the human protein atlas yielded differential immunohistochemical staining patterns between normal and cancerous prostate tissues. Primary focus was on the top ranking 7 PTPs (**Table 4.2**), in addition to five other PTP that were shown to have good signals by PRM assays. A snapshot of differential immunohistochemical staining for all top 12 biomarkers confirmed by PRM is documented in **Annexure II**. Vitamin K-dependent protein S (PROS1) demonstrated moderate immunohistochemical staining of prostatic tissues, but demonstrated low staining of normal prostate tissue (**Figure 4.4**).

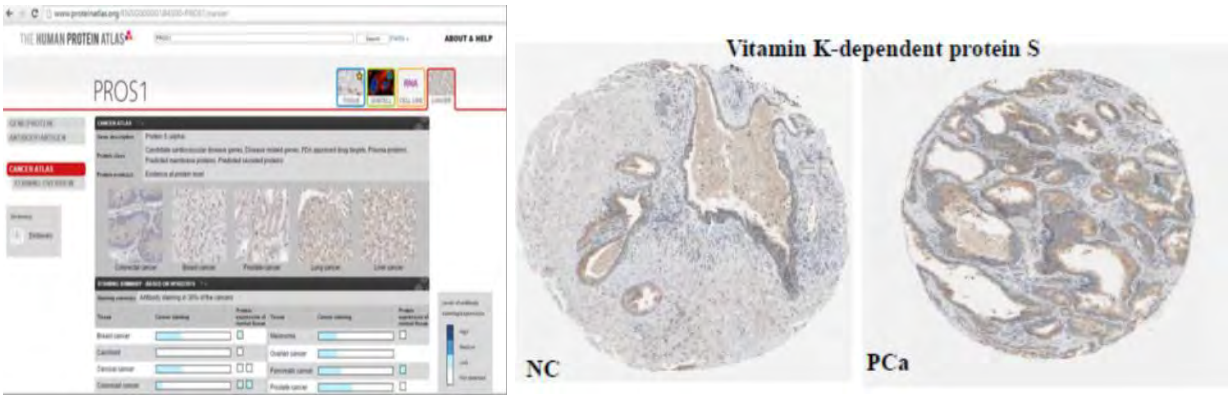


Figure 4.4. Differential staining pattern for Vitamin K-dependent protein S between prostate cancer and normal prostate tissues

N-acetylmuramoyl-L-alanine amidase (PGLYRP2) showed heavy staining ductal acinar cells with a non-specific staining of PCa stromal tissues, whilst there was low staining of normal prostate tissues for this biomarker. For prostate cancer tissues, Haptoglobin related protein (HPR) stained lightly positive, but no staining was observed for normal prostate tissues (**Figure 4.5**).

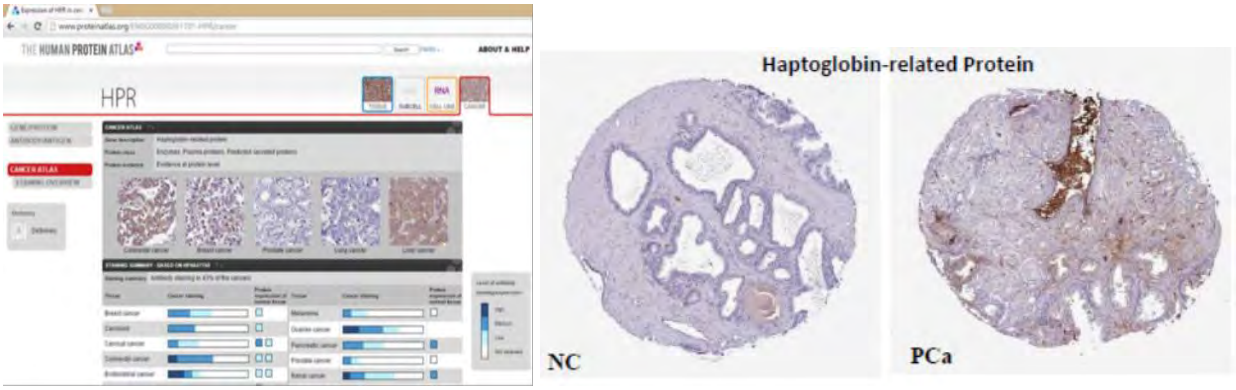


Figure 4.5. Differential staining pattern for Haptoglobin-related between prostate cancer and normal prostate tissues

Carboxypeptidase N catalytic chain (CPN1) stained positively in both cancerous and normal prostate; but this staining was stronger for prostate cancer. PCa was found to stain positive for Alpha-actinin-1 (ACTN1), but this biomarker stained negative in normal prostate. Pregnancy zone protein (PZP) stained positive in prostate cancer but negative for normal prostate tissue; albeit light staining of the stromal of normal prostate tissues was observed. It was found that Myocilin (MYOC) stained positive for PCa and negative for normal prostate tissues. In addition to these seven top ranking biomarkers, the following five potential protein biomarkers described hereafter had good transition signals during the PRM. Using the human protein atlas, Prostatic acid phosphatase (ACPP) was unable to immunohistochemically distinguish between normal tissue and PCa; showing diffuse positivity in prostatic epithelium under both conditions with mild stromal stains in PCa (**Figure 4.6**).

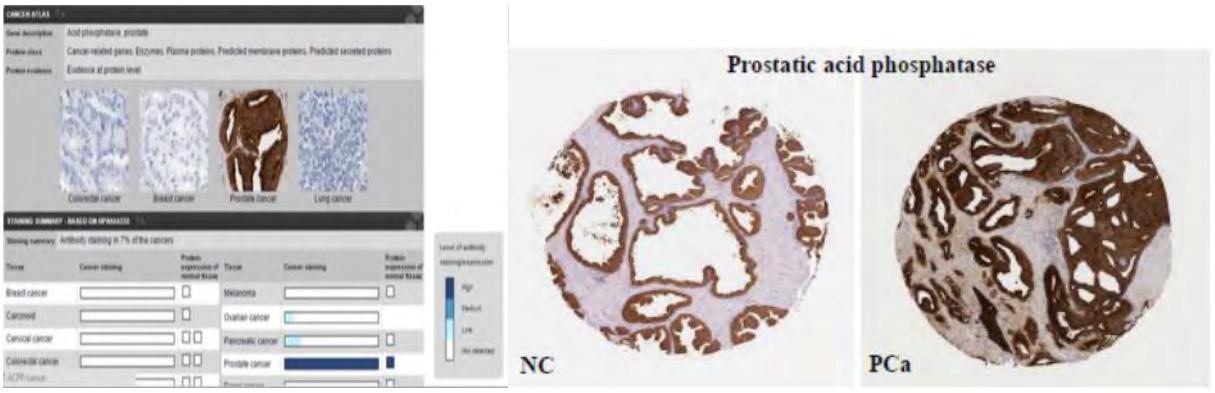


Figure 4.6. Differential staining pattern for prostatic acid phosphatase between prostate cancer and normal prostate tissues

To a reasonable extent, Prostate-specific antigen (KLK3) was able to distinguish between normal and cancerous prostate tissues. KLK3 stained positive in PCa, but negative in normal prostate

epithelium (**Figure 4.7**). The staining pattern observed for Nidogen-1 (NID1) was non-specific both for normal and cancerous prostate tissues.

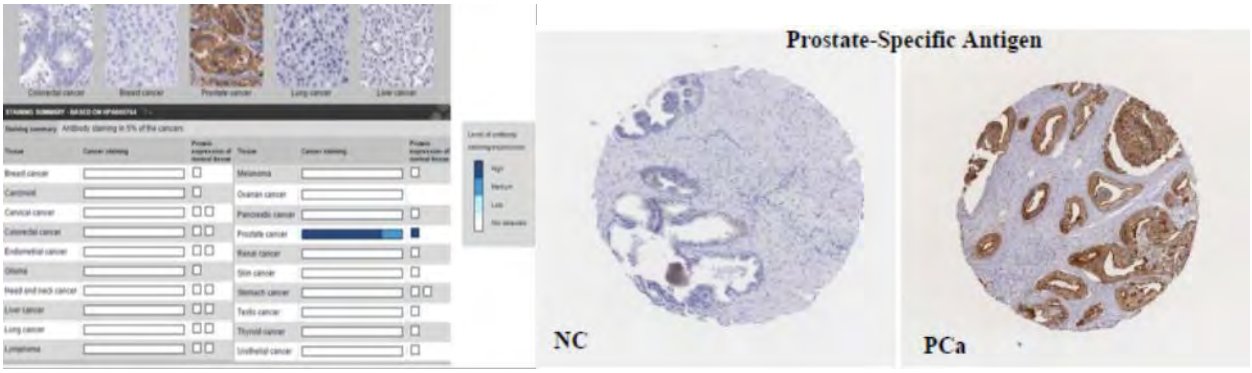


Figure 4.7. Differential staining pattern for prostatic-specific antigen between prostate cancer and normal prostate tissues

Although higher levels of luminal and stromal staining was observed for normal as compared with cancerous prostate tissues. Cathepsin Z (CTSZ) stained positive in PCa but negative for normal prostatic tissues. Besides MYOC, one other race-based potential biomarker with good signal in our PRM was Slain motif-containing protein 1 (SLAIN1) which stained remarkably for PCa tissue but stained mildly for stromal tissues in normal prostate tissue.

4.3.3. Evaluation of Biomarkers using SRMAtlas Database

Selected 32 top ranking potential prostate biomarkers were carefully searched with the SRMAtlas. The Best out of 3 proteotypic peptide for each of the top ranking potential biomarker was verified in the SRMAtlas database. This database automatically ranks the best *b*- and *y*- ion transitions making it easy to identify the "best flying" transitions to target. Mass detection at Q1 and Q3 were reported. The amino acid following the C-terminus (Fol-AA) and the one before the N-terminus (Pre-AA) were demonstrated. Sequence specific retention times (SSRT) for hydrophobicity and relative intensities (RI) of peaks in the collision induced dissociation (CID) spectra were shown. Also reported in the SRMAtlas database was frequency of peptide mapping (N-Map) to proteins in target proteomes. Weighting of predicted *vs* empirical suitability score results in the reported adjusted suitability score (AdjSS) of transitions. SRMAtlas results for selected proteotypic peptides KLK3, PROS1, ACPP and HPR are demonstrated below (**Table 4.3**).

Haptoglobin-related protein

Protein	External Links	Pre AA	Sequence	Fol AA	Adj SS	Source	Q1_mz	Q1_ chg	Q3_mz	Q3_ chg	Ion	Rank	RI	SSRT	N_map	IonTrap
P00739		R	TEGDGVYTLNDK	K	0.9	IonTrap	656.31	2	753.38	1	y6	1	10000	16.8	2	
P00739		R	TEGDGVYTLNDK	K	0.9	IonTrap	656.31	2	1081.52	1	y10	2	7980	16.8	2	
P00739		R	TEGDGVYTLNDK	K	0.9	IonTrap	656.31	2	590.32	1	y5	3	2802	16.8	2	
P00739		R	TEGDGVYTLNDK	K	0.9	IonTrap	656.31	2	909.47	1	y8	4	2694	16.8	2	
P00739		K	SPVGVQPILNEHTFC[160]VGMSK	Y	0.89	IonTrap	734.03	3	958.97	2	y17	1	10000	37.5	1	
P00739		K	SPVGVQPILNEHTFC[160]VGMSK	Y	0.89	IonTrap	734.03	3	816.9	2	y14	2	8733	37.5	1	
P00739		K	SPVGVQPILNEHTFC[160]VGMSK	Y	0.89	IonTrap	734.03	3	1008.51	2	y18	3	3035	37.5	1	
P00739		K	SPVGVQPILNEHTFC[160]VGMSK	Y	0.89	IonTrap	734.03	3	930.46	2	y16	4	2062	37.5	1	
P00739		K	VVLHPNYHQVDIGLIK	L	0.84	IonTrap	615.68	3	698.88	2	y12	1	10000	35.8	1	
P00739		K	VVLHPNYHQVDIGLIK	L	0.84	IonTrap	615.68	3	823.45	1	b7	2	4901	35.8	1	
P00739		K	VVLHPNYHQVDIGLIK	L	0.84	IonTrap	615.68	3	549.64	3	y14	3	3476	35.8	1	
P00739		K	VVLHPNYHQVDIGLIK	L	0.84	IonTrap	615.68	3	767.41	2	y13	4	2632	35.8	1	
P00739		R	VGYVSGWGQSDNFK	L	0.78	IonTrap	772.36	2	1125.5	1	y10	1	10000	28.2	1	
P00739		R	VGYVSGWGQSDNFK	L	0.78	IonTrap	772.36	2	795.36	1	y7	2	9822	28.2	1	
P00739		R	VGYVSGWGQSDNFK	L	0.78	IonTrap	772.36	2	1038.46	1	y9	3	6394	28.2	1	
P00739		R	VGYVSGWGQSDNFK	L	0.78	IonTrap	772.36	2	981.44	1	y8	4	3403	28.2	1	

P00739		K	VTSIQHWVQK	T	0.78	IonTrap	613.34	2	666.36	1	b6	1	9450	22	1
P00739		K	VTSIQHWVQK	T	0.78	IonTrap	613.34	2	1079.56	1	b9	2	6666	22	1
P00739		K	VTSIQHWVQK	T	0.78	IonTrap	613.34	2	560.32	1	y4	3	6538	22	1
P00739		K	VTSIQHWVQK	T	0.78	IonTrap	613.34	2	951.51	1	b8	4	4972	22	1

Vitamin K-dependent protein S

Protein	External Links	Pre AA	Sequence	Fol AA	Adj SS	Source	Q1_mz	Q1_chg	Q3_mz	Q3_chg	Ion	Rank	RI	SSRT	N_map	IonTrap
P07225		R	QSTNAYPDLR	S	1.83	IonTrap	582.79	2	500.28	1	y4	1	6135	15	1	
P07225		R	QSTNAYPDLR	S	1.83	IonTrap	582.79	2	663.35	1	y5	2	3556	15	1	
P07225		R	QSTNAYPDLR	S	1.83	IonTrap	582.79	2	288.2	1	y2	3	3366	15	1	
P07225		R	QSTNAYPDLR	S	1.83	IonTrap	582.79	2	877.37	1	b8	4	1217	15	1	
P07225		R	SFQTGLFTAAR	Q	1.72	IonTrap	599.81	2	565.31	1	y5	1	5846	27.6	1	
P07225		R	SFQTGLFTAAR	Q	1.72	IonTrap	599.81	2	482.76	2	y9	2	4445	27.6	1	
P07225		R	SFQTGLFTAAR	Q	1.72	IonTrap	599.81	2	836.46	1	y8	3	4000	27.6	1	
P07225		R	SFQTGLFTAAR	Q	1.72	IonTrap	599.81	2	735.42	1	y7	4	3428	27.6	1	
P07225		K	ASFTC[160]TC[160]KPGWQGEK	C	1.63	IonTrap	586.26	3	799.86	2	y13	1	10000	21.3	1	
P07225		K	ASFTC[160]TC[160]KPGWQGEK	C	1.63	IonTrap	586.26	3	533.58	3	y13	2	8925	21.3	1	
P07225		K	ASFTC[160]TC[160]KPGWQGEK	C	1.63	IonTrap	586.26	3	726.32	2	y12	3	6450	21.3	1	
P07225		K	ASFTC[160]TC[160]KPGWQGEK	C	1.63	IonTrap	586.26	3	675.8	2	y11	4	3963	21.3	1	
P07225		R	IQALSLC[160]SDQQSHLEFR	V	1.54	IonTrap	1016.5	2	1044.52	1	y8	1	8391	33.5	1	

P07225		R	IQALSLC[160]SDQQSHLEFR	V	1.54	IonTrap	1016.5	2	788.41	1	y6	2	7575	33.5	1	
P07225		R	IQALSLC[160]SDQQSHLEFR	V	1.54	IonTrap	1016.5	2	1246.58	1	y10	3	7214	33.5	1	
P07225		R	IQALSLC[160]SDQQSHLEFR	V	1.54	IonTrap	1016.5	2	701.37	1	y5	4	4922	33.5	1	
P07225		**K**	**HC[160]LVTVEK**	**G**	**1.51**	**IonTrap**	**985.51**	**1**	**839.41**	**1**	**b7**	**1**	**10000**	**17.8**	**1**	
P07225		**K**	**HC[160]LVTVEK**	**G**	**1.51**	**IonTrap**	**985.51**	**1**	**710.37**	**1**	**b6**	**2**	**9473**	**17.8**	**1**	
P07225		**K**	**HC[160]LVTVEK**	**G**	**1.51**	**IonTrap**	**985.51**	**1**	**510.25**	**1**	**b4**	**3**	**3956**	**17.8**	**1**	
P07225		**K**	**HC[160]LVTVEK**	**G**	**1.51**	**IonTrap**	**985.51**	**1**	**611.3**	**1**	**b5**	**4**	**2611**	**17.8**	**1**	

Prostatic acid phosphatase

Protein	External Links	Pre AA	Sequence	Fol AA	Adj SS	Source	Q1_mz	Q1_ chg	Q3_mz	Q3_ chg	Ion	Rank	RI	SSRT	N_map	IonTrap
P15309		K	LSGLHGQDLFGIWSK	V	2	IonTrap	553.3	3	590.33	1	y5	1	10000	42.9	1	
P15309		K	LSGLHGQDLFGIWSK	V	2	IonTrap	553.3	3	772.9	2	y14	2	9519	42.9	1	
P15309		K	LSGLHGQDLFGIWSK	V	2	IonTrap	553.3	3	737.4	1	y6	3	9507	42.9	1	
P15309		K	LSGLHGQDLFGIWSK	V	2	IonTrap	553.3	3	729.38	2	y13	4	7676	42.9	1	
P15309		**R**	**SPIDTFPTDPIK**	**E**	**1.5**	**IonTrap**	**665.85**	**2**	**357.25**	**1**	**y3**	**1**	**10000**	**28.7**	**1**	
P15309		**R**	**SPIDTFPTDPIK**	**E**	**1.5**	**IonTrap**	**665.85**	**2**	**1033.52**	**1**	**y9**	**2**	**9823**	**28.7**	**1**	
P15309		**R**	**SPIDTFPTDPIK**	**E**	**1.5**	**IonTrap**	**665.85**	**2**	**817.45**	**1**	**y7**	**3**	**6686**	**28.7**	**1**	
P15309		**R**	**SPIDTFPTDPIK**	**E**	**1.5**	**IonTrap**	**665.85**	**2**	**918.49**	**1**	**y8**	**4**	**5844**	**28.7**	**1**	
P15309		K	GEYFVEMYYR	N	1.44	IonTrap	678.8	2	761.33	1	y5	1	10000	33.9	1	
P15309		K	GEYFVEMYYR	N	1.44	IonTrap	678.8	2	860.4	1	y6	2	5715	33.9	1	
P15309		K	GEYFVEMYYR	N	1.44	IonTrap	678.8	2	632.29	1	y4	3	4274	33.9	1	
P15309		K	GEYFVEMYYR	N	1.44	IonTrap	678.8	2	501.25	1	y3	4	4106	33.9	1	

P15309		R	LQGGVLVNEILNHMK	R	1.09	IonTrap	832.96	2	998.51	1	y8	1	10000	35.7	1	
P15309		R	LQGGVLVNEILNHMK	R	1.09	IonTrap	832.96	2	1210.66	1	y10	2	8080	35.7	1	
P15309		R	LQGGVLVNEILNHMK	R	1.09	IonTrap	832.96	2	712.39	2	y13	3	7855	35.7	1	
P15309		R	LQGGVLVNEILNHMK	R	1.09	IonTrap	832.96	2	1097.58	1	y9	4	6489	35.7	1	
P15309		R	ELSELSLLSLYGIHK	Q	0.45	IonTrap	851.48	2	817.46	1	y7	1	10000	48.3	1	
P15309		R	ELSELSLLSLYGIHK	Q	0.45	IonTrap	851.48	2	930.54	1	y8	2	6059	48.3	1	
P15309		R	ELSELSLLSLYGIHK	Q	0.45	IonTrap	851.48	2	1130.66	1	y10	3	5954	48.3	1	
P15309		R	ELSELSLLSLYGIHK	Q	0.45	IonTrap	851.48	2	730.43	1	y6	4	4466	48.3	1	

Prostate-specific antigen

Protein	External Links	Pre AA	Sequence	Fol AA	Adj SS	Source	Q1_mz	Q1_chg	Q3_mz	Q3_chg	Ion	Rank	RI	SSRT	N_map	IonTrap
P07288		**R**	**LSEPAELTDAVK**	**V**	**9.24**	**IonTrap**	**636.84**	**2**	**943.51**	**1**	**y9**	**1**	**10000**	**23.6**	**1**	
P07288		**R**	**LSEPAELTDAVK**	**V**	**9.24**	**IonTrap**	**636.84**	**2**	**472.26**	**2**	**y9**	**2**	**2013**	**23.6**	**1**	
P07288		**R**	**LSEPAELTDAVK**	**V**	**9.24**	**IonTrap**	**636.84**	**2**	**533.29**	**1**	**y5**	**3**	**1223**	**23.6**	**1**	
P07288		**R**	**LSEPAELTDAVK**	**V**	**9.24**	**IonTrap**	**636.84**	**2**	**536.78**	**2**	**y10**	**4**	**1143**	**23.6**	**1**	
P07288		K	FMLC[160]AGR	W	1.49	IonTrap	427.7	2	576.29	1	y5	1	10000	21.1	1	
P07288		K	FMLC[160]AGR	W	1.49	IonTrap	427.7	2	463.21	1	y4	2	2287	21.1	1	

P07288	K	FMLC[160]AGR	W	1.49	IonTrap	427.7	2	303.18	1	y3	3	556	21.1	1
P07288	K	FMLC[160]AGR	W	1.49	IonTrap	427.7	2	392.2	1	b3	4	508	21.1	1
P07288	R	AVC[160]GGVLVHPQWVLTA AHC[160]IR	N	1.4	IonTrap	782.08	3	1007.54	2	y18	1	7574	40.7	1
P07288	R	AVC[160]GGVLVHPQWVLTA AHC[160]IR	N	1.4	IonTrap	782.08	3	727.37	1	y6	2	4732	40.7	1
P07288	R	AVC[160]GGVLVHPQWVLTA AHC[160]IR	N	1.4	IonTrap	782.08	3	900.99	2	y15	3	2777	40.7	1
P07288	R	AVC[160]GGVLVHPQWVLTA AHC[160]IR	N	1.4	IonTrap	782.08	3	794.91	2	y13	4	2193	40.7	1
P07288	K	LQC[160]VDLHVISNDVC[160] AQVHPQK	V	1.21	IonTrap	820.74	3	1030.02	2	y18	1	2289	32	1
P07288	K	LQC[160]VDLHVISNDVC[160] AQVHPQK	V	1.21	IonTrap	820.74	3	980.49	2	y17	2	2270	32	1
P07288	K	LQC[160]VDLHVISNDVC[160] AQVHPQK	V	1.21	IonTrap	820.74	3	967.48	1	y8	3	1987	32	1
P07288	K	LQC[160]VDLHVISNDVC[160] AQVHPQK	V	1.21	IonTrap	820.74	3	866.42	1	b7	4	1600	32	1
P07288	R	FLRPGDDSSHDLMLLR	L	1.08	IonTrap	468.74	4	537.94	3	y14	1	10000	35	1
P07288	R	FLRPGDDSSHDLMLLR	L	1.08	IonTrap	936.48	2	686.36	1	b6	2	10000	35	1
P07288	R	FLRPGDDSSHDLMLLR	L	1.08	IonTrap	936.48	2	1227.54	1	b11	3	9114	35	1
P07288	R	FLRPGDDSSHDLMLLR	L	1.08	IonTrap	936.48	2	1186.59	1	y10	4	8425	35	1

Table 4.3. SRMAtlas result for PROS, KLK3, ACPP and HPR showing verified proteotypic peptides for each biomarker in red highlight

4.3.4. Parallel Reaction Monitoring Prevalidation

Preliminary targeted PRM experiments were performed to potentially corroborate the in-silico biomarker screening results and identify high confidence potential biomarkers before proceeding to large scale, multi-platform hypothesis testing using targeted proteomics experiment. Although ELISA is widely acclaimed to be the gold standard for protein quantification, its use can sometimes be limited by the availability of high quality antibodies. Targeted proteomics has been shown to demonstrate performances similar to ELISA assays [403-405]. The focus for parallel reaction monitoring was on the high ranking proteotypic peptides detected from biologic screening and in-silico verification of potential biomarkers. Transitions from this experiment were identified with good quantifiable peak area under curve (AUC) (**Figure 4.8**).

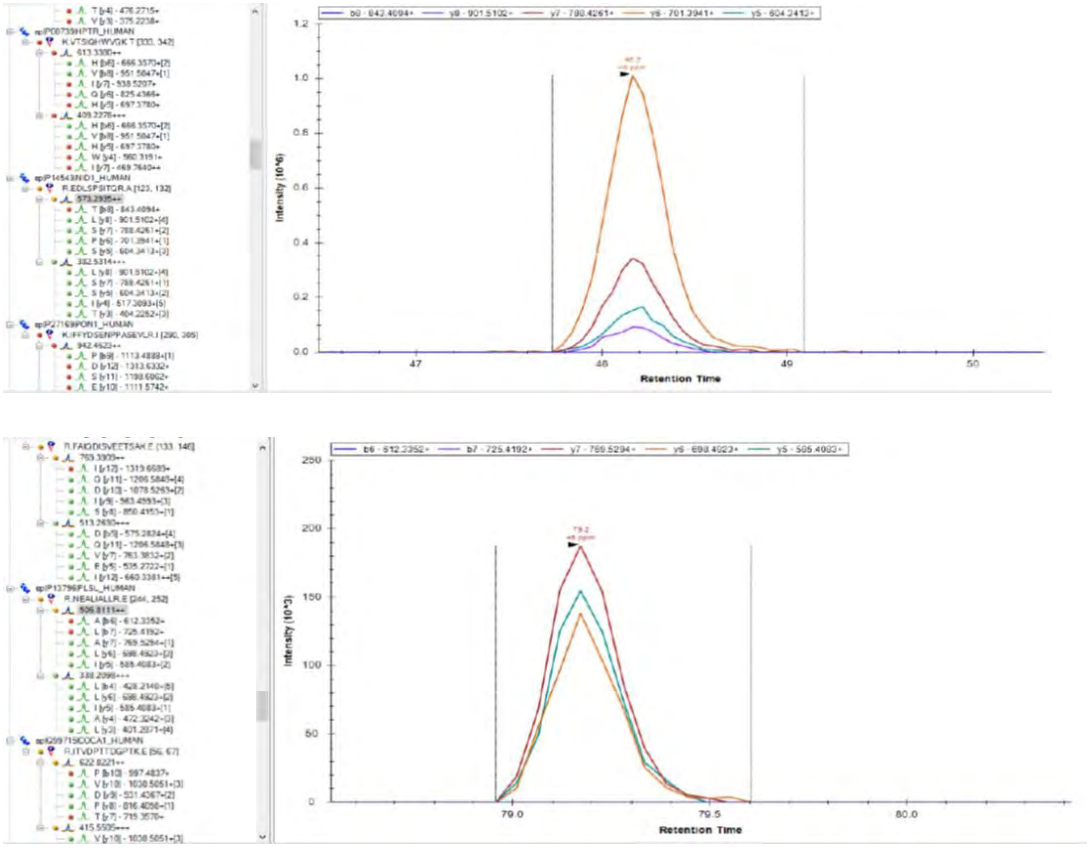


Figure 4.8. Quantitation of target transition of potential biomarkers by parallel reaction monitoring. All peaks within the vertical bracket represent the multiple transitions that were used to target a peptide.

Differential expression was observed for *N*-acetylmuramoyl-*L*-alanine amidase (PGLYRP2) showing a higher expression in PCa as compared to normal controls. This expression difference is more remarkable in the +2 charge state as compared to the +3 charge state. Replicate comparison of peak AUC showed difference majorly in *y*-type ions for +3 charge states while a single *b*-type ion expressed difference in the 2+ charge state. There was greater than 2 fold of difference in peptide transitions between PCa and normal controls for Vitamin K-dependent protein S (PROS1). Transition ions of the *b*-type were observed to be higher for both +2 and +3 charge states in PCa in comparison to normal controls. The same trend was found for Haptoglobin related protein (HPR) which had more than a 3-fold difference in AUC between normal control and PCa regardless of charge states with *b*-ions expressed differentially for both +2 and +3 charges. Carboxypeptidase N catalytic chain (CPN1) transitions were differentially expressed with +2 peptide transitions being lower in normal control and higher in PCa. However, a greater AUC was found for CPN1 in normal controls compared to prostate cancer +3 charge states with *y*-ion dominance. Pregnancy Zone Proteins (PZP) demonstrated greater AUC signals in the normal controls as compared with PCa for both +2 and +3 charges; and were predominantly *y*-type ions. In normal controls, Alpha-actinin-1 (ACTN1) was more highly expressed in comparison to PCa in the +3 charge state. Five *y*-ions fragments were detected in the +2 charge state for NC which was not detected at all in the +3 charge. Similarly, detected transitions for Myocilin (MYOC) in normal controls were higher than those found in PCa; with mostly *b*-type and *y*-type ions being responsible for the difference in the +2 and +3 charge states respectively. Better signals were found for prostatic acid phosphatase (ACPP) in normal controls as compared to PCa, with *y*-ion fragment predominance. Prostate-specific antigen (KLK3) showed higher transition signals for PCa in comparison to normal controls, albeit the difference was not great. Replicate comparisons of peak AUC showed differential signals with mixed *b*- and *y*-ion fragments. The performance of Nidogen-1 (NID1) in parallel reaction monitoring assay was spectacular, demonstrating over 6-fold difference between normal control (higher) and PCa transitions for both +2 and +3 charge states. Replicate comparison for NID-1 showed y-ions more differentially detected. Cathepsin Z (CTSZ) had higher in transition signal for normal

control as compared to PCa for both +2 and +3 charge states. No transitions were detected for CTSZ in the +3 charge states while *y*-ions dominated both charge states. The race-based biomarker slain motif-containing protein 1 (SLAIN1) showed significant differential expression between PCa and normal controls. For SLAIN1, PCa had over 2-fold difference in transition signals in comparison to normal controls.

4.3.5. Confirmation of Candidate Biomarkers in Databases

The top 12 potential biomarkers were re-assessed retrospectively in the data from the previous shotgun urinary proteomics experiments that was performed. Frequency of occurrence of the top ranking biomarkers was assessed in this database as well as 14 other publicly available urinary proteomics databases. In the previous shotgun urinary proteomics data from our PCa cohort, it was observed that HPR, PROS1, SLAIN1 and PZP were reliable biomarkers of PCa (**Figure 4.9**). PZP, PROS1, and HPR were observed only in PCa but found absent in BPH and normal control. Greater than 2-fold of difference in expression was found for SLAIN1 in PCa as compared with normal control and BPH. NID1, CTSZ, ACPP and KLK3 were also moderately good biomarkers of PCa because of their higher expression in normal controls compared with BPH and PCa. NID1 was highly observed in normal controls, slightly present in BPH, but not found in PCa.

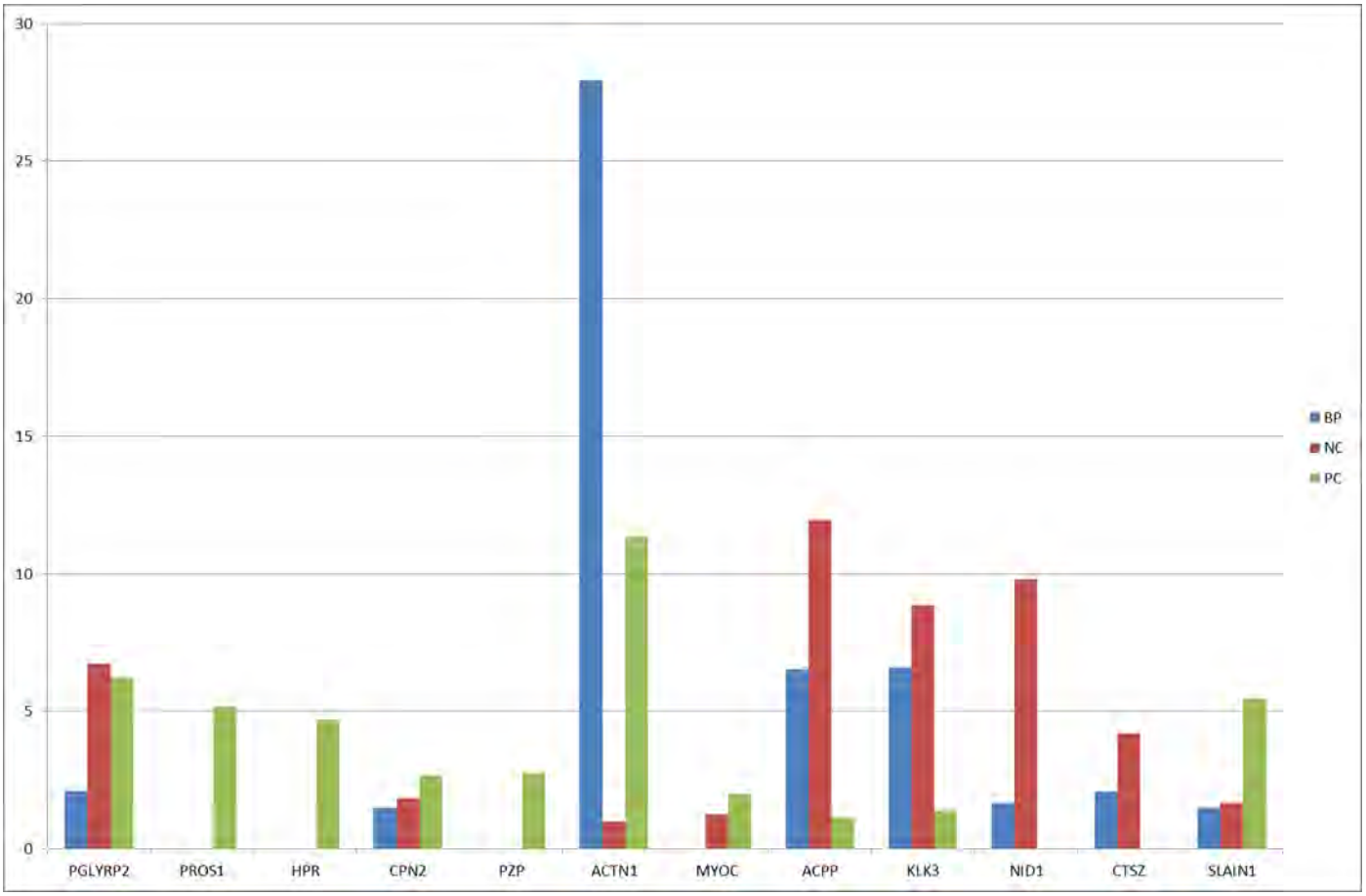


Figure 4.9. Confirmation of top ranking potential biomarkers in shotgun urinary proteomics data. On the x-axis are the top ranking potential biomarkers and on the y-axis are the log2-transformed IBAQ values

KLK3 and ACPP were both highly observed in BPH and normal controls as compared with PCa. CTSZ was observed as highest in NC, slightly present in BPH but completely absent in PCa. All other top ranking biomarkers queried did not demonstrate a clear-cut pattern in their frequency of occurrence. They either demonstrated similar pattern of expression across PCa, BPH and normal control or similar expression in normal control and PCa. These top ranking 12 potential biomarkers of PCa were further queried against 14 other publicly available urinary proteomics databases to assess the degree of concordance with our urinary proteomics database. Six PCa, 2 BPH, 6 healthy (NC), and 1 multiple condition (MC) urinary databases were used (**Table 4.4**). PGLYRP2 was absent in PCa, MC and BPH databases but was present in 5 (83.3%) of the NC databases. PROS1 was found in 4 (66.7%) of the NC databases, but found absent in all BPH, MC and PCa databases. Three (50%) of the PCa databases demonstrated HPR as well as one (16.7%) of the NC databases. MC and BPH databases were not found to demonstrate the presence of

HPR. CPN1 was observed to be present in 5 (83.3%) of the NC and 2 (33.3%) of the PCa databases, whilst absent in MC and BPH databases. PZP was noticed to be absent in BPH, NC and PCa databases, albeit found in the MC database. ACTN1 was observed in 3 (50%) of the NC databases and absent in the BPH, MC and PCa databases. MYOC was found in 2 (33.3%) of the NC and not present in MC, BPH and PCa databases. ACPP was observed in 5 (83.3%) of the NC and 3 (50%) of the PCa databases; it was also present in the MC and BPH databases. KLK3 was observed in 5 (87.3%) of the NC and 1 (16.7%) of the PCa databases; while absent in MC databases but present in BPH. NID1 was found in 6 (100%) of the NC databases and absent (0%) in PCa, MC and BPH databases. CTSZ was found in 4 (66.7%) of the NC and 1 (16.7%) of the PCa databases; whilst absent for both BPH and MC databases. None of the 14 databases contained SLAIN1.

Databases	Condition	Sample	PGLYRP2	PROS1	HRP	CPN1	PZP	ACTN1	MYOC	ACPP	KLK3	NID1	CTSZ	SLAIN1
Youhe Gao et al 2011	Multiple (MC)	Urine	(-)	(-)	(-)	(-)	(+)	(-)	(-)	(+)	(-)	(-)	(-)	(-)
Hong-Lin Cheng et al 2012	Benign Prostate (BPH)	Urine	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(+)	(+)	(-)	(-)	(-)
Ihor Batruch et al 2011	Healthy (NC)	Urine	(+)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(+)	(-)	(-)
Xuejiao Liu et al 2012	Healthy (NC)	Urine	(+)	(+)	(-)	(+)	(-)	(-)	(-)	(+)	(+)	(+)	(+)	(-)
Qing-Run Li et al 2010	Healthy (NC)	Urine	(+)	(+)	(-)	(+)	(-)	(-)	(-)	(+)	(+)	(+)	(-)	(-)
HUPO Max Plank 2014	Healthy (NC)	Urine	(-)	(+)	(-)	(+)	(-)	(+)	(+)	(+)	(+)	(+)	(+)	(-)
Jun Adachi et al 2006	Healthy (NC)	Urine	(+)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(+)	(+)	(+)	(-)
Akilesh Pandey et al 2011	Healthy (NC)	Urine	(+)	(-)	(+)	(+)	(-)	(+)	(+)	(+)	(+)	(+)	(+)	(-)
Dan Theodorescu et al 2008	PCa *vs* Normal (PCa)	Urine	(-)	(-)	(+)	(+)	(-)	(-)	(-)	(+)	(+)	(-)	(-)	(-)
Katarina Davalieva et al 2015	PCa vs Normal (PCa)	Urine	(-)	(-)	(+)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
Cordelia Geisler et al 2015	PCa vs Normal (PCa)	Urine	(-)	(-)	(-)	(+)	(-)	(-)	(-)	(+)	(-)	(-)	(+)	(-)
Chunhui Li et al 2015	PCa vs Normal (PCa)	Urine	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
Andrew Percy et al 2015	PCa vs Normal (PCa)	Urine	(-)	(+)	(+)	(-)	(-)	(-)	(-)	(+)	(-)	(-)	(-)	(-)
Tujin Shi et al 2014	PCa vs Normal (PCa)	Urine	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)

Table 4.4. Confirmation of 12 top ranking potential biomarkers in 14 publicly available urinary proteomics databases [402]

4.3.6. Functional network association

Functional genetic pathway analysis was performed using GENEMANIA for the 12 top ranking biomarkers in addition to PTEN and ETS genes which are well established to be involved in PCa pathogenesis. A network with a predicted coexpression of 79.19%, physical interaction of 17.26%, co-localization of 2.73% and shared protein domain of 0.83% was generated (**Figure 4.10**).

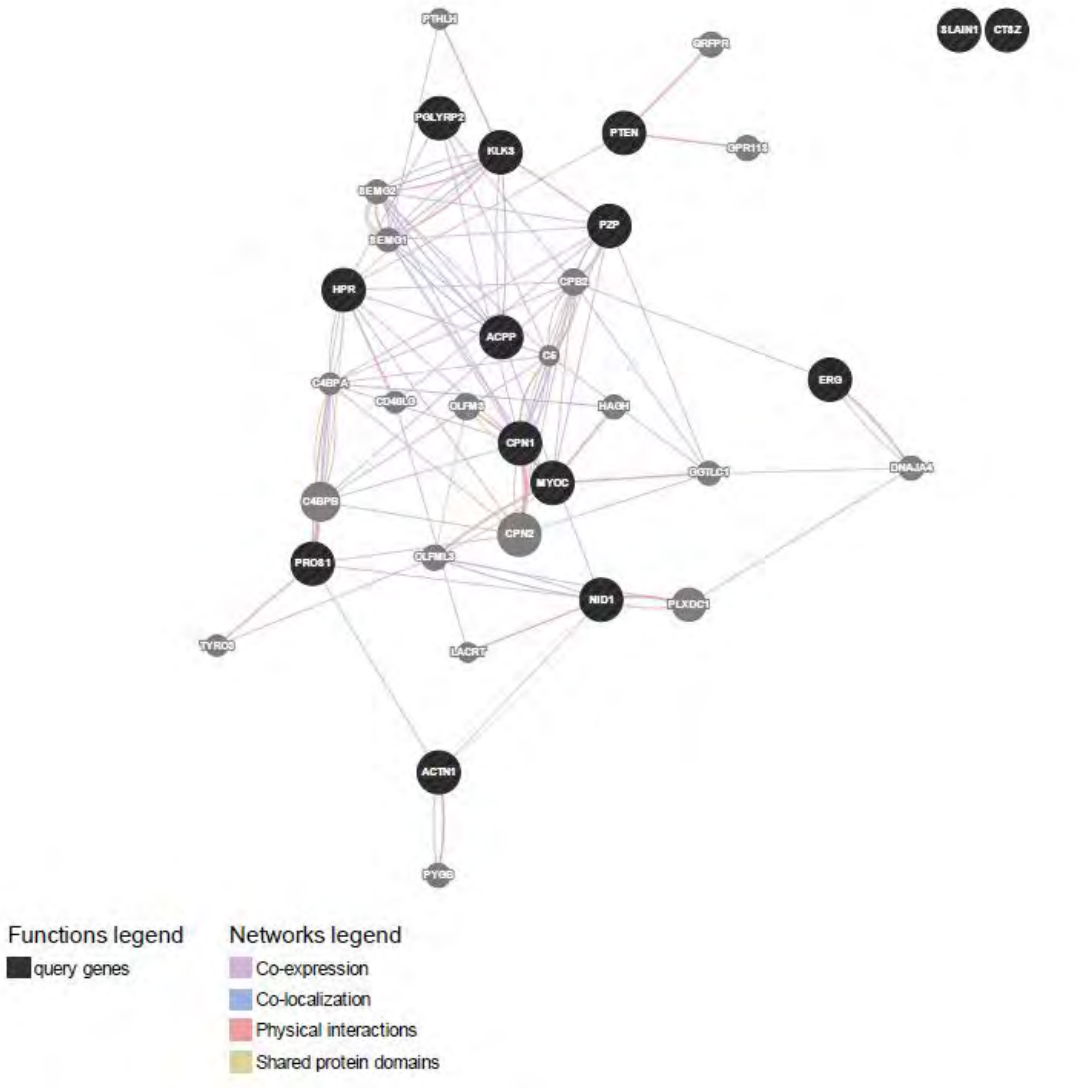


Figure 4.10. Functional network association between 12 top ranking potential biomarkers, PTEN and ETS

4.4. Discussion

Targeted validation of potential biomarkers generated from hypothesis generating shotgun proteomics experiments needs rigorous assay optimization and careful planning [314, 375, 406]. Unfortunately, the expanding number of discovered candidate biomarkers for many diseases, does not tally with clinically applicable or approved disease biomarker**s** [406, 407]. Some reports have suggested the establishment of standards for biomarker discovery and validation of candidate biomarkers for clinical use [116, 375, 408]. An essential requirement of the validation phase of the biomarker pipeline is the availability of numerous clinical samples that are randomized [406]. An orthogonal approach was used in this study to screen 82 potential biomarkers that were previously discovered for prostate cancer in our South African cohort [319]. Database searching using in-silico approaches is rapidly becoming an crucial adjunct to high-throughput systems biology [409-411], although there are some limitations to its use [412]. Over half (51.5%) of our selected proteotypic peptides were shown to have been previously assayed by targeted SRM on an ion trap instrument. This coverage indicates that the selected PTP were reliable. Another subset (47.7%) has not been reported previously in targeted proteomics assays and can be added to this database. Close to 80% coverage has been previously reported when comparing experimental transition identification to in-silico identification [234], indicating the prospects in-silico method application in the biomarker pipeline.

Finding differentially expressed urinary biomarkers in the Human Protein Atlas database confirms their tissue specificity. Due to its vast coverage (11,200 unique proteins) of differential immunoproteomics profile of normal *vs* cancerous tissue, as well as its tissue-specific expression profiles; the human proteomics atlas has been branded as a vital tool for histopathologic assessment [413] and biomarkers discovery [414]. This database has contributed huge proteogenomic information to the development of the Human Proteome Project [415, 416]. An important caution for immunoproteomics evaluation of potential biomarkers is avidity and sequence specificity of the antibody when targeting a specific variant; this may lead to non-specific binding. Another down side to antibody proteomics is the fact that it is may be dependent on commercial availability of the specific antibody; many of which are not

variant/isoform specific [417, 418]. However, we found the use of this database quite helpful in sorting out biomarkers of PCa in our cohort.

Considering the fact that all product ions are concurrently available with minimum interference during the full scan duration, PRM is reliable for multiplexed targeted proteomics even without foreknowledge or prior selection of target transition. This drastically reduces optimization time and effort for every single target transition, albeit poor transition signals may be caused by the stochastic nature of ion selection in PRM. Conversely, a duty cycle of nearly 100% as seen in the QQQ instrument which uses electron multiplier-based detection offers a better sensitivity compared to Orbitrap instrument image current-based detection [419]. This brings to consideration, the inherent tradeoff between the sensitivity and high mass accuracy between PRM and SRM experiments [238, 401, 419]. PRMs are typically used in the screening/survey mode of targeted proteomics experiments, whilst SRM may be more suited for absolute/ precise quantitation of analytes across samples [419]. In this study, we used PRM as a screening tool for the 32 top ranking potential biomarkers generated by in-silico database screening.

The frequency of detection of the 12 top ranking potential biomarkers of prostate cancer was confirmed retrospectively and differential expression was observed in the previous urinary shotgun proteomics experiments. Putative biomarkers of prostate cancer like KLK3 and ACPP were found in PCa, normal control and BPH urinary databases. This indicates that these putative biomarkers are specific for prostate disease and not exclusive to PCa. SLAIN1, a Race-based potential biomarker was more observed in PCa compared to NC and BPH, although no information was found for it in the SRMAtlas. This may be accounted for by the paucity of targeted proteomics data emanating from African PCa studies. Good biomarkers observed by database reconfirmation were PROS1, HRP, SLAIN1, PZP, and NID1. Notably, NID1 was observed only in the healthy urine samples; indicating a fairly constant performance across both in-silico and experimental biomarker analyses. The high level of co-expression (79.19%) as well as physical interaction (17.26%) of shared protein domains found in the network association and pathway analysis of the 12 top ranking biomarkers with PTEN and ETS pathways requires further study. Using a pragmatic penultimate screening approach prior to definitive large scale targeted proteomic validation, this study identified a panel of 12 top ranking potential biomarkers of PCa in a South African cohort of patients.

CHAPTER 5: CANCER ANTIGEN MICROARRAY -BASED DISCOVERY OF NOVEL POTENTIAL BIOMARKERS OF PROSTATE CANCER

5.1. Background

Among cancers in Africa, PCa has the highest incidence, mortality and 5-year prevalence as reported by the WHO/IARC GLOBOCAN 2012 database [32]. The basis for PCa disparity in Africa is poorly understood, although higher levels of serum testosterone [176, 420], androgen receptor variations [176, 421, 422], dietary factors [423, 424], familial [193] and genetic mutations [193, 425-427] have been previously suggested. More reliable novel biomarkers of PCa are needed in Africa considering that PSA as a blood-based biomarker is neither able to differentiate benign from malignant prostate disease [428] nor distinguish aggressive from indolent phenotypes of the disease [98, 429]. Another blood-based biomarker, prostatic acid phosphatase (ACPP), is poorly sensitive for localized PCa [71, 430]. In addition, ACPP can be found in extraprostatic tissue as well, making it a poor clinical biomarker for PCa [72, 431]. Derivatives of prostate-specific antigen such as free-to-bound PSA ratio [432], velocity, density, doubling time, proPSA, kallikrein-marker panels and Prostate Health Index (PHI) have been observed to improve the prognostic and diagnostic value of PSA [432, 433]. In addition, TMPRSS2:ERG translocation has been shown to be prognostic marker for PCa aggression in Western studies [432]. Another blood-based predictor that has been used to predict prostate cancer distant metastasis is circulating tumour cells (CTCs) [74]. Urinary levels of TMPRSS2:ERG fusion, PCA3 are also some of the emerging biomarkers for stratification of PCa risk [74, 432]. Clearly, a combination of good clinical assessment and reliable PCa biomarkers would offer more benefit in the diagnosis and management of PCa. For example, PHI and PSA isoforms have been combined with the use of multiparametric-MRI targeted prostate biopsy; and has improved the stratification of PCa risk in comparison to urinary biomarkers or PSA alone [434].

An emerging miniaturized technology for multiplexed high-throughput discovery of molecular pathways, cancer biomarkers, and immunotherapeutic targets is the protein microarray technology [435]. Seeing that a vital that hallmark of natural immunity and cancer is the

production of autoantibodies [436]. Various forms of protein microarrays exist, including bead based arrays, antibody arrays, reverse phase protein arrays, microspot ELISA array, and pathway arrays systems. Various platforms exist for the formulation of protein microarrays such as glass slides, membranes, and beaded platforms [435]. This technology is potentially capable of detecting variations in antibody response to immunization and can identify target antibodies in large cohorts of individuals [437]. The fact that a target subset of the proteome can be functionally interrogated in a systems-oriented manner, the consistency of the technical layout of the array and economy of ligand or reaction quantity permits meaningful comparison and quantitation of protein microarray results across huge datasets [438]. Many authors have shown that antibody-antigen microarray platforms used is beneficial for early diagnosis of cancer and discovery of biomarkers [257, 439-443].

Cancer-Testis antigens (CTAs) are a heterogeneous family of tumour associated antigens (TAAs) that are solely found naturally in adult ovarian and testicular germ cells as well as placental embryonic/trophoblast membranes; and pathologically in cancers [277, 444-446]. The fact that these antigens are only found in cancer or gonadal germline tissues makes them an attractive target for immunotherapy and diagnosis of cancer. In-depth understanding of these tumour specific and very immunogenic antigens is currently evolving, although over a hundred of CTAs have been described in scientific literature [446, 447]. Cancer-testis antigens have been reported previously with variable expression in melanomas, cancers involving the lung, prostate, breast, bladder, colon, kidneys; as well as malignancies involving the hematologic and lymphoreticular systems [447]. The theranostic possibilities of cancer-testis antigen for urological malignancies were explored in a review by Kulkarni *et al* [448], who examined the cancer-testis antigens as therapeutic targets and diagnostic biomarkers in an array of urologic malignancies including renal, prostate, testicular germ cell and bladder malignancies.

Analysis of serum for humoral adaptive response to cancer using protein microarray technology is a promising scientific approach to achieve a bidirectional outcome of potential diagnostic biomarker and vaccine target identification. Using protein microarrays to identify cancer-specific humoral immune responses; vaccine targets and potential CTA biomarkers were found for non-small cell lung cancer [449], as well as pancreatic and ovarian cancer [437]. Cancer testis antigen have been previously described as potential biomarkers of aggression [450, 451], progression of

disease [452], staging [453] and biochemical recurrence [454] of PCa. As an emerging technology, protein microarrays are not without their challenges. For instance, pH alteration may render capture molecule unstable, and they may demonstrate variable specificity for their target antigens. These problems can be compounded by dynamic range issues found with the plasma proteome. Additionally, in post-translationally modified protein, non-active conformations of the arrayed proteins may affect exposure of the desired epitope [435]. Furthermore, many of the bioinformatics data analysis software for protein microarrays were adapted from genomic microarray computational workflows, which may not necessary be amenable for variations in individual unique immune response as observed in antigen microarrays [435, 437].

Despite the developments in protein microarray technology, there is a dearth of literature on the cancer-testis antigen expression pattern in any African PCa cohort. In addition, racial disparities in CTA expression in a heterogeneous cohort of PCa patients are under-explored. Hence, a novel blood-based approach to identification of immunotherapeutic target and potential PCa biomarkers using a protein microarray comprised of CTAs and a cocktail of putative TAAs was described in this this study. In addition, racial distribution of TAAs and CTAs were explored in our multi-cultural cohort of South African PCa patients.

5.2. Methodology

5.2.1. Characterization of patient and sample cohort

Blood samples (*N*=67) were collected in our cohort over a period of 2 year period from benign prostatic hyperplasia (BPH) (n=32), PCa patients (PCa) (n=20), and individuals who were symptomatic but were suffering from other genitourinary diseases or screened negative for malignancy or BPH with histopathologic assessment (DC) (n=15). Clinicopathologic and demographic data of all 67 participants were documented (**Table 5.1**). The mean ages for PCa, BPH and DC were 68.95, 67.75, and 58.3 respectively; while the median ages for these groups were 68, 68, and 61 respectively. The mean PSA values for PCa, BPH and DC were 158.61, 8.67, and 4.63 respectively; while the median PSA values for these groups were 27.52, 4.87, and 3.7 respectively. Blood was collected and processed as described in **Sections 2.1** & **2.2** respectively.

Code	Diagnosis	Age	PSA level	Race Group	Gleason Score
PC1	CANCER	61	11.6	MA	6
PC2	CANCER	71	5.1	MA	7
PC3`	CANCER	61	100	B	9
PC4	CANCER	74	31	MA	7
PC5	CANCER	61	16.4	MA	6
PC6	CANCER	67	350	MA	9
PC7	CANCER	80	39.4	B	7
PC8	CANCER	68	1091	MA	9
PC9	CANCER	69	5.9	W	7
PC10	CANCER	64	9.1	W	6
PC11	CANCER	64	1.1	W	6
PC12	CANCER	76	195	W	10
PC13	CANCER	74	1.5	MA	6
PC14	CANCER	66	34	MA	9
PC15	CANCER	76	315	MA	6
PC16	CANCER	62	17.05	MA	7

PC17	CANCER	63	24.05	MA	7
PC18	CANCER	71	2.9	W	9
PC19	CANCER	74	184	W	10
PC20	CANCER	77	738	B	8
BPH1	BENIGN	72	3.6	MA	NA
BPH2	BENIGN	64	2.8	MA	NA
BPH3	BENIGN	70	5.4	W	NA
BPH4	BENIGN	61	2.9	MA	NA
BPH5	BENIGN	75	3.7	MA	NA
BPH6	BENIGN	58	19.5	W	NA
BPH7	BENIGN	70	48.4	MA	NA
BPH8	BENIGN	68	5.7	MA	NA
BPH9	BENIGN	69	7.1	MA	NA
BPH10	BENIGN	56	5.03	MA	NA
BPH11	BENIGN	53	1.24	W	NA
BPH12	BENIGN	63	4.7	W	NA
BPH13	BENIGN	86	10.6	B	NA
BPH14	BENIGN	56	9.6	MA	NA
BPH15	BENIGN	75	7.3	B	NA
BPH16	BENIGN	61	4.5	B	NA
BPH17	BENIGN	81	25.8	B	NA
BPH18	BENIGN	64	6.6	B	NA
BPH19	BENIGN	74	2.6	B	NA
BPH20	BENIGN	57	3.64	MA	NA
BPH21	BENIGN	69	3.3	MA	NA
BPH22	BENIGN	70	1.2	W	NA
BPH23	BENIGN	75	0.83	W	NA
BPH24	BENIGN	65	27	B	NA
BPH25	BENIGN	68	5.8	MA	NA
BPH26	BENIGN	62	1.4	W	NA
BPH27	BENIGN	78	0.1	B	NA
BPH28	BENIGN	66	3.6	W	NA
BPH29	BENIGN	68	0.75	B	NA
BPH30	BENIGN	70	9.1	MA	NA

BPH31	BENIGN	77	37.1	MA	NA
BPH32	BENIGN	67	6.43	MA	NA
DC1	NFM	44	0.7	MA	NA
DC2	NFM	74	7.8	MA	NA
DC3	NFM	62	3.7	MA	NA
DC4	NFM	52	NA	MA	NA
DC5	NFM	34	NA	B	NA
DC6	NFM	63	6.3	MA	NA
DC7	NFM	67	5.9	MA	NA
DC8	NFM	65	3.4	MA	NA
DC9	NFM	59	18.9	MA	NA
DC10	NFM	69	0.4	MA	NA
DC11	NFM	61	8	MA	NA
DC12	NFM	76	7	B	NA
DC13	NFM	56	0.9	MA	NA
DC14	NFM	35	NA	B	NA
DC15	NFM	58	6.5	MA	NA

PSA= Prostate-specific antigen; MA= Mixed Ancestry; W= White; B= Black; NFM= Negative for malignancy

Table 5.1. Clinicopathologic information on all participants

5.2.2. Nexterion H-slide microarray derivatization

The specifications of the Nexterion® Slide H (SCHOTT GmBH, Jena, Germany) used as well as its streptavidin derivatization protocol is described in details in **Section 2.4.1**. Derivatized slides were stored at -20° C until use.

5.2.3. Quality check for streptavidin-derivatized slides

Quality check (QC) was carried out to make sure that streptavidin is homogeneously spread on the Nexterion H-slides. For the QC assessment, the last slide of every batch derivatised was incubated with 10 µg/ml Cy5-biotinylated BSA in PBS, washed in washing buffer and then

scanned using the Tecan LS Reloaded™ microarray scanner (Tecan Group Ltd, Männedorf, Germany). Only slides with CV ≤ 5% across the slide surface were used for the assay.

5.2.4. Fabrication of CT100+ Antigen Microarray

Genetix QArray2 robotic arrayer (Genetix Ltd., New Milton, UK), a high throughput microarray printer was used to print antigen-containing crude insect lysates on the streptavidin-coated Nexterion H-slides as described in **Section 2.4.2**. Heterologous expression with *Escherichia coli* (E.coli) has been generally fraught with expression problems like defects in folding, absence of posttranslational modifications (PTMs), and protein solubility [278]. Therefore, *Spodoptera frugiperda* SF21 insect cells were used to express the human antigens of interest because this method presents a simple eukaryotic-like expression system that preserves PTM and native protein folding better [438]. Full details of the optimized insect expression system and lysate preparation have been well documented in previous works [277, 278, 438]. The physical origin of ligand binding assay system is comprised of spatially-defined array of immobilised, purified cancer-testis antigens printed on a surface which permits capture of cognate autoantigens in serum sample. Through fluorescently labelled anti-human IgG, which are captured on the microarray surface, anti-CT antigen expression is then detected. Biotinylated human IgG and biotinylated human serum were used as positive control and a crude cell lysate containing the BCCP-tag alone with no recombinant fusion partner was the negative control.

5.2.5. QC in Microarray Fabrication

Each slide was carefully inspected prior to the printing process. Quality checks were used to confirm that there were no forms of contaminants like fingerprints or dust particles on the slide. The Qarray2 arrayer has a 300 μM solid pins to print replicate CT100 arrays in a 4-plex format. An equal volume of 40 % sucrose was mixed with lysates and spotted in triplicates for each array. The settings for microarray printing are as described in **Section 2.4.2**. After printing, slides were stored in storage buffer (Blocking buffer with 50 % glycerol) at -20 °C until when the CT100+ assays were done.

5.2.6. CT100+ Microarray Assay

In total, 70 serum samples were assayed, 3 of which were used for the purposes of quality control. These quality controls included negative control serum, positive control serum, and a mouse-anti-c-Myc-Cy3 antibody. The positive control which was derived pooled multiple cancer sera showed reactivity to a number of antigens on the array, whereas the negative control which were pooled from normal healthy patients' sera showed no reactivity to antigens on the microarray. The anti-c-myc-Cy3 assay was used to confirm that all the 123 individual antigens (**Table 5.2**) were successfully immobilised to the array during printing with layout shown below (**Figure 5.1**). The steps involved in the CT100+ assay are described in **Section 2.4.3**.

List of 123 antigens printed on the CT100+ antigen microarray platform

123 NY-ESO-1 ORF2	MAGEA3	p53 M133T	THEG
MAGEA4 v2	BAGE2	TPTE	97 p53 L344P
MAGEA4 v3	BAGE3	TSGA10	98 CYP3A4
MAGEA4 v4	BAGE4	TSSK6	99 CYPR
MAGEA5	BAGE5	TYR	100 EGFR
MAGEB1	CCDC33	XAGE-2	101 5T4/TPBG
MAGEB5	CEP290	XAGE3a v1	102 XAGE1B
MAGEB6	COL6A1	XAGE3a v2	103 SOX2
MART-1/MLANA	COX6B2	ZNF165	104 ACVR2A
MICA	CSAG2	AKT1	105 ACVR2B
NLRP4	CT47.11	CDK2	106 ITGB1
NXF2	CT62	CDK4	107 MAP9
NY-CO-45	CTAG2/ NY-ESO-2	CDK7	PIM1
NY-ESO-1	CXorf48.1	FES	TKTL1 Isoform a
OIP5	DDX53	FGFR2	SPATS1 isoform 1
p53	MMA1	MAPK1	DPPA2
PBK	FTHL17	MAPK3	SOX1
RELT	GAGE1	PRKCZ	ROPN1A
ROPN1	GAGE2A	RAF	CEACAM1 isoform 1

SGY-1/	**GAGE4**	**SRC**	**POU5F1**
SILV	**GAGE5**	**CALM1**	**NANOG**
SPAG9	**GAGE6**	**CDC25A**	**BORIS B0**
SPANXA1	**GAGE7**	**CREB1**	**DPPA4**
SPANXB1	**GRWD1**	**CTNNB1**	**DPPA3**
SPANXC	**HORMAD1**	**p53 S6A**	**GDF3**
SPANXD	**LDHC**	**p53 C141Y**	**CTAG2 / LAGE-1b / LAGE-1L**
SPO11	**LEMD1**	**p53 S15A**	**CAMEL**
SSX1	**LIPI**	**p53 T18A**	**p53 S392A**
SSX2a	**MAGEA1**	**p53 Q136X**	**MAGEA2**
SSX4	**MAGEA10**	**p53 S46A**	**SYCP1**
SYCE1	**MAGEA11**	**p53 K382R**	

Table 5.2. List of 123 antigens printed on CT100+ microarray

On each slide, individual arrays were scanned using the Tecan LS Reloaded microarray scanner with 10 μm resolution, and fluorescence was detected. The images were then saved as a 16-bit Tagged image file (TIF) format (**Figure 5.2**). Any sample that shows spot merging of Cy5-biotin-BSA channels was re-assayed afresh.

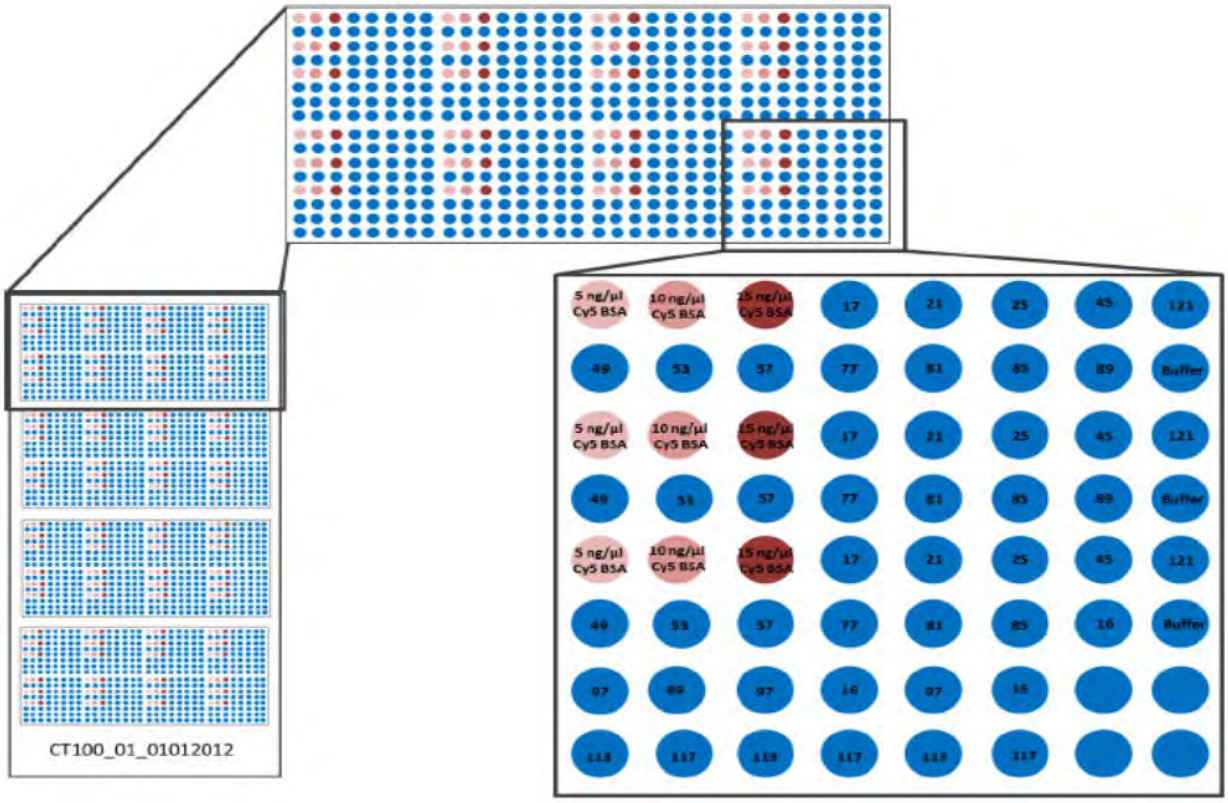


Figure 5.1. Layout of CT100+ microarray with each slide printed in a 4-plex format which is further subdivided into 8 subarrays

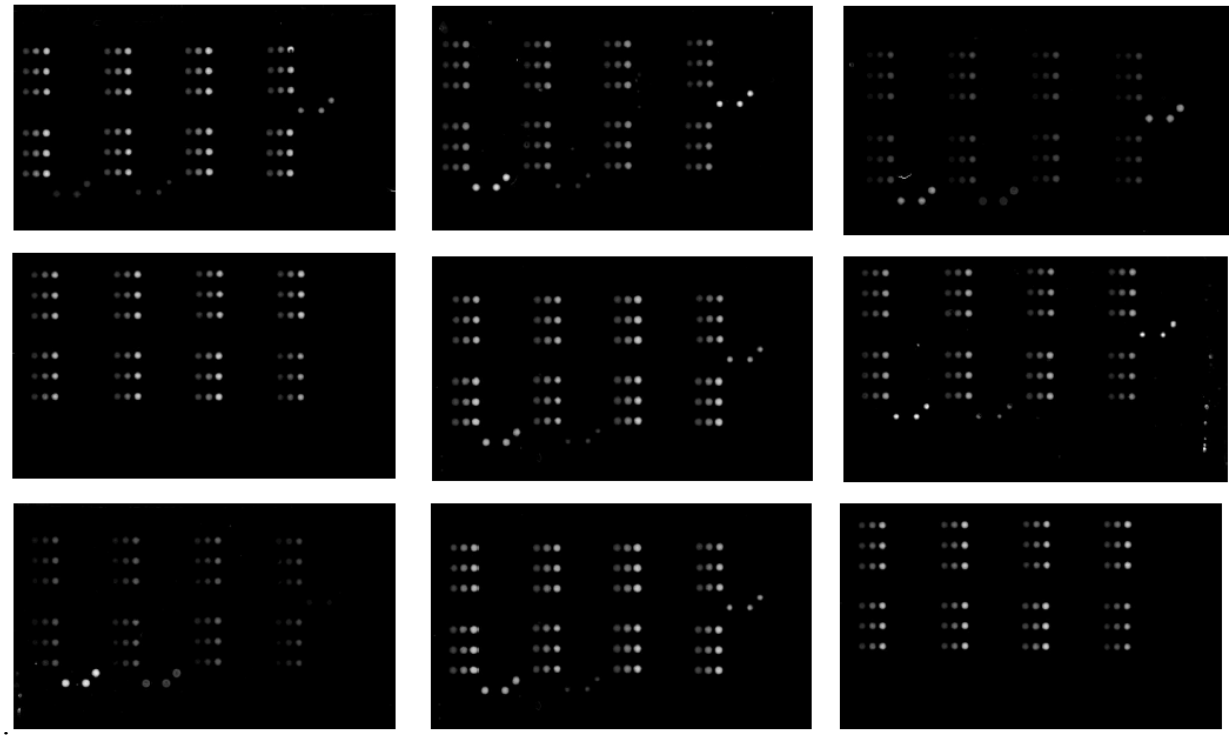


Figure 5.2. Tagged image file format (TIFF) images generated with Tecan LS Reloaded microarray scanner after the CT100+ microarray assay

5.2.7. Data extraction from TIFF images

The TIFF files generated from scanning were inspected visually inspection as a form of preliminary quality assessment. The triplicate BSA spots at 5, 10, and15 μg/ml were confirmed to be visibly present within each subarray of the microarray. Human IgG which detected by fluorescently labelled secondary antibody as well as human anti-IgG which were detected only on the addition of plasma or serum to the slide were employed as positive controls to evaluate image signal intensity. Each microarray was further visually assessed for high background, presence of interfering dust particles, excessive speckling, and evidence of protein spot coalescing; and such slides were excluded. If any of the microarrays demonstrated the aforementioned properties, the assay was repeated form the sera in a fresh slide until the anomaly is corrected. Array-Pro 4.5 (Media Cybernetics, Inc., Maryland, USA) was used for data extraction. A grid with the identity of each antigen and control was positioned such that it encircled each control and antigen on the TIFF image. Further protocol for TIFF image data extraction is documented in **Section 2.4.4**. Once the data was extrapolated by the ArrayPro Analyzer Software version 4.5 (Media Cybernetics, Inc, Maryland, USA), it was filtered and normalized using an in-house developed software (CT100+ programme) with additional data analysis.

5.2.8. CT100+ programme and data normalization

One of the main goals of the CT100+ programme was to resolve some of the bottlenecks of normalization, class comparison and qualitative clustering of protein microarray data. The current programme developed in-house [278] aims to address various parameters as discussed herein. For quality control, technical variations between triplicate signals from spot to spot should be identical. Multiple spots may bleed into each other due to proximity of the spots as well as the use of the wrong buffer at a specific spot. Another issue is that the spotting pin may get stuck during print runs or due to wrong calibration and artefact formation due to inadequate cleaning of print-heads between runs. Homogeneous pixel intensity is expected from individual spot signals across spot surface. Improper handling and storage conditions as well as pin height

errors may result in “doughnut effects”, high intensity spots, uneven spot intensity, and printed spot evaporation. Homogeneity was estimated by measuring the percentage mean intensity and coefficient of variation (CV) within a spot; and spots that fall short of the expected quality are repeated.

5.2.9. Background signal intensity normalization and data filtering

There should be minimal variation in the background signal across array spots. Improper handling, dust particles, poor storage and presence of artefacts can result in high background signal at a print spot. This reduction in signal to noise ratio (SNR) makes it difficult to identify genuine signals (net spot intensity) because there is no distinct difference in the intensity of the high background noise and the actual foreground signal intensity. For QC, only SNR of less than 2 is acceptable for further statistical evaluation. Spots across the array should not contain saturated pixels as this can affect the scanner reading. When there is a faulty array the whole process of sample preparation, microarray printing and CT100+ assay is repeated. If this high signal still persists after repeat assays, automatic gain control (AGC) is used to ensure that the signal lies within the reference range of 200-65,550 RFU. If AGC signal is still too strong after this, the measurement may go into the non-linear range of the platform; hence the antigen is flagged automatically. The same scanner was used for all the slides throughout this study. Further QC was performed to filter out noise and array defects before bioinformatics evaluation. The data quality is improved by filtering because poor quality and disputable spot or arrays can be easily detected. Spots with saturation levels beyond our saturation cutoff were excluded from further analysis as well as triplicate spot intensity with high CV. When this occurs, the spot with the high intensity is excluded and analysis is carried out on the remaining two spots (S1 and S2). The variability is now defined by the equation $(|S\ 1 - S\ 2|)/(S\ 1 + S\ 2)$. Before data analysis, normalization was carried out using a customized composite normalization method for antigen arrays. This method is robust enough to accommodate a relatively smaller number of positive controls to inter-spot and inter-array spot variation. An assumption that there is equal distribution of intensity in all positive controls across array was made. Based on this proviso, a combination of quantile normalization and total intensity normalization was carried out to eliminate systematic bias on the array. In the quantile normalization workflow, the Cy5-labelled positive

controls were assumed to share similar distribution across arrays. Through this distribution, baseline "house-keeping" intensities across arrays were defined. The data was then reorganized to accommodate outlier spots in the positive control dataset. In the total intensity normalization workflow, summation of all the positive controls on each array is presumed to be constant. If a "house-keeping" positive control spot intensity is deemed to be an outlier, this method of normalization ensures that the same spot is regarded as such across all arrays.

5.2.10. Differential and linear statistical analysis

Following the preprocessing statistical step with the CT100+ programme, spot intensities were compared between BPH, PCa, and DC groups and a set of top ranking antigens in each group using linear fold over cut-off differences in relation to the interquartile values per array for each group were identified. For differential expression and multivariate analysis, background-corrected raw intensity data for the 67 samples were then loaded into the Perseus Software (*version 1.4.0.20*) followed by logarithmic normalization, data filtering and further analyses. Differentially expressed antigen between PCa, BPH and DC were identified (FDR = 0.01) using an independent sample t-test with Bonferroni correction for multiple testing. In addition, unsupervised hierarchical clustering was carried out to identify unique antigen signatures for PCa, BPH and DC. Multivariate testing, using principal component analysis (PCA) was also performed to see how the group clustered along each principal component. Unsupervised hierarchical clustering was also performed using *k*-mean clustering on Cluster (*Version 3.0*) software in conjunction with Java TreeView (*version 3.0*) software. Futhermore, the top ranking 50 antigens were assessed with a 3-way Venn diagram plotted using the software known as Venny (as described earlier), a freely available Venn diagram plotting resource. Antigens unique to DC, PCa, and BPH were identified and compared with those identified by differential and linear expression analyses. The top ranking 20 antigens with the highest signal intensities in PCa, BPH and DC were also analysed. Racial variation in PCa antigen expression was examined and high ranking antigen expressions in Mixed Ancestry, Africans, and Caucasians in our PCa cohort were identified. The presence of potential antigen biomarkers of PCa was also confirmed in the previous shotgun urinary PCa proteomic data [319].

5.2.11. Functional pathway analysis

Functional Pathway enrichment analyses of highly expressed antigen by differential expression, linear expression, and Venn diagram analysis were performed using the GeneMANIA software (http://www.genemania.org/), a free online gene interaction pathway analysis tool. Functionally enriched pathways and genes were then further confirmed using the STRING (*version 9.1*) functional protein association network software (freely available at http://string-db.org/).

5.3. Results

5.3.1. Linear analysis for potential biomarkers

Given that most approaches to protein microarray analysis are modified from gene microarrays and considering that statistical methods for protein microarrays are currently evolving; absolute quantification of antigen for standardized comparison between different individuals can be quite challenging. A series of 67 patients' sera were analyzed for autoantibody response to 123 Antigens, composed primarily of a cocktail of CTAs and a few other TAAs. A union of all identified potential biomarkers by linear, Top 20, Venn diagram, and differential expression analysis yielded a total of 41 antigens with differential autoantibody response in our prostate cancer cohort (**Table 5.3**).

S/N	Potential PCa antigen Biomarkers	Analyses	High/ Low autoantibody titre in PCa	Ethnic Distribution
1.	DPPA4	Differential & Venn	High	MA
2.	CEACAM1 Isoform 1	Differential	High	CA
3.	NY-ESO-1	Differential & Top 20	High	MA
4.	FGFR2	Differential, Venn & Shotgun	High	CA
5.	RAF	Differential	High	MA
6.	ZNF165	Differential	High	
7.	TKTL1 (Isoform a)	Differential	High	IA
8.	MAPK3	Differential & Top 20	High	CA*
9.	CAMEL	Differential & Top 20	High	MA
10.	LDHC	Differential , Venn & Top 20	High	
11.	BORIS BO	Venn	High	
12.	SPANXA1	Linear & Venn	High	MA
13.	ROPN1A	Linear & Top 20	High	MA*
14.	p53 S392A	Venn	High	
15.	p53 L344P	Differential	High	IA
16.	p53 C141Y	Top 20	High	MA

17.	p53 K328R	Differential	High	MA
18.	p53 S15A	Linear & Top 20	Low	MA*
19.	p53 T18A	Top 20	High	MA*
20.	CDK2	Differential	High	IA*
21.	MAGEA11	Differential	High	CA
22.	FES	Differential & Top 20	High	
23.	OIP5	Differential & Top 20	High	MA*
24.	SSX2A	Differential	High	
25.	GAGE5	Differential	Low	IA
26.	MAGEB5	Differential & Top 20	Low	CA
27.	EGFR	Differential	Low	
28.	CCDC33	Differential	Low	CA
29.	CSAG2	Differential & Venn	Low	CA
30.	DDX53	Differential & Venn	Low	IA
31.	CT47.11	Differential & Venn	Low	IA
32.	p53	Differential & Venn	Low	
33.	p53 Q136X	Differential & Venn	Low	
34.	MAGEB6	Differential	Low	CA
35.	PBK	Differential	Low	IA*
36.	CAML1	Shotgun	High	IA*
37.	COL6A1	Shotgun	High	IA*
38.	GAGE1	Linear & Top 20	High	MA
39.	PRKCZ	Linear & Top 20	High	MA
40.	p53 S46A	Linear & Top 20	low	MA
41	MAGEB1	Linear & Top 20	Low	

MA= Mixed Ancestry; IA= Indigenous African; CA= Caucasian African; S/N= Serial number; *=Significant

Table 5.3. List of 41 antigens with differential autoantibody response in prostate cancer using linear, differential and Venn diagram analysis.

It was observed that the positive control sera showed reactivity to many of the antigens on the array (**Figure 5.3**), whereas the negative control sera showed no reactivity to antigens on the microarray (**Figure 5.4**).

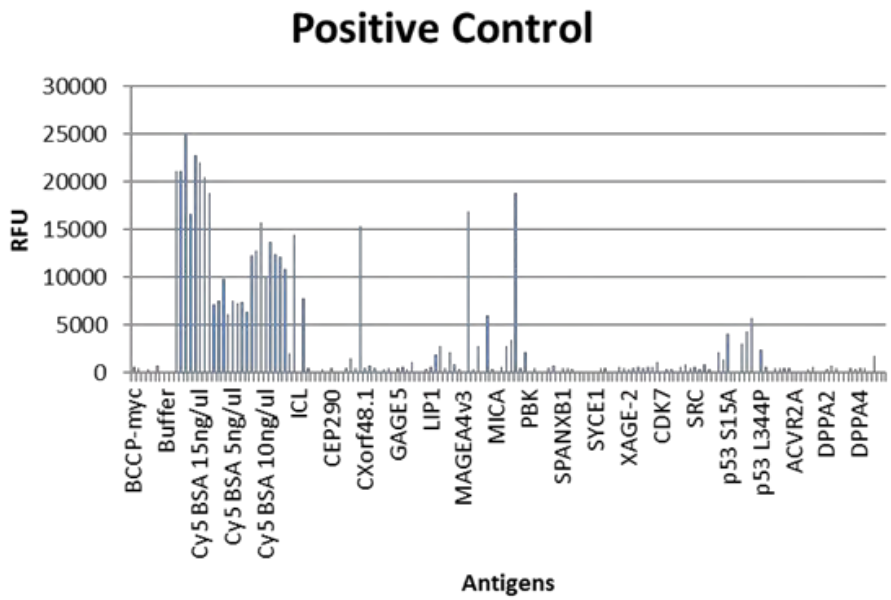


Figure 5.3. Bar graph demonstrating that positive control sera showed reactivity to antigens on the array

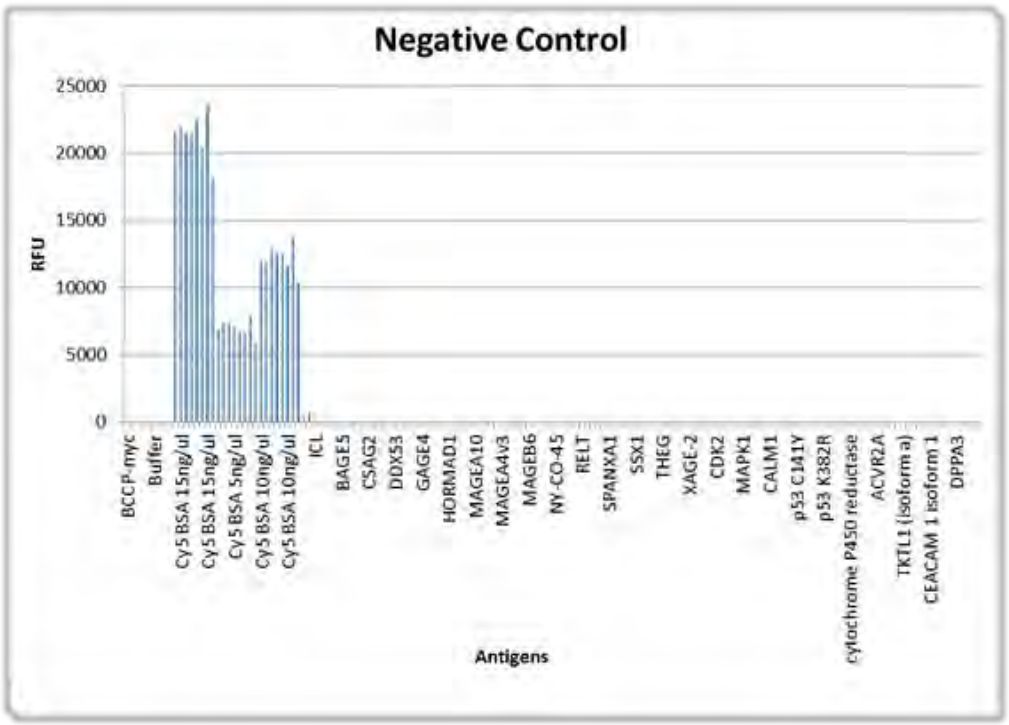


Figure 5.4. Bar graph demonstrating that negative control sera showed no reactivity to antigens on the array

The anti-c-myc-Cy3 assay was used to ensure that individual antigens were successfully immobilised to the array during printing (**Figure 5.5**). All spots demonstrated in **Figure 5.5** are

not of the same intensity because all 123 recombinant proteins are expressed to different degrees in the insect lysate and biologic processes like rate of degradation is not similar for all proteins on this array. They are linked to biotin and cMyc (positive control) to confirm their presence on the array. Albeit all signals are not discernible by naked eyes, the signals of the apparently invisible spots are read off by the ArrayPro Analyzer software and scored. (a bar graph of the c-Myc assay can be found in **Annexure IV**)

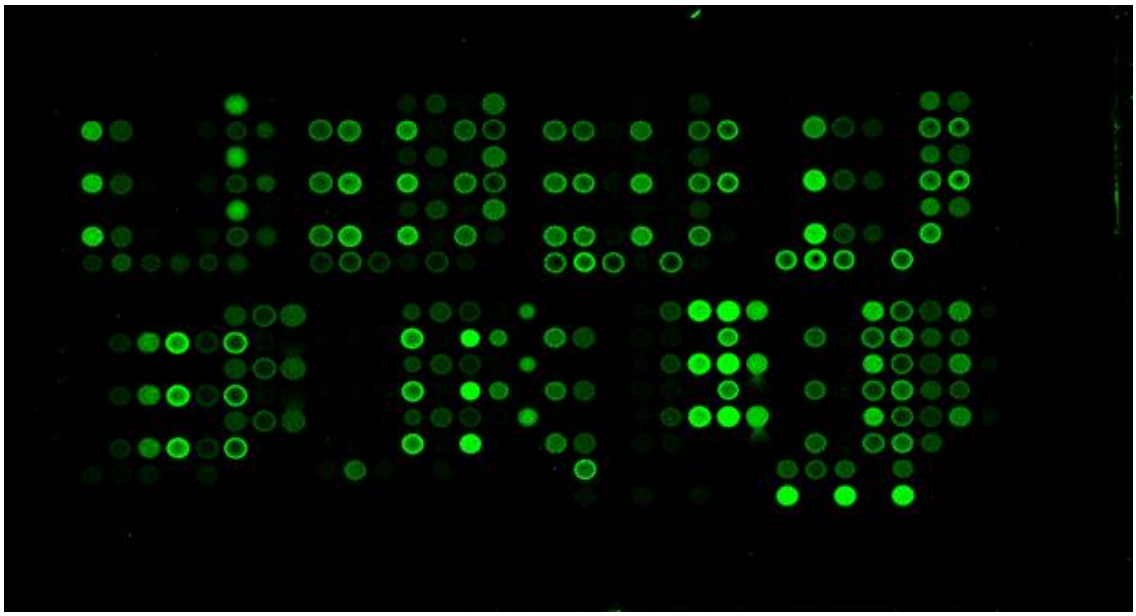


Figure 5.5. TIFF image showing successful immobilization of antigens on the array

For prostate cancer samples, higher autoantibody titres were found to ROPN1, GAGE1, PRKCZ, and SPANXA1 relative to other antigens (**Figure 5.6**); while two mutant p53 antigens, p53 S46A and p53 S15A had the highest autoantibody titres in DC samples (**Figure 5.7**).

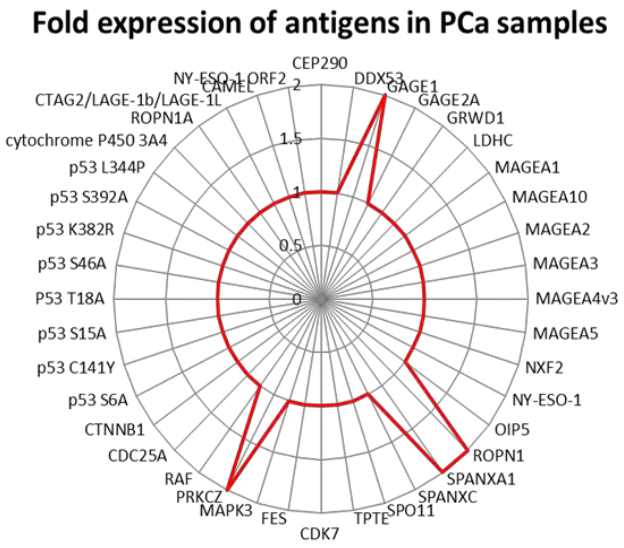


Figure 5.6. Radar plot showing ROPN1, PRKCZ, SPANXA1 and GAGE1 having higher autoantibody titer in prostate cancer

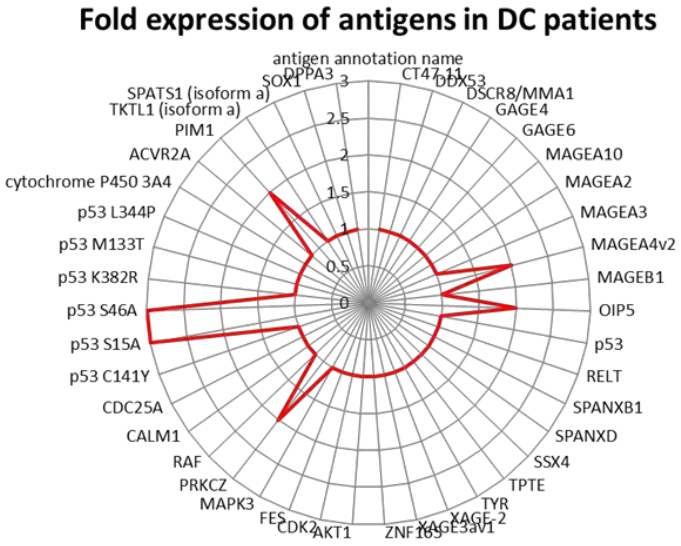


Figure 5.7. . Radar plot showing p53 S46A and p53 S15A having higher autoantibody titer in disease control sera

MAGEB1 and PRKCZ were the antigens found to have the highest autoantibody titres in BPH samples (**Figure 5.8**).

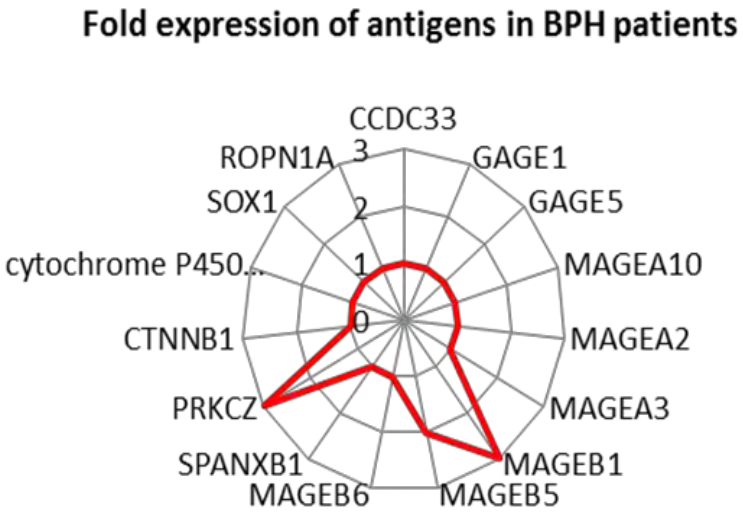


Figure 5.8. Radar plot showing MAGEB1 and PRKCZ having higher autoantibody titer in benign prostatic hyperplasia sera

There was a general variation in mean autoantibody response to TAAs, observed between PCa, BPH and DC samples (**Figure 5.9**). These highly differentially expressed autoantibodies were confirmed by ranking the autoantibody responses based on to their mean signal intensities and selecting the "top 20" intensities in each of the three categories for further analysis (**Table 5.4**). By using this approach, p53 S46A and SPANXA1 were not found for DC and PCa groups respectively, possibly due to the fact that this method focuses on the signal strength and not necessarily on presence or absence of autoantibody response.

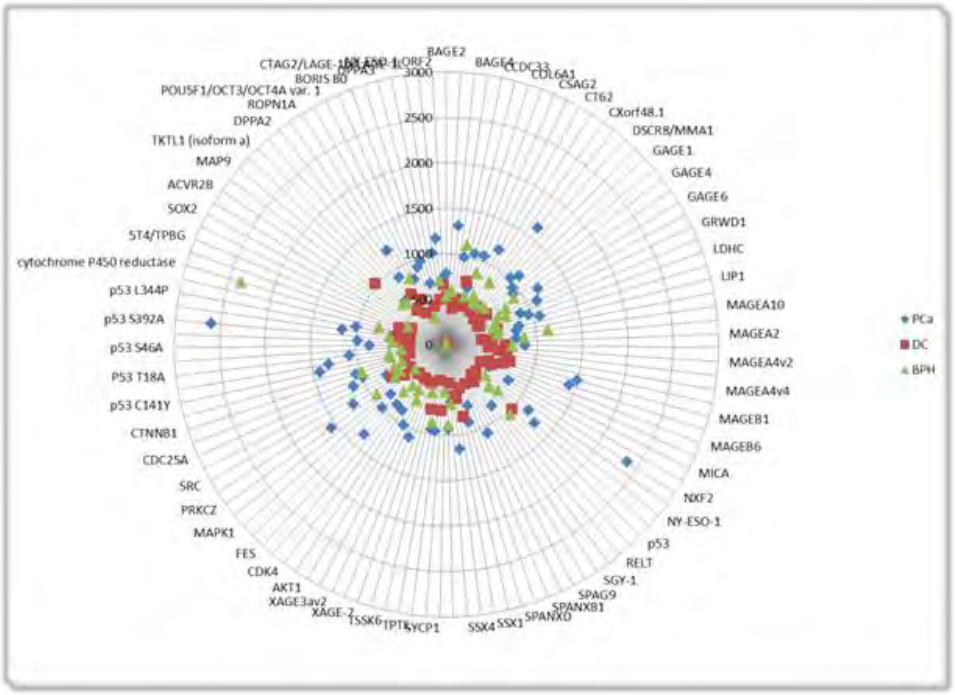


Figure 5.9. Radar plot showing difference in autoantibody response signals between PCa, BPH and DC

PCa		**BPH**		**DC**	
Antigens	Mean Intensity	Antigens	Mean Intensity	Antigens	Mean Intensity
PRKCZ	1048.3	PRKCZ	907.0	DDX53	813.5
p53 S15A	938.8	MAGEB1	846.6	MAGEB1	778.1
OIP5	849.1	GAGE2A	778.3	PRKCZ	757.3
MAPK3	846.2	GAGE1	766.9	GAGE4	668.6
MAGEB1	820.3	CCDC33	764.3	GAGE1	650.7
NY-ESO-1	816.6	GAGE5	734.1	GAGE6	648.4
ROPN1A	810.2	LEMD1	715.9	p53 S15A	616.0
p53 C141Y	702.5	OIP5	686.0	OIP5	600.2
P53 T18A	692.5	CTNNB1	625.9	MAPK3	598.9
p53 S46A	692.0	GAGE7	612.5	MAGEB5	584.3
LDHC	690.8	RAF	604.4	MAGEA2	583.7
CDC25A	683.8	MAGEB5	591.6	PBK	570.5
CTNNB1	678.2	MAGEA10	569.4	p53 Q136x	568.3
CAMEL	665.2	GAGE6	558.6	CALM1	565.1
GAGE1	653.9	CDC25A	555.1	LEMD1	560.5

p53 S6A	632.3	p53 S15A	551.1	MAGEB6	551.1
XAGE3av1	632.0	p53 C141Y	535.9	MAGEA10	545.2
GAGE6	629.7	CREB1	526.4	CT47.11	540.9
FES	607.1	GAGE4	500.3	P53 T18A	532.4
MAGEB5	600.1	MAGEA2	496.2	NY-ESO-1	529.9

Table 5.4. List of top ranking 20 antigens in PCa, BHP and DC by mean intensities

A third method of linear analysis which does not have intensity components was carried using a three-way Venn diagram for the "top 50" autoantibody responses ranked by mean pixel intensity. This analysis revealed that 25 (50%) of the antigens were common to PCa, BPH and DC groups, 9 (18%) of the antigens were only common to PCa and BPH. There were 6 (12%) antigens only common to BPH and DC, whilst only 4 (8%) antigens were common to PCa and DC. In total, 12 (24%), 15 (28%), and 10 (20%) antigens were found unique to PCa, DC and BPH respectively (**Figure 5.10**).

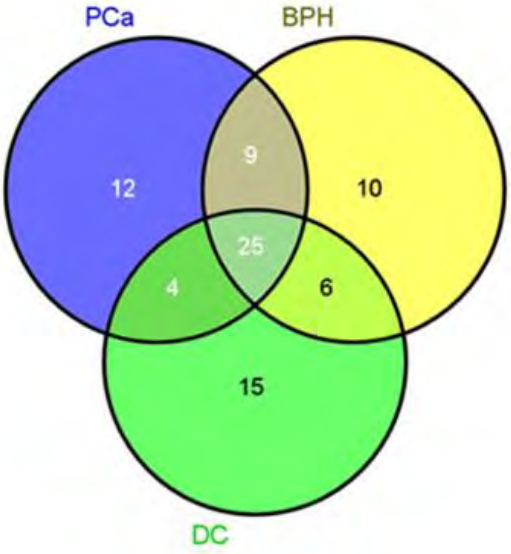


Figure 5.10. Three-way Venn diagram showing overlap in "Top 50" antigens by mean intensities

Interestingly, 6 (50%) of the 12 antigens found unique to PCa were found highly expressed by linear and differential expression analyses. These 6 antigens were LDHC, FGFR2, SPANXA1, p53 S392A, DPPA4, and BORIS BO. Similarly CSAG2, which was one of the 10 antigens found unique to BPH, was also observed to be differentially expressed between PCa and BPH. Wild-

type p53, DDX53, CT47.11 and p53 Q136X (a mutant p53) were found among the 15 antigens unique to DC (**Table 5.5**).

Unique to PCa	Unique to BPH	Unique to DC
CDK7	CSAG2	DDX53
SPANXA1	SSX4	p53 Q136x
p53 S392A	SPATS1 (isoform a)	p53
DPPA4	BAGE2	CALM1
LDHC	CDK4	CT47.11
FGFR2	ACVR2A	NY-ESO-1 ORF2
BORIS B0	XAGE3av2	DPPA2
TPTE	GAGE7	DPPA3
DSCR8/MMA1	AKT1	HORMAD1
ACVR2B	CTAG2	MART-1
SPO11		COL6A1
THEG		cytochrome P450 reductase
		SRC
		TSSK6
		CTAG2/LAGE-1b/LAGE-1L

Table.5.5. Venn diagram analysis showing autoantigens common to linear and differential analyses results

5.3.2. Differential expression analysis for Potential biomarkers

Perseus Software (*version 1.4.0.20*), Cluster (*Version 3.0*) and TreeView (*Version 3.0*) were employed for further bioinformatics analysis and visualization of the protein microarray data. Using an independent sample t-test with Bonferroni correction for multiple testing on the background corrected raw intensity data post-normalization; DPPA, NY-ESO-1, CEACAM1 isoform1, GAGE5, P53 L344P, EGFR, MAGEB5, CSAG2 and CCDC33 were differentially expressed (±2SD) between BPH and PCa (**Table 5.6**).

Upregulated	Downregulated
1. DPPA4	1. GAGE5
2. CECAM1 Isoform 1	2. MAGEB5
3. NY-ESO-1	3. EGFR
4. P53 L344P	4. CCDC33
	5. CSAG2

Table 5.6. Differentially expressed antigens between PCa and BPH by t-test

Using the same test for PCA *vs* DC, 8 differentially expressed (±2SD) antigens were identified including; P53 L344P, GAGE5, DPPA4, wild-type p53, ZNF, RAF, CT47.11, and DDX53 (**Table 5.7**).

Upregulated	Downregulated
1.P53 L344P	1.GAGE5
2.RAF	2.DDX53
3.DPPA4	3.CT47.11
4.ZNF165	4.P53
5.TKTL1 (Isoform a)	5.P53 Q136X

Table 5.7. Differentially expressed antigens between PCa and DC by t-test

Given that BPH is technically speaking, a form of disease control as well, all benign samples were further combined i.e. DC & BPH, and differentially expressed antigens between Benign and Malignant samples were explored. A total of 17 TAAs were found with higher autoantibody titres in PCa in comparison to benign conditions, whilst 6 antigens were found with lower antibody titres accordingly (**Table 5.8)**.

Upregulated	Downregulated
1.DPPA4	1.P53 Q136X
2.P53 L344P	2.MAGEB6
3.MAPK3	3.MAGEB5
4.CAMEL	4.GAGE5
5.RAF	5.CSAG2
6.TKTL1 (Isoform a)	6.PBK
7.LDHC	
8.P53 C141Y	
9.NY-ESO-1	
10.P53 K328R	
11.P53 S15A	
12.CDK2	
13.MAGE11	
14.FES	
15.P53 T18A	
16.OIP5	
17.SSX2A	

Table 5.8. Differentially expressed antigens between PCa and benign by t-test

Initial unsupervised hierarchical clustering of all antigens using either Perseus or Cluster revealed moderate molecular signature overlap between DC, PCa, and BPH (**Figure 5.11**). This overlapping patter was also observed in multivariate testing using principal component analysis (PCA), which revealed that individuals in distinct groups clustered haphazardly with other groups along the principal components both at 1-D and 3-D reconstruction (**Figure 5.12**).

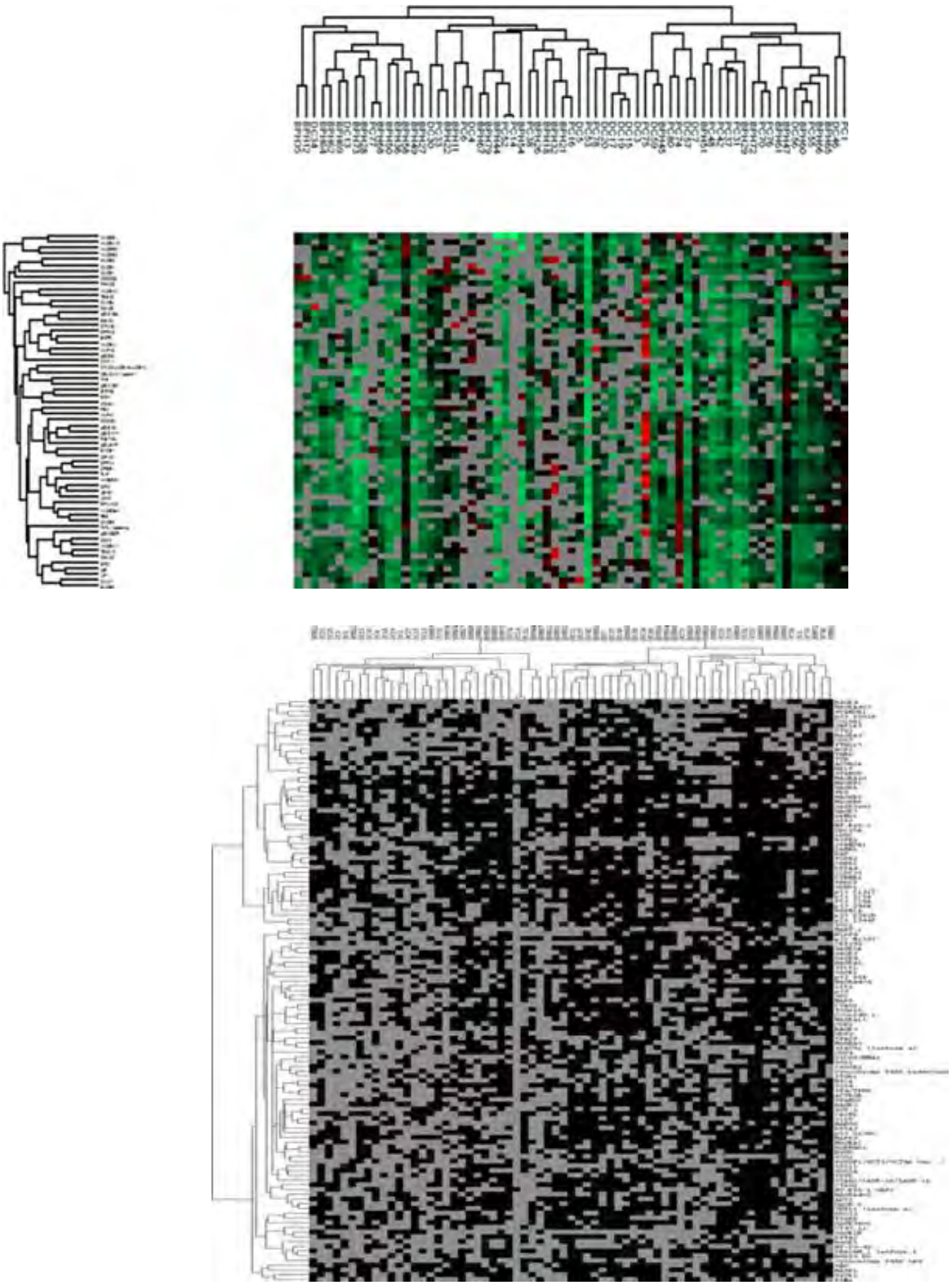


Figure 5.11. Unsupervised hierarchical clustering showing overlap in molecular signature on Perseus (top) and Cluster (bottom) analyses

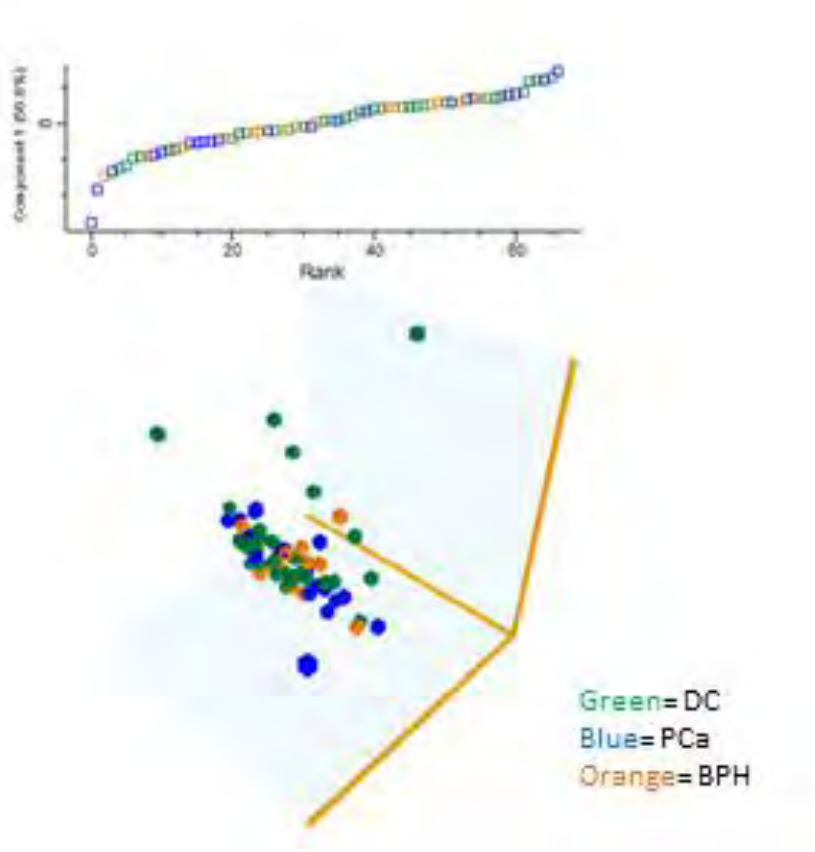


Figure 5.12. Overlap in molecular signature by multivariate test using 1D (left) and 3D (right) principal component analysis [455]

However, when the top ranking 10 TAAs with the highest autoantibody titres and presence in the 20 PCa samples were analysed, distinct grouping patterns were identified for BPH, PCa, and DC both by PCA and hierarchical clustering (**Figure 5.13 &5.14**).

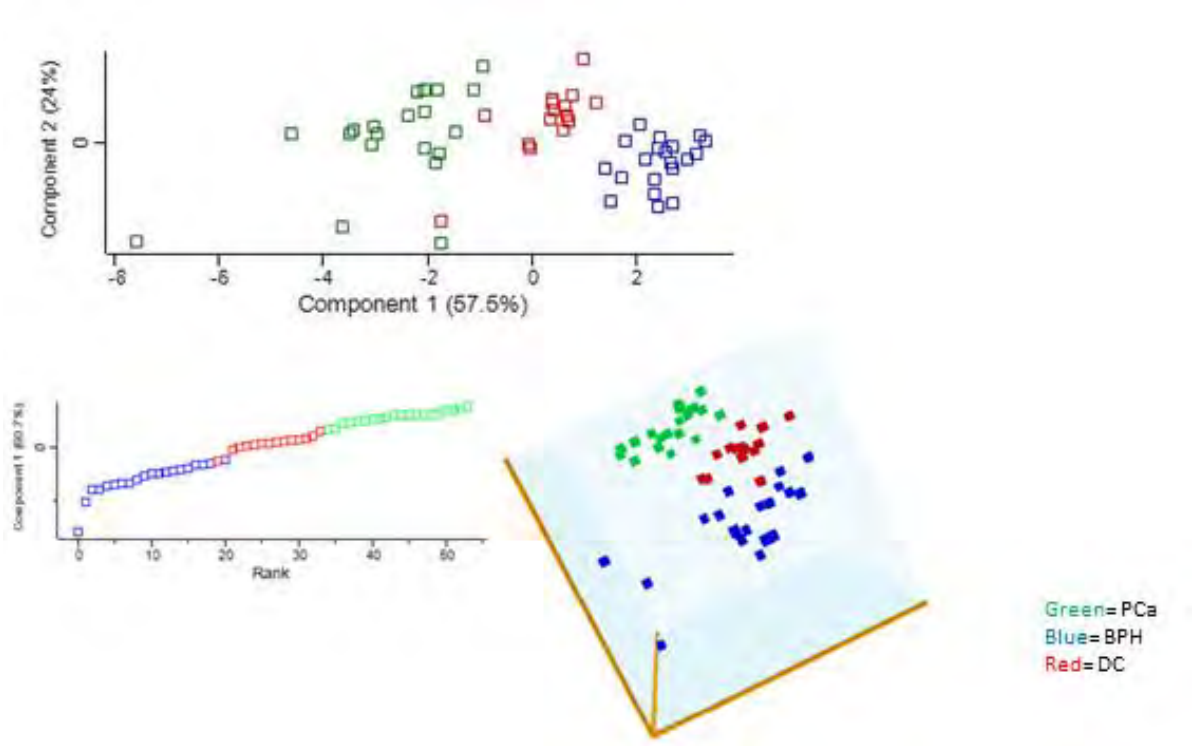


Figure 5.13. 1D (top left), 2D (top right), and 3D (bottom) principal component analysis of top 10 TAAs with highest titres in PCa [455]

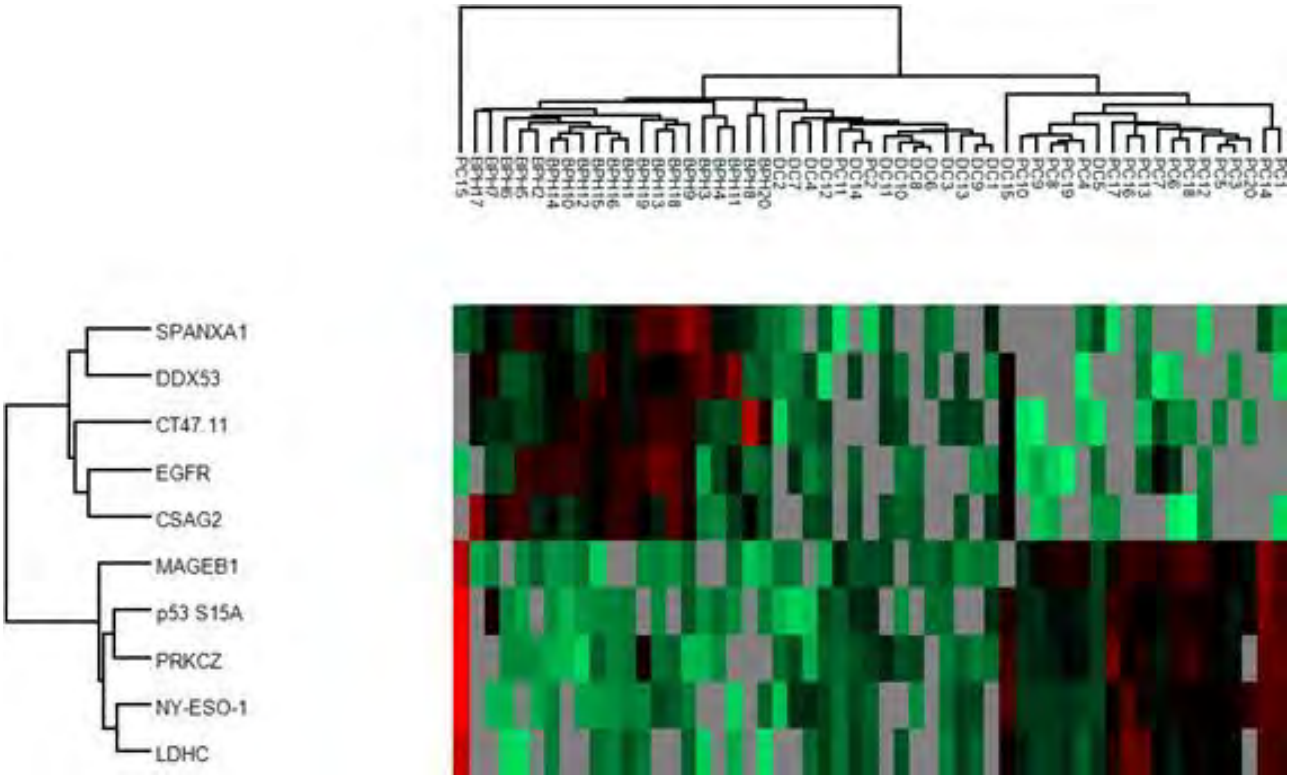


Figure 5.14. Unsupervised hierarchical clustering analysis of top 10 TAAs with highest titres in PCa

For the hierarchical clustering analysis, 17 (85%) out of the 20 PCa patients clustered together, whilst two PCa patients (PC11 & PC2) clustered with DC and one PCa patient (PC15) clustered separately, albeit closer to the BPH clusters. The two PCa samples that clustered with DC were re-evaluated and observed to have relatively lower PSA levels than other PCa cases. PC15 had a very high PSA level (315ng/mL), in spite of a moderate total Gleason score of 6. All BPH samples clustered distinctly together, while 13 (86.7%) of the 15 DC samples clustered together. Two of the DC samples (DC5 & DC15) clustered with PCa. On re-assessment of these samples, DC5 was found to be of a younger age (34 years) and of Africa ethnicity (Black), however his PSA level was not available; DC15 was a 58 year old patient with a PSA level of 6.8ng/mL and of Mixed-Ancestry ethnicity.

5.3.3. Racial variations in autoantibody response to potential TAA biomarkers

To determine if there was any variation in autoantibody response to TAAs between PCa patients drawn from three major ethnicities of which our cohort was composed; differences in autoantibody response to TAAs between the racial groups were evaluated. There were 6 from Caucasian African (White), 3 samples from Indigenous African (Black), and 11 from Mixed-Ancestry (Coloured) PCa patients. All the 41 potential antigens biomarkers were examined in our PCa patient cohort. We observed variation in autoantibody response to TAAs between the Caucasian African (PCa_C), Indigenous Africans (PCa_B) and Mixed Ancestry (PCa_M) prostate cancer patients (**Figure 5.15**).

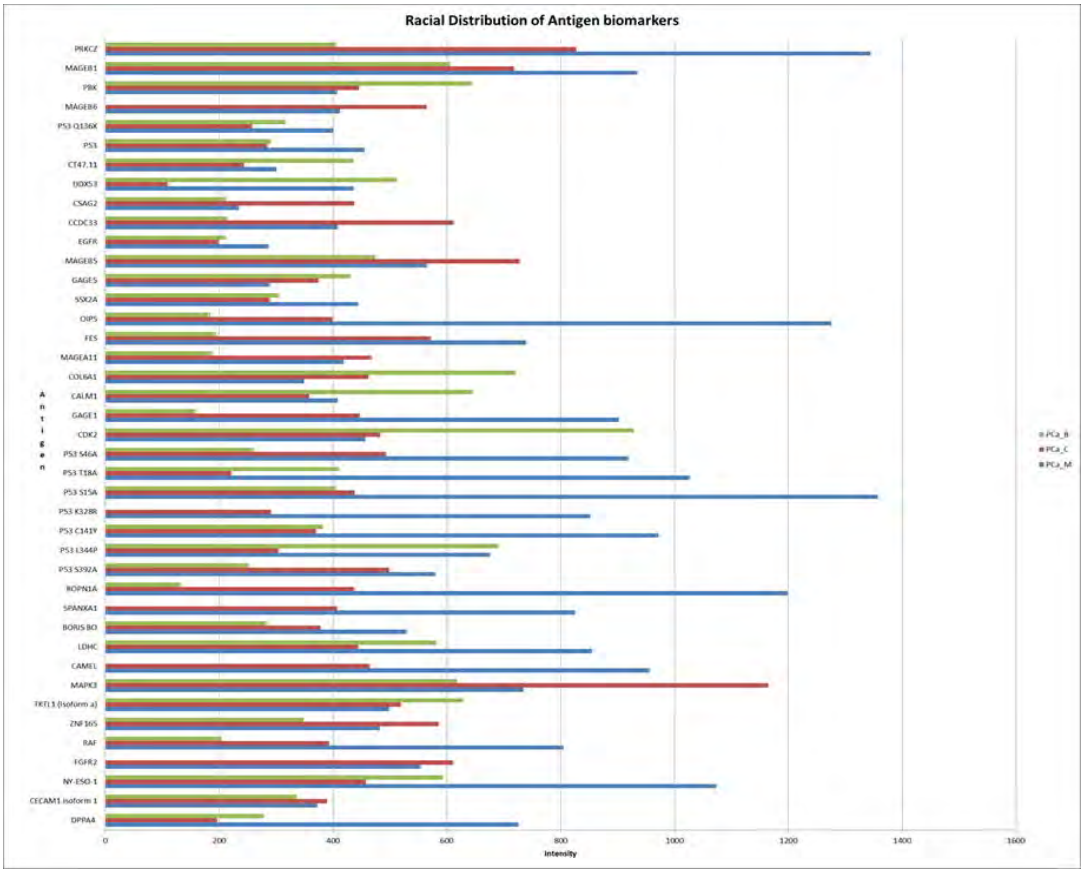


Figure 5.15. Variations in autoantibody response to 41 candidate TAA biomarkers in Caucasian, African and Mixed-Ancestry PCa patients

The mixed ancestry population had the highest autoantibody expression, which included DPPA4, NY-ESO-1, CAMEL, RAF, SPANXA1, GAGE1, ROPN1A, OIP5, PRKCZ and a subset of mutant p53 antigen including p53 T18A, p53 S15A, p53 S46A, p53 C141Y, and p53 K328R. From these 14 antigen, OIP5, p53 S15A, ROPN1A, and p53 T18A had the highest autoantibody response in PCa_M as compared with PCa_C or PCa_B. There were eight highly expressed autoantibodies for PCa_C in comparison to PCa_M or PCa_B, which included MAGEB6, CSAG2, MAGEB5, CCDC33, MAGEA11, FGFR2, MAKP3, and CEACAM1 Isoform 1. MAPK3 was found to be significantly more expressed in PCa_C as compared with others. Nine antigens were observed to have a higher autoantibody titer in PCa_B, which included PBK, DDX53, CT47.11, GAGE5, CALM1, COL6A1, CDK2, p53 L344P and TKTL Isoform a. Among these, COL6A1, PBK, CALM1 and CDK2 were found with the highest autoantibody titer in PCa_B.

5.3.4. Functional pathway analysis of potential PCa antigen biomarkers

The possibility of differentially expressed TAAs and CTAs being involved in a common or multiple biologic pathways was investigated to better understand and explore functional pathways in prostate cancer pathogenesis. Gene names or symbols for potential TAA biomarkers for prostate cancer were uploaded into GENEMANIA (http://www.genemania.org/) and bioinformatics enrichment analysis for the GO terms cellular component, molecular function and biological process was automatically carried out. Functionally clustering the "biological process" as the major Gene Ontology (GO) term for enrichment in the differentially expressed antigens revealed that they act via analogous signaling pathways involving: fibroblast growth receptor, Fc receptor, ERBB, epidermal growth factor receptor, neurotrophin, and ERK1/ERK2. These 6 signaling pathways were found similar to the lists of differentially expressed antigen between PCa and DC (**Table 5.9**) or between PCa and all controls (DC and BPH) (**Table 5.10**).

	Function	FDR	Coverage
1	query genes	n/a	11 / 11
2	Fc receptor signaling pathway	2.74E-7	9 / 219
3	epidermal growth factor receptor signaling pathway	2.74E-7	9 / 213
4	ERBB signaling pathway	2.74E-7	9 / 216
5	Fc-epsilon receptor signaling pathway	7.81E-7	8 / 166
6	neurotrophin signaling pathway	7.81E-7	9 / 277
7	neurotrophin TRK receptor signaling pathway	7.81E-7	9 / 274
8	cellular response to fibroblast growth factor stimulus	7.81E-7	8 / 183
9	fibroblast growth factor receptor signaling pathway	7.81E-7	8 / 170
10	response to fibroblast growth factor	7.81E-7	8 / 183
11	immune response-regulating cell surface receptor signaling pathway	1.23E-6	9 / 297
12	ERK1 and ERK2 cascade	5.76E-5	6 / 120
13	protein autophosphorylation	1.82E-4	6 / 149
14	positive regulation of MAPK cascade	1.82E-4	7 / 253
15	response to steroid hormone	2.46E-4	5 / 81
16	positive regulation of ERK1 and ERK2 cascade	2.6E-4	5 / 83
17	blood vessel morphogenesis	3.15E-4	6 / 170
18	regulation of ion homeostasis	5.83E-4	5 / 100
19	blood vessel development	5.9E-4	6 / 193
20	cell-matrix adhesion	6.04E-4	5 / 103
21	regulation of ERK1 and ERK2 cascade	8.33E-4	5 / 111
22	response to estrogen	8.48E-4	4 / 47
23	skin morphogenesis	9.29E-4	3 / 13

Table 5.9. Top ranking biologic processes between PCa and DC

When the antigens unique to PCa using the Venn diagram analysis were queried on GeneMANIA, cellular processes involved in reproduction were the major pathways involved (**Table 5.11**). Similarly, cellular process involved in reproduction was also one of the top 25 pathways common to "PCa *vs* all controls". To be able to determine whether this was the true picture at the protein expression level, we queried the same complement of genes using STRING (*version 9.1*) functional protein association network software.

	Function	FDR	Coverage
1	query genes	n/a	19 / 19
2	Fc-epsilon receptor signaling pathway	3.74E-4	7 / 166
3	Ras protein signal transduction	3.74E-4	7 / 167
4	Fc receptor signaling pathway	1.59E-3	7 / 219
5	fibroblast growth factor receptor signaling pathway	2.83E-3	6 / 170
6	neurotrophin TRK receptor signaling pathway	2.83E-3	7 / 274
7	immune response-regulating cell surface receptor signaling pathway	2.83E-3	7 / 297
8	cellular response to fibroblast growth factor stimulus	2.83E-3	6 / 183
9	neurotrophin signaling pathway	2.83E-3	7 / 277
10	response to insulin	2.83E-3	6 / 186
11	response to fibroblast growth factor	2.83E-3	6 / 183
12	cellular response to insulin stimulus	2.83E-3	6 / 182
13	small GTPase mediated signal transduction	2.83E-3	7 / 291
14	insulin receptor signaling pathway	2.83E-3	6 / 153
15	epidermal growth factor receptor signaling pathway	5.27E-3	6 / 213
16	ERBB signaling pathway	5.27E-3	6 / 216
17	ERK1 and ERK2 cascade	5.27E-3	5 / 120
18	cellular response to peptide hormone stimulus	9.98E-3	6 / 244
19	cellular response to peptide	1.01E-2	6 / 247
20	response to peptide hormone	1.15E-2	6 / 255
21	response to peptide	1.22E-2	6 / 260
22	positive regulation of ERK1 and ERK2 cascade	2.14E-2	4 / 83
23	cellular process involved in reproduction	5.06E-2	5 / 207
24	regulation of ERK1 and ERK2 cascade	6.1E-2	4 / 111
25	activation of MAPKK activity	7.66E-2	3 / 44

Table 5.10. Top ranking biologic processes between PCa and "all controls" (BPH and DC)

	Function	FDR	Coverage
1	query genes	n/a	12 / 12
2	synapsis	3.2E-2	3 / 17
3	chromosome organization involved in meiosis	3.2E-2	3 / 18
4	cellular process involved in reproduction	1.38E-1	5 / 207
5	meiosis I	3.71E-1	3 / 50
6	synaptonemal complex assembly	3.82E-1	2 / 10
7	synaptonemal complex organization	3.82E-1	2 / 10
8	meiotic nuclear division	8.59E-1	3 / 84
9	meiotic cell cycle	8.59E-1	3 / 87
10	synaptonemal complex	8.59E-1	2 / 18

Table 5.11. Top ranking biologic processes in PCa by Venn diagram analysis

By this approach, similar molecular signaling pathways were found involved both at the gene and protein expression level. (**Table 5.12**).

Term	NumberOfGenes	p-value	p-value_fdr	p-value_bonferroni
fibroblast growth factor receptor signaling pathway	8	1.68E-07	2.12E-03	2.12E-03
cellular response to fibroblast growth factor stimulus	8	4.64E-07	2.33E-03	5.85E-03
response to fibroblast growth factor	8	5.54E-07	2.33E-03	6.98E-03
epidermal growth factor receptor signaling pathway	8	1.17E-06	2.67E-03	1.48E-02
ERBB signaling pathway	8	1.27E-06	2.67E-03	1.60E-02
reproduction	17	1.78E-06	2.67E-03	2.24E-02
neurotrophin TRK receptor signaling pathway	9	1.79E-06	2.67E-03	2.25E-02
neurotrophin signaling pathway	9	1.84E-06	2.67E-03	2.32E-02
cell division	12	2.19E-06	2.67E-03	2.76E-02
cell cycle	17	2.30E-06	2.67E-03	2.91E-02
Fc receptor signaling pathway	9	2.33E-06	2.67E-03	2.94E-02
single organism reproductive process	15	3.63E-06	3.02E-03	4.57E-02
response to epidermal growth factor	4	3.64E-06	3.02E-03	4.59E-02
positive regulation of cyclin-dependent protein serine/threonine kinase activity involved in G1/S transition of mitotic cell cycle	3	3.64E-06	3.02E-03	4.59E-02
positive regulation of phosphorylation	13	3.73E-06	3.02E-03	4.71E-02
Fc-epsilon receptor signaling pathway	8	3.84E-06	3.02E-03	4.84E-02
response to oxygen-containing compound	16	4.86E-06	3.60E-03	6.12E-02
regulation of cyclin-dependent protein serine/threonine kinase activity involved in G1/S transition of mitotic cell cycle	3	5.44E-06	3.80E-03	6.86E-02
regulation of cell fate commitment	4	5.73E-06	3.80E-03	7.22E-02
synapsis	4	6.59E-06	4.02E-03	8.30E-02
cellular process involved in reproduction	10	6.69E-06	4.02E-03	8.44E-02
cell fate specification	5	7.48E-06	4.07E-03	9.43E-02
cellular response to mechanical stimulus	5	7.48E-06	4.07E-03	9.43E-02
positive regulation of cell cycle	6	7.75E-06	4.07E-03	9.77E-02
positive regulation of MAPK cascade	9	8.34E-06	4.21E-03	1.05E-01

Table 5.12. Top ranking biologic process top ranking TAA biomarkers using STRING software

Network generated from STRING demonstrated that CDK2, TP53, PBK, MAPK3, EGFR and CEACAM1 were strongly linked TAAs by neighborhood evidence (**Figure 5.16**). Although

some CTAs were not observed to be interacting physically with this TAA network, the possibility of their co-expression or predicted linkage is plausible. Furthermore, CTAG1A, GAGE1, CTAG2 and SSX7 were observed to be linked by neighborhood evidence. Co-occurrence, neighborhood and experimental evidences for TKTL1 and LDHC were also observed.

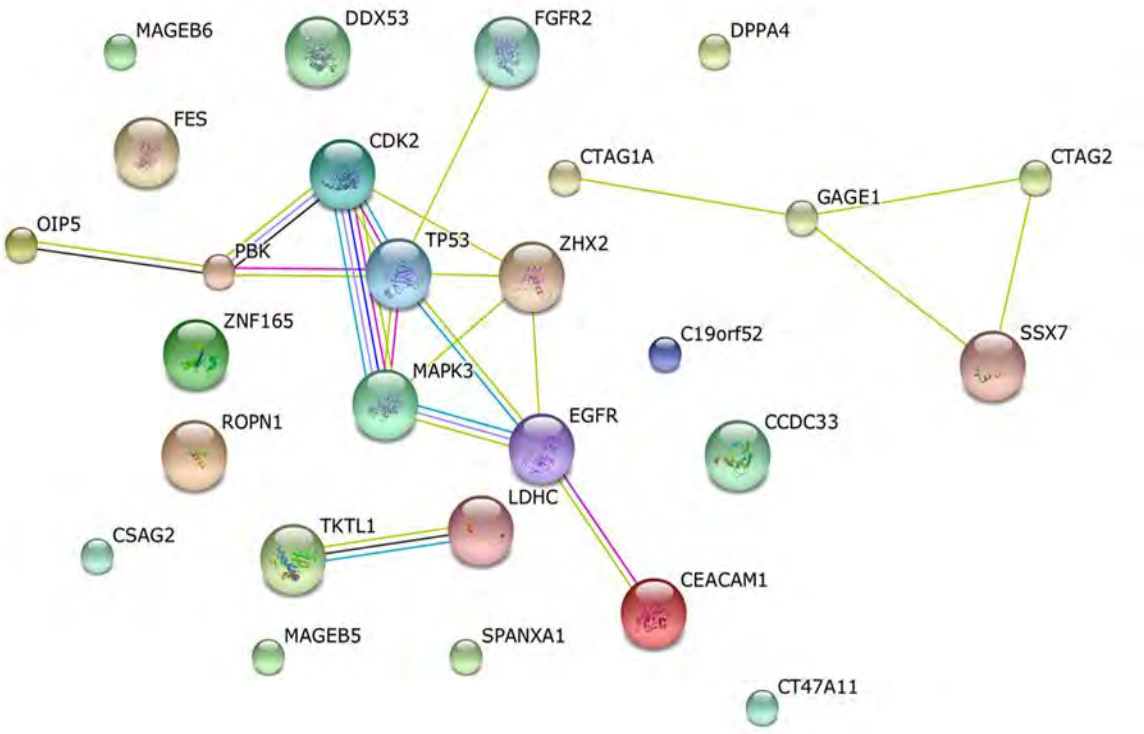


Figure 5.16. STRING Protein network interaction between top ranking prostate cancer TAA

5.3.5. Verification of TAAs in urinary shotgun proteomics data

The presence of TAA and CTA expression in urine of PCa patients in our cohort was assessed. Previously acquired urinary shotgun proteomics data [319] was assessed for the presence of all antigens on the CT100+ antigen array. Of the 123 antigens on the array, only 11 (8.9%) antigens were found in the urine of prostate cancer patient in our cohort. These included MAPK1,

FGFR2, COL6A1, CALM1, SOX1, SRC, LIP1, CEACAM1 isoform 1, RELT, ITGB1 and EGFR (**Table 5.13**). Further focusing on antigens without splice variants in the shotgun data, only 5 (4.5%) antigens namely COL6A1, RELT, CALM1, FGFR2 and ITGB1 were found in this database.

Antigen	Homologous genes detected	Presence in cancer antigen microarray data			Presence in Shotgun data	Presence in PCa shotgun	Biomarker Category
		BP	DC	PC			
FGFR2	FGFR2	low	low	high	yes	yes	Moderate
MAPK1	yes	low	low	high	yes	no	weak
COL6A1	COL6A1	low	low	very high	yes	Yes	Strong
SOX1	yes	low	low	high	yes	no	Weak
CALM1	CALM1	low	low	high	yes	Yes	Moderate
SRC	yes	low	high	low	yes	no	Weak
LIP1	yes	high	high	high	yes	no	Weak
CEACAM 1 isoform 1	yes	low	high	low	yes	no	Weak
RELT	RELT	low	high	high	yes	no	Weak
EGFR	yes	medium	high	low	yes	no	weak
ITGB1	ITGB1	high	low	low	yes	no	Weak

BP= Benign prostatic hyperplasia; **DC**= Disease control; **PC**= Prostate cancer

Table 5.13. Verification of TAAs in shotgun urinary proteomics data

Initial classification of the antigens identified based on their differential expression both in the shotgun proteomics experiment and on the cancer antigen array showed that out of these 5 antigens, only FGFR2, COL6A1, and CALM1 demonstrated promise for use as biomarkers of PCa by both methods.

5.4. Discussion

It is becoming evident that autoimmunity plays an important role in the process of carcinogenesis. Three plausible hypotheses for the origin of autoimmune response in cancer are: firstly, when cancer cells die, their content and membrane remnants are exposed, driving an autoimmune response in the body; secondly, cancer cells may carry aberrant immunogenic glycoproteins on their membranes which are displayed as Major Histocompatibility Complex (MHC) I molecule for antigen presentation to cytotoxic T Cells; thirdly, cancer cells or the remnants thereof may be engulfed by macrophages or antigen presenting cells and displayed as MHC II molecules for antigen presentation to T-helper cells.

In this study, a novel cancer antigen microarray platform was employed to evaluate serum autoantibodies to TAAs as potential biomarkers and immunotherapeutic targets for PCa in our South Africa cohort. Protein detection in serum ranges from fractions of a picogram to the gram range and on stimulation, their abundance can increase by as much as 10,000 folds. Owing to wide variations in the size and abundance of proteins present in serum [456], it is vital to evaluate the dynamic range and sensitivity of the microarray used in this study. Earlier protein microarray platforms were fraught by narrow dynamic range and low sensitivity for analytes of interest [457, 458]. However, remarkable improvement in dynamic range of newer protein microarrays and their sensitivity in the femtomolar range have increased their relevance [458]. Using a two-step linearity and dynamic range assay as described in previous work [277], a linear response in dynamic range of over 3 orders of magnitude was demonstrated. In addition, using a noise threshold of 2 standard deviations (of the background), a detection limit of ca. 1:1,000,000 serum dilution was observed, which corresponded to an autoantibody titer of about 190pg/mL [277].

Autoantibodies are more stable in serum than proteins or polynucleotides which may be rapidly degraded soon after their release by tumour cells [459]. Autoantibodies are capable of remaining in the blood long after the removal of antigenic stimulus [460], and they represent a promising measurable biosensor that can be correlated with different disease and health states; albeit we recognize the fact that autoantibody response between individuals is dynamic. Hence, measurement of autoantibody response to TAAs is a better approach than the direct measurement

of serum autoantigen. Due the detection of a few of the arrayed TAAs in our previous shotgun urinary proteomics data, we assumed that it is plausible that an increased autoantibody titer may be indirectly associated with increased urinary expression of the autoantigen.

Even though, the renal glomerular filtration barrier was anecdotally suggested to prevent the passage of macromolecules such as proteins due to a smaller pore size in comparison to serum protein size [461], there is overwhelming literature evidence for the presence of pathogen-specific antigens in urine [313, 462-471]. Suggested barriers to protein filtration include pore size, charge selectivity, shape of protein and potential for downstream tubular reabsorption [472, 473]. Although, it was reported that the human nephron size selectivity is about 50KDa [474]; the molecular weight of the overlapping TAAs detected in urine in this study ranged from 39-140KDa. This alludes to the possibility of an alternative mechanism such as shape/ structure might explain their presence in urine. In addition, it is plausible that detected overlapping TAAs in urine from shotgun proteomics data may be short proteolytic fragments or "degradome" of longer proteins. Emerging evidence suggests that low molecular mass proteins/antigens are able to pass through this filtration barrier and are concentrated in the proximal tubule of the kidney [474]; and then sequestered in the renal lymph nodes and dendritic cells. Such antigens stimulate PDL1-mediated apoptosis of antigen presenting T-cell. Hence there is physiologic hyporesponsiveness, inactivation or clonal deletion of cytotoxic T-cells against circulating innocuous proteins/antigen [474-476]. It is therefore plausible that that the low overlap between serum and urinary antigen stems from a combination of lability of serum antigens, renal filtration barrier, dendritic cells/ renal lymph node sequestration, inter alia. It therefore seems reasonable that urinary cancer antigen testing, just like pathogen-specific antigen testing may become more routinely used in clinics for cancer diagnosis.

Many antigens on our platform have been specifically documented to be dysregulated in PCa, albeit all are cancer-associated. For instance, promoter methylation of GAGE1 has been reported in PCa cell lines [477], while ROPN1 has been found to be expressed in a subset of acute myeloid leukemia (AML) patients [478]. SPANXA1 a member of the SPANX-A/D cluster of SPANX gene located on Xq27 [479] has been found highly expressed in cancer as well. Recently, a splice variant of PRKCZ has also been shown to be an emerging biomarker of malignant prostatic epithelium as well as PCa cell lines [480].

It is unclear why immunoreactivity of BPH sera to PRKCZ was high in our cohort; possibly indicating a malignant biotransformation potential of BPH. In addition, MAGEB1, which also had a high BPH sera immunoreactivity in our study was previously found to exhibit low immunoreactivity to PCa after administration of Lenalidomide (a thalidomide analogue) to PCa patients in a randomized phase II trial [481]. Generally p53 missence mutations were found more frequently in PCa and BPH sera, while wild-type p53 was found more commonly in DC. Two missence mutations p53 S15A and p53 S46A found in DC have been previously reported to be responsible for inactivation of constitutive phosphorylation of p53 [482] and dysregulation of apoptotic target genes, cell cycle, senescence as well as suppression of ERK activation [483, 484], respectively. Phosphorylation of p53 S46 appears to be important for maintaining genomic stability of cells after DNA damage and its high autoantibody response in our DC sera may indicate that a follow-up should be directed at patients who, despite presenting with symptom of prostate disease, screened negative histopathologically. There is evidence that anti-p53 (auto) antibodies isolated from cancer patients are sometimes unable to recognize wild-type p53; and most commercial antibodies and vaccines are directed to the p53 terminal domain epitopes which is similar in both wild-type and mutant p53 rather than the "core" DNA-binding domain where most of the mutations occur [485]. These observations support the idea that differential signals to p53 variants on our array are real. We found considerable overlap between potential antigen biomarkers of PCa identified by different analyses, notably BORIS BO which was unique to PCa in our 3-way Venn diagram analysis has been reported recently to correlate at mRNA level with prostate cancer aggression and is potentially able to activate the androgen receptor genes [453].

Class comparison of differentially expressed antigens in this study showed that DPPA4, a nuclear chromatin associated embryonic stem cell protein was more upregulated in PCa sera compared to BPH and DC. Expression of DPPA4 has been reported in PCa cell lines but its role in somatic cancer is largely unknown [486]. Higher titer of NY-ESO-1 antigen in PCa serum in comparison to BPH and DC demonstrates that this antigen which is known to generate autoantibody response in many human cancers, could be a part of diagnostic antigen biomarker panel for PCa. We found the autoantibody response to CSAG2 antigen to be lower in our PCa sera in comparison to BPH, albeit previous reports did not find any difference in CSAG2 gene expression between PCa and normal prostate tissue [454]. GAGE5 antigen which had a low expression in our PCa sera compared to BPH and DC was reported not to be significantly

expressed by SKOV-3 human ovarian cancer cell lines [487]. Differential antigen expression detected in this study represents a potential niche for diagnostic biomarker and PCa immunotherapeutic target detection for African populations.

Variation in antigen expression was observed here between indigenous African, Mixed Ancestry and Caucasian PCa patients in our cohort. Notably, ROPN1A, OIP5 and two mutant p53 antigens (S15A & T18A) were highly expressed in PCa patients of Mixed Ancestry origin. Mutation of Thr18 to alanine in p53 using site-directed mutagenesis has been reported to result in the reduction of thioredoxin reductase expression, DNA damage, increased expression of p21, and increased production of reactive oxygen species [488]. In a report by Lee *et al* [489], OIP5 has been described as a targetable CTA for colon and gastric cancer using cell lines and cancer tissues. MAPK3 was found to be significantly highly expressed in Caucasian patients in our cohort in comparison to Indigenous Africans or Mixed Ancestry PCa patient. We observed that CSAG2, even though previously reported to be highly expressed in prostate cancer [450], was generally higher in our BPH as compared with PCa sera. However, focusing on our PCa data alone we observed that CSAG2 was higher in the Caucasian sera, as compared with other racial groups. This difference is possibly due to the population from which these samples were drawn and the fact that formalin fixed paraffin embedded (FFPE) sections were used in previous studies [450]. In addition, our cohort may not be large enough to conclude that CSAG2 is lower in African PCa patients as compared to their Caucasian counterparts. Hence, this observation warrants further studies using samples drawn from Africa PCa patients' populations. We also found COL6A1, CALM1, PBK and CDK2 to be highly expressed in Indigenous Africa PCa patients in comparison to Caucasian and Mixed Ancestry PCa patients in our cohort. PBK and CSAG were found to correlate significantly with histopathologic progression using Gleason's score in the same study [450]. Taking advantage of racial variation in TAA antigen expression may be a plausible approach to PCa theranostics in the future.

It was evident from pathway analysis that the strong link between CDK2, MAPK3, EGFR, CEACAM1, PBK, OIP5, FGFR and TP53 seem to play crucial roles in the pathogenetic mechanism of PCa development. In addition, other networks such as CTAG2, GAGE1, SSX7 and CTAG1A; as well as TKTL1 and LDHC need further study *vis-à-vis* PCa pathogenesis. It is also noteworthy that some of the potential serological TAA biomarkers of PCa were also

identified in our previous urinary shotgun discovery proteomics data [319]. This observation makes sense because the majority of our potential urinary biomarkers were predicted membrane and intercellular matrix/space proteins and the process of cancer development involves disruption of cell-cell contact and discohesive tendencies of the normal prostate epithelial architecture.

As demonstrated in a previous study using CT100+ cancer antigen array platform [277], results generated were comparable to an independently generated ELISA assay. Hence, an interesting approach to antigen array result validation is to use a multiplexed mini-array of fully validated biomarker subsets for in-vitro clinical immunodiagnostics. A panel of antigens on the full array to which autoantibody response is found can be miniaturized for point-of-care (POC) diagnostics. Such mini arrays have been previously used for confirmatory identification of differential autoantibody response signature for Hepatocellular [490], Ovarian [491], Breast [492], and Colon [493] cancer. The cohort of differential autoantibody response signature discovered for prostate cancer in the current study could serve as a primer for its immunodiagnostic mini-array fabrication.

In this study, potential serologic diagnostic biomarkers and immunotherapeutic targets for PCa were explored in sub-Saharan Africa (SSA), a sub-continent where infection plays a pivotal role in cancer pathogenesis [494, 495]. It is envisaged that cancer vaccination would play an important role in reduction of the cancer burden in SSA. Emerging cancer immunotherapy techniques such as checkpoint blocking antibodies [496] and genetically engineered T-cells [497] are beginning to take center stage, although great attention needs to be paid to possible side effects of their use. There are various novel vaccination-based therapies currently in the pipeline for different stages of PCa [498], but several bottlenecks need to be addressed to bring these to routine clinical utility [10]. This study serves as a foundation for further immunological prostate cancer theranostic research in Africa and development of point-of-care (POC) tools for timely diagnosis and treatment monitoring of PCa, as well as for patient stratification prior to therapeutic vaccination.

CHAPTER 6: CONCLUSION AND RECOMMENDATIONS

This study provides better insight into various aspects of prostate cancer biology, diagnosis and treatment in Southern African subjects using proteomics methodologies. Without fractionation, this study identified close to 1,800 protein groups, which is close to some of the largest urinary proteome coverages identified in Western studies (even with fractionation) [322]. Even though, no study so far has directly proved the causal relationship between prostate cancer and BPH, studies have shown that BPH often co-exist with PCa at diagnosis and increases the potential for diagnosis of incidental cancers [499]. In a 27 year Danish follow-up study of over 3 million men, BPH was found to increase the risk of PCa incidence and mortality by up to three and eight fold respectively [500], albeit the authors disclaimed causality relationship between the BPH and PCa. Other studies have suggested similar immunologic [501] and molecular [502] pathogenesis of BPH and PCa, but did not establish causality relationship between the two. The observation of proteomics overlap in this study provides further evidence for malignant biotransformation potential of BPH. The observed overlap may also point to the disparity between histopathological and molecular level of diagnosis; suggesting that treatment should not be solely based on histopathologic classification.

Because of the vastness of its database, the Human Protein Atlas has been recommended as a very effective proteomics tool for biomarker discovery [414], clinical biomarker studies [503], and also an important tool for pathologic studies [413]. SRMAtlas on the other hand, has also been described as essential for *in-silico* prediction of the best target transitions for biomarker validation [504], because its database contains a compendium of MS-detectable transitions for a host of target peptides. In addition, SRMatlas has been found beneficial as adjunct in biomarker discovery and validation for various diseases [505-508]. More than half of the selected proteotypic peptides have been previously assayed, and have adequate information in the SRMAtlas database. This indicates that using SRMAtlas prior to definitive targeted validation is a reasonable target transition screening method.

Even though, cancers due to environmental exposure are on the decrease in high-income nations, it is reported that such cancers are on the increase in low and middle income countries [509].

Chronic infection has been linked with cancer development and Africa bears a great burden of such infection-related cancers in comparison to other regions of the world [510-514]. Chronic infection is known to lead to the development of Burkitt's lymphoma (EBV), Kaposi sarcoma (HHV8), hepatocellular carcinoma (HBV), gastric cancer (H.*Pylori*) and cervical cancer (HPV). In addition, Acquired Immunodeficiency Syndrome (HIV) has been known to be associated with the development of various cancers. It seems plausible that immunotherapy and vaccine development may play an important role in cancer control in Africa. Several studies have shown a link between infection to PCa [515-517], but none has established causal relationship yet.

Similar to what was found in the urinary shotgun proteomics, some of the autoantibody signatures overlapped between BPH and PCa. For instance, PRKCZ which has been previously described as highly upregulated in PCa [480], was also detected as upregulated in BPH in this study. This further alludes to the potential precursor relationship between BPH and PCa. BORIS BO which we identified as a potential biomarker of PCa has been recently identified as a marker of androgen receptor activation as well as aggressive progression in prostate cancer [453]. High autoantibody titres to putative CTA biomarkers of PCa such as NY-ESO-1, DPPA4 and ROPN1A in this study indicate that our panel of differentially expressed autoantigen is reliable. In addition, the link between CDK2, MAPK3, EGFR, CEACAM1, PBK, OIP5, FGFR and TP53 by functional network analysis needs further elucidation. These pathways seem to be highly related for prostate carcinogenesis in the African population. The higher autoantibody expression level to CTAs observed in the mixed ancestry population in comparison to men of African or Caucasian descent warrants further studies.

This study identified novel non-invasive potential urinary and serological biomarkers of PCa in a heterogeneous cohort of South African patients. Draft physiologic and PCa-specific urinary proteomes were established for our South African cohort. Ethnic trends were identified in potential urinary and serologic biomarkers signatures which may ameliorate the development of ethnic-tailored biomarkers for PCa diagnosis and management in South Africa. Based on these preliminary findings, it is recommended that more global funding should be directed towards cancer research in Africa. Not least, further research should be directed towards understanding the potential role of prostate cancer immunotherapy in African patients. It is also suggested, that

a centralized proteome repository should be built for various diseases in Africa, as this would provide a resource for better research and understanding of disease pathogenesis in the continent.

CHAPTER 7: REFERENCES

1. Leissner KH, Tisell LE. The weight of the human prostate. *Scandinavian journal of urology and nephrology* 13(2), 137-142 (1979).
2. Schulze H, Isaacs JT, Coffey DS. A critical review of the concept of total androgen ablation in the treatment of prostate cancer. *Progress in clinical and biological research* 243A, 1-19 (1987).
3. Kumar VL, Majumder PK. Prostate gland: Structure, functions and regulation. *International urology and nephrology* 27(3), 231-243 (1995).
4. Bhavsar A, Verma S. Anatomic imaging of the prostate. *BioMed research international* 2014, 728539 (2014).
5. Hiraoka Y, Akimoto M. Anatomy of the prostate from fetus to adult - origin of benign prostatic hyperplasia. *Urology Research* 15(3), 177-180 (1987).
6. Mcneal JE. The zonal anatomy of the prostate. *The Prostate* 2(1), 35-49 (1981).
7. Emami N, Deperthes D, Malm J, Diamandis EP. Major role of human klk14 in seminal clot liquefaction. *The Journal of biological chemistry* 283(28), 19561-19569 (2008).
8. Sitzmann BD, Urbanski HF, Ottinger MA. Aging in male primates: Reproductive decline, effects of calorie restriction and future research potential. *Age* 30(2-3), 157-168 (2008).
9. Izadpanahi MH, Honarmand R, Khorrami MH, Najarzadegan MR, Sichani MM, Alizadeh F. A comparison of bladder neck preservation and bladder neck reconstruction for urinary incontinence after radical retro pubic prostatectomy. *Journal of research in medical sciences : the official journal of Isfahan University of Medical Sciences* 19(12), 1140-1144 (2014).
10. Baxevanis CN, Papamichail M, Perez SA. Prostate cancer vaccines: The long road to clinical application. *Cancer immunology, immunotherapy : CII* 64(4), 401-408 (2015).
11. Arnold JT, Isaacs JT. Mechanisms involved in the progression of androgen-independent prostate cancers: It is not only the cancer cell's fault. *Endocrine-related cancer* 9(1), 61-73 (2002).
12. Sun Y, Niu J, Huang J. Neuroendocrine differentiation in prostate cancer. *American journal of translational research* 1(2), 148-162 (2009).
13. Terry S, Beltran H. The many faces of neuroendocrine differentiation in prostate cancer progression. *Frontiers in oncology* 4, 60 (2014).
14. Hu CD, Choo R, Huang J. Neuroendocrine differentiation in prostate cancer: A mechanism of radioresistance and treatment failure. *Frontiers in oncology* 5, 90 (2015).
15. Mosca A, Berruti A, Russo L, Torta M, Dogliotti L. The neuroendocrine phenotype in prostate cancer: Basic and clinical aspects. *Journal of endocrinological investigation* 28(11 Suppl International), 141-145 (2005).
16. Wernert N. [immunohistochemistry of the prostate and prostate carcinomas]. *Veroffentlichungen aus der Pathologie* 135, 1-163 (1991).
17. Leung CS, Srigley JR. Distribution of lipochrome pigment in the prostate gland: Biological and diagnostic implications. *Human pathology* 26(12), 1302-1307 (1995).
18. Sajjad Y. Development of the genital ducts and external genitalia in the early human embryo. *The journal of obstetrics and gynaecology research* 36(5), 929-937 (2010).
19. White CW, Xie JH, Ventura S. Age-related changes in the innervation of the prostate gland: Implications for prostate cancer initiation and progression. *Organogenesis* 9(3), 206-215 (2013).
20. Cunha GR, Donjacour AA, Cooke PS *et al*. The endocrinology and developmental biology of the prostate. *Endocrine reviews* 8(3), 338-362 (1987).

21. Baan R, Grosse Y, Straif K *et al*. A review of human carcinogens--part f: Chemical agents and related occupations. *The lancet oncology* 10(12), 1143-1144 (2009).
22. Wogan GN, Hecht SS, Felton JS, Conney AH, Loeb LA. Environmental and chemical carcinogenesis. *Seminars in cancer biology* 14(6), 473-486 (2004).
23. Hodgson S. Mechanisms of inherited cancer susceptibility. *Journal of Zhejiang University. Science. B* 9(1), 1-4 (2008).
24. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 100(1), 57-70 (2000).
25. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell* 144(5), 646-674 (2011).
26. Floor SL, Dumont JE, Maenhaut C, Raspe E. Hallmarks of cancer: Of all cancer cells, all the time? *Trends in molecular medicine* 18(9), 509-515 (2012).
27. Ward PS, Thompson CB. Metabolic reprogramming: A cancer hallmark even warburg did not anticipate. *Cancer cell* 21(3), 297-308 (2012).
28. Gerlinger M. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing (vol 366, pg 883, 2012). *New Engl J Med* 367(10), 976-976 (2012).
29. Magee JA, Piskounova E, Morrison SJ. Cancer stem cells: Impact, heterogeneity, and uncertainty. *Cancer cell* 21(3), 283-296 (2012).
30. Pollock PA, Ludgate A, Wassersug RJ. In 2124, half of all men can count on developing prostate cancer. *Current oncology* 22(1), 10-12 (2015).
31. Ganz PA, Barry JM, Burke W *et al*. Nih state-of-the-science conference statement: Role of active surveillance in the management of men with localized prostate cancer. *NIH consensus and state-of-the-science statements* 28(1), 1-27 (2011).
32. Ferlay J, Soerjomataram I, Dikshit R *et al*. Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012. *International journal of cancer. Journal international du cancer* 136(5), E359-386 (2015).
33. De Marzo AM, Meeker AK, Zha S *et al*. Human prostate cancer precursors and pathobiology. *Urology* 62(5A), 55-62 (2003).
34. Chandler PD, Scott JB, Drake BF *et al*. Impact of vitamin d supplementation on inflammatory markers in african americans: Results of a four-arm, randomized, placebo-controlled trial. *Cancer prevention research* 7(2), 218-225 (2014).
35. Gann PH, Ma J, Giovannucci E *et al*. Lower prostate cancer risk in men with elevated plasma lycopene levels: Results of a prospective analysis. *Cancer research* 59(6), 1225-1230 (1999).
36. Cohen JH, Kristal AR, Stanford JL. Fruit and vegetable intakes and prostate cancer risk. *Journal of the National Cancer Institute* 92(1), 61-68 (2000).
37. Heinonen OP, Albanes D, Virtamo J *et al*. Prostate cancer and supplementation with alpha-tocopherol and beta-carotene: Incidence and mortality in a controlled trial. *Journal of the National Cancer Institute* 90(6), 440-446 (1998).
38. Clark LC, Dalkin B, Krongrad A *et al*. Decreased incidence of prostate cancer with selenium supplementation: Results of a double-blind cancer prevention trial. *British journal of urology* 81(5), 730-734 (1998).
39. Yan L, Spitznagel EL. Soy consumption and prostate cancer risk in men: A revisit of a meta-analysis. *The American journal of clinical nutrition* 89(4), 1155-1163 (2009).
40. Macinnis RJ, English DR. Body size and composition and prostate cancer risk: Systematic review and meta-regression analysis. *Cancer causes & control : CCC* 17(8), 989-1003 (2006).
41. Zheng J, Yang B, Huang T, Yu Y, Yang J, Li D. Green tea and black tea consumption and prostate cancer risk: An exploratory meta-analysis of observational studies. *Nutrition and cancer* 63(5), 663-672 (2011).
42. Rose DP, Connolly JM. Omega-3 fatty acids as cancer chemopreventive agents. *Pharmacology & therapeutics* 83(3), 217-244 (1999).

43. Qin LQ, Xu HY, Wang PY, Tong H, Hoshi K. Milk consumption is a risk factor for prostate cancer in western countries: Evidence from cohort studies. *Asia Pac J Clin Nutr* 16(3), 467-476 (2007).
44. Rohrmann S, Linseisen J, Key TJ *et al*. Alcohol consumption and the risk for prostate cancer in the european prospective investigation into cancer and nutrition. *Cancer Epidem Biomar* 17(5), 1282-1287 (2008).
45. Platz EA, Leitzmann MF, Rimm EB, Willett WC, Giovannucci E. Alcohol intake, drinking patterns, and risk of prostate cancer in a large prospective cohort study. *American journal of epidemiology* 159(5), 444-453 (2004).
46. Huncharek M, Haddock KS, Reid R, Kupelnick B. Smoking as a risk factor for prostate cancer: A meta-analysis of 24 prospective cohort studies. *American journal of public health* 100(4), 693-701 (2010).
47. Kasper JS, Giovannucci E. A meta-analysis of diabetes mellitus and the risk of prostate cancer. *Cancer Epidem Biomar* 15(11), 2056-2062 (2006).
48. Nelson WG, De Marzo AM, Isaacs WB. Mechanisms of disease: Prostate cancer. *New Engl J Med* 349(4), 366-381 (2003).
49. Nelson CP, Kidd LCR, Sauvageot J *et al*. Protection against 2-hydroxyamino-1-methyl-6-phenylimidazo[4,5-b]pyridine cytotoxicity and DNA adduct formation in human prostate by glutathione s-transferase p1. *Cancer research* 61(1), 103-109 (2001).
50. De Marzo AM, Marchi VL, Epstein JI, Nelson WG. Proliferative inflammatory atrophy of the prostate - implications for prostatic carcinogenesis. *American Journal of Pathology* 155(6), 1985-1992 (1999).
51. Hayes RB, Pottern LM, Strickler H *et al*. Sexual behaviour, stds and risks for prostate cancer. *British journal of cancer* 82(3), 718-725 (2000).
52. Choi YS, Kim KS, Choi SW *et al*. Microbiological etiology of bacterial prostatitis in general hospital and primary care clinic in korea. *Prostate international* 1(3), 133-138 (2013).
53. Mcconnell JD. The pathophysiology of benign prostatic hyperplasia. *Journal of andrology* 12(6), 356-363 (1991).
54. Brawer MK. Prostatic intraepithelial neoplasia: An overview. *Reviews in urology* 7 Suppl 3, S11-18 (2005).
55. Humphrey PA. Histological variants of prostatic carcinoma and their significance. *Histopathology* 60(1), 59-74 (2012).
56. Shah RB. Current perspectives on the gleason grading of prostate cancer. *Archives of pathology & laboratory medicine* 133(11), 1810-1816 (2009).
57. Humphrey PA. Gleason grading and prognostic factors in carcinoma of the prostate. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 17(3), 292-306 (2004).
58. Bulut S, Aktas BK, Gokkaya CS *et al*. Association between pre-biopsy white blood cell count and prostate biopsy - related sepsis. *Central European journal of urology* 68(1), 86-90 (2015).
59. Underwood DJ, Zhang J, Denton BT, Shah ND, Inman BA. Simulation optimization of psa-threshold based prostate cancer screening policies. *Health care management science* 15(4), 293-309 (2012).
60. Vickers AJ, Brewster SF. Psa velocity and doubling time in diagnosis and prognosis of prostate cancer. *Br J Med Surg Urol* 5(4), 162-168 (2012).
61. Thomsen FB, Roder MA, Hvarness H, Iversen P, Brasso K. Active surveillance can reduce overtreatment in patients with low-risk prostate cancer. *Danish medical journal* 60(2), A4575 (2013).
62. Moul JW, Wu H, Sun L *et al*. Early versus delayed hormonal therapy for prostate specific antigen only recurrence of prostate cancer after radical prostatectomy. *J Urol* 171(3), 1141-1147 (2004).

63. Sullivan KF, Crawford ED. Targeted focal therapy for prostate cancer: A review of the literature. *Therapeutic advances in urology* 1(3), 149-159 (2009).
64. Madan RA, Gulley JL. The current and emerging role of immunotherapy in prostate cancer. *Clinical genitourinary cancer* 8(1), 10-16 (2010).
65. Suzman DL, Boikos SA, Carducci MA. Bone-targeting agents in prostate cancer. *Cancer metastasis reviews* 33(2-3), 619-628 (2014).
66. Xie X, Kong Y, Tang H, Yang L, Hsu JL, Hung MC. Targeted bikdd expression kills androgen-dependent and castration-resistant prostate cancer cells. *Molecular cancer therapeutics* 13(7), 1813-1825 (2014).
67. El-Amm J, Aragon-Ching JB. The changing landscape in the treatment of metastatic castration-resistant prostate cancer. *Therapeutic advances in medical oncology* 5(1), 25-40 (2013).
68. Schweizer MT, Antonarakis ES. Abiraterone and other novel androgen-directed strategies for the treatment of prostate cancer: A new era of hormonal therapies is born. *Therapeutic advances in urology* 4(4), 167-178 (2012).
69. Bielekova B, Martin R. Development of biomarkers in multiple sclerosis. *Brain : a journal of neurology* 127(Pt 7), 1463-1478 (2004).
70. Madu CO, Lu Y. Novel diagnostic biomarkers for prostate cancer. *Journal of Cancer* 1, 150-177 (2010).
71. Velonas VM, Woo HH, Dos Remedios CG, Assinder SJ. Current status of biomarkers for prostate cancer. *International journal of molecular sciences* 14(6), 11034-11060 (2013).
72. Graddis TJ, Mcmahan CJ, Tamman J, Page KJ, Trager JB. Prostatic acid phosphatase expression in human tissues. *International journal of clinical and experimental pathology* 4(3), 295-306 (2011).
73. Huang J.G. CN, Goldenberg S.L. Psa and beyond: Biomarkers in prostate cancer. *BC Medical Journal* 55(7), 334-341 (2014).
74. Dijkstra S, Mulders PF, Schalken JA. Clinical use of novel urine and blood based prostate cancer biomarkers: A review. *Clinical biochemistry* 47(10-11), 889-896 (2014).
75. Prensner JR, Rubin MA, Wei JT, Chinnaiyan AM. Beyond psa: The next generation of prostate cancer biomarkers. *Science translational medicine* 4(127), 127rv123 (2012).
76. Prensner JR, Rubin MA, Wei JT, Chinnaiyan AM. Beyond psa: The next generation of prostate cancer biomarkers. *Sci Transl Med* 4(127), (2012).
77. Pettersson A, Graff RE, Bauer SR *et al*. The tmprss2:Erg rearrangement, erg expression, and prostate cancer outcomes: A cohort study and meta-analysis. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 21(9), 1497-1509 (2012).
78. Rubin MA, Bismar TA, Andren O *et al*. Decreased alpha-methylacyl coa racemase expression in localized prostate cancer is associated with an increased rate of biochemical recurrence and cancer-specific death. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 14(6), 1424-1432 (2005).
79. Berger MF, Lawrence MS, Demichelis F *et al*. The genomic complexity of primary human prostate cancer. *Nature* 470(7333), 214-220 (2011).
80. Beretov J, Wasinger VC, Millar EK, Schwartz P, Graham PH, Li Y. Proteomic analysis of urine to identify breast cancer biomarker candidates using a label-free lc-ms/ms approach. *PloS one* 10(11), e0141876 (2015).
81. Xing X, Huang Y, Wang S *et al*. Dataset for the quantitative proteomics analysis of the primary hepatocellular carcinoma with single and multiple lesions. *Data in brief* 5, 226-240 (2015).

82. Matta A, Ralhan R, Desouza LV, Siu KW. Mass spectrometry-based clinical proteomics: Head-and-neck cancer biomarkers and drug-targets discovery. *Mass spectrometry reviews* 29(6), 945-961 (2010).
83. Hudler P, Kocevar N, Komel R. Proteomic approaches in biomarker discovery: New perspectives in cancer diagnostics. *TheScientificWorldJournal* 2014, 260348 (2014).
84. Banez LL, Srivastava S, Moul JW. Proteomics in prostate cancer. *Current opinion in urology* 15(3), 151-156 (2005).
85. Fredolini C, Liotta LA, Petricoin EF. Application of proteomic technologies for prostate cancer detection, prognosis, and tailored therapy. *Critical reviews in clinical laboratory sciences* 47(3), 125-138 (2010).
86. Novelli G, Ciccacci C, Borgiani P, Papaluca Amati M, Abadie E. Genetic tests and genomic biomarkers: Regulation, qualification and validation. *Clinical cases in mineral and bone metabolism : the official journal of the Italian Society of Osteoporosis, Mineral Metabolism, and Skeletal Diseases* 5(2), 149-154 (2008).
87. Hunter DJ, Losina E, Guermazi A, Burstein D, Lassere MN, Kraus V. A pathway and approach to biomarker validation and qualification for osteoarthritis clinical trials. *Current drug targets* 11(5), 536-545 (2010).
88. Pepe MS, Etzioni R, Feng Z *et al*. Phases of biomarker development for early detection of cancer. *Journal of the National Cancer Institute* 93(14), 1054-1061 (2001).
89. Larkin SE, Zeidan B, Taylor MG *et al*. Proteomics in prostate cancer biomarker discovery. *Expert Rev Proteomics* 7(1), 93-102 (2010).
90. Ornstein DK, Tyson DR. Proteomics for the identification of new prostate cancer biomarkers. *Urologic oncology* 24(3), 231-236 (2006).
91. Davalieva K, Polenakovic M. Proteomics in diagnosis of prostate cancer. *Prilozi / Makedonska akademija na naukite i umetnostite, Oddelenie za bioloski i medicinski nauki = Contributions / Macedonian Academy of Sciences and Arts, Section of Biological and Medical Sciences* 36(1), 5-36 (2015).
92. Downes MR, Byrne JC, Dunn MJ, Fitzpatrick JM, Watson RW, Pennington SR. Application of proteomic strategies to the identification of urinary biomarkers for prostate cancer: A review. *Biomarkers : biochemical indicators of exposure, response, and susceptibility to chemicals* 11(5), 406-416 (2006).
93. Pin E, Fredolini C, Petricoin EF, 3rd. The role of proteomics in prostate cancer research: Biomarker discovery and validation. *Clinical biochemistry* 46(6), 524-538 (2013).
94. Davalieva K, Kiprijanovska S, Komina S, Petrusevska G, Zografska NC, Polenakovic M. Proteomics analysis of urine reveals acute phase response proteins as candidate diagnostic biomarkers for prostate cancer. *Proteome science* 13(1), 2 (2015).
95. Lin JF, Xu J, Tian HY *et al*. Identification of candidate prostate cancer biomarkers in prostate needle biopsy specimens using proteomic analysis. *International journal of cancer. Journal international du cancer* 121(12), 2596-2605 (2007).
96. Rizzardi AE, Rosener NK, Koopmeiners JS *et al*. Evaluation of protein biomarkers of prostate cancer aggressiveness. *BMC cancer* 14, 244 (2014).
97. Shipitsin M, Small C, Choudhury S *et al*. Identification of proteomic biomarkers predicting prostate cancer aggressiveness and lethality despite biopsy-sampling error. *British journal of cancer* 111(6), 1201-1212 (2014).
98. Al-Ruwaili JA, Larkin SE, Zeidan BA *et al*. Discovery of serum protein biomarkers for prostate cancer progression by proteomic analysis. *Cancer genomics & proteomics* 7(2), 93-103 (2010).
99. Rolfo C, Castiglia M, Hong D *et al*. Liquid biopsies in lung cancer: The new ambrosia of researchers. *Biochimica et biophysica acta* 1846(2), 539-546 (2014).

100. Francis G, Stein S. Circulating cell-free tumour DNA in the management of cancer. *International journal of molecular sciences* 16(6), 14122-14142 (2015).
101. Tsujiura M, Ichikawa D, Konishi H, Komatsu S, Shiozaki A, Otsuji E. Liquid biopsy of gastric cancer patients: Circulating tumor cells and cell-free nucleic acids. *World journal of gastroenterology : WJG* 20(12), 3265-3286 (2014).
102. Schwarzenbach H. Circulating nucleic acids as biomarkers in breast cancer. *Breast cancer research : BCR* 15(5), 211 (2013).
103. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: Monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol* 10(8), 472-484 (2013).
104. Brock G, Castellanos-Rizaldos E, Hu L, Coticchia C, Skog J. Liquid biopsy for cancer screening, patient stratification and monitoring. *Transl Cancer Res* 4(3), 280-290 (2015).
105. Cree IA. Liquid biopsy for cancer patients: Principles and practice. *Pathogenesis* 2(1-2), 1-4 (2015).
106. Pantel K, Alix-Panabieres C. Real-time liquid biopsy in cancer patients: Fact or fiction? *Cancer Res* 73(21), 6384-6388 (2013).
107. Nannini M, Astolfi A, Urbini M, Biasco G, Pantaleo MA. Liquid biopsy in gastrointestinal stromal tumors: A novel approach. *Journal of translational medicine* 12, 210 (2014).
108. Siravegna G, Bardelli A. Genotyping cell-free tumor DNA in the blood to detect residual disease and drug resistance. *Genome biology* 15(8), 449 (2014).
109. Alix-Panabieres C, Pantel K. Real-time liquid biopsy: Circulating tumor cells versus circulating tumor DNA. *Ann Transl Med* 1(2), 18 (2013).
110. Karachaliou N, Mayo-De-Las-Casas C, Molina-Vila MA, Rosell R. Real-time liquid biopsies become a reality in cancer treatment. *Annals of translational medicine* 3(3), 36 (2015).
111. Heitzer E, Auer M, Ulz P, Geigl JB, Speicher MR. Circulating tumor cells and DNA as liquid biopsies. *Genome medicine* 5(8), 73 (2013).
112. Galletti G, Portella L, Tagawa ST, Kirby BJ, Giannakakou P, Nanus DM. Circulating tumor cells in prostate cancer diagnosis and monitoring: An appraisal of clinical potential. *Molecular diagnosis & therapy* 18(4), 389-402 (2014).
113. Marrinucci D, Bethel K, Kolatkar A *et al*. Fluid biopsy in patients with metastatic prostate, pancreatic and breast cancers. *Physical biology* 9(1), 016003 (2012).
114. Brock G. C-RE, Hu L., Coticchia C., Skog J. Liquid biopsy for cancer screening, patient stratification and monitoring. *Transl Cancer Res* 4(3), 280-290 (2015).
115. Sobhani K. Urine proteomic analysis: Use of two-dimensional gel electrophoresis, isotope coded affinity tags, and capillary electrophoresis. *Methods in molecular biology* 641, 325-346 (2010).
116. Decramer S, Gonzalez De Peredo A, Breuil B *et al*. Urine in clinical proteomics. *Molecular & cellular proteomics : MCP* 7(10), 1850-1862 (2008).
117. Gutierrez G, Reines HD, Wulf-Gutierrez ME. Clinical review: Hemorrhagic shock. *Critical care* 8(5), 373-381 (2004).
118. Pisitkun T, Johnstone R, Knepper MA. Discovery of urinary biomarkers. *Molecular & cellular proteomics : MCP* 5(10), 1760-1771 (2006).
119. Galasko D, Golde TE. Biomarkers for alzheimer's disease in plasma, serum and blood - conceptual and practical problems. *Alzheimer's research & therapy* 5(2), 10 (2013).
120. Goldknopp IL, Sheta E, Bryson J *et al*. Blood serum biomarkers for differential diagnosis of parkinson's disease. *Faseb Journal* 20(4), A64-A64 (2006).
121. Jesneck JL, Mukherjee S, Yurkovetsky Z *et al*. Do serum biomarkers really measure breast cancer? *BMC cancer* 9, (2009).

122. Allen RE, Rogozinska E, Cleverly K, Aquilina J, Thangaratinam S. Abnormal blood biomarkers in early pregnancy are associated with preeclampsia: A meta-analysis. *Eur J Obstet Gyn R B* 182, 194-201 (2014).
123. Reuschenbach M, Von Knebel Doeberitz M, Wentzensen N. A systematic review of humoral immune responses against tumor antigens. *Cancer immunology, immunotherapy : CII* 58(10), 1535-1544 (2009).
124. Thakur A, Littrup P, Paul EN, Adam B, Heilbrun LK, Lum LG. Induction of specific cellular and humoral responses against renal cell carcinoma after combination therapy with cryoablation and granulocyte-macrophage colony stimulating factor: A pilot study. *Journal of immunotherapy* 34(5), 457-467 (2011).
125. Ullenhag GJ, Frodin JE, Jeddi-Tehrani M *et al*. Durable carcinoembryonic antigen (cea)-specific humoral and cellular immune responses in colorectal carcinoma patients vaccinated with recombinant cea and granulocyte/macrophage colony-stimulating factor. *Clinical cancer research : an official journal of the American Association for Cancer Research* 10(10), 3273-3281 (2004).
126. Michils A, Yernault JC, Noel E, Gossart B, Servais G, Duchateau J. Abnormal humoral immune response to mucosal antigenic stimulation in patients with lung cancer. *Cancer* 69(9), 2252-2257 (1992).
127. Fossa A, Berner A, Fossa SD, Hernes E, Gaudernack G, Smeland EB. Ny-eso-1 protein expression and humoral immune responses in prostate cancer. *The Prostate* 59(4), 440-447 (2004).
128. Sreekumar A, Laxman B, Rhodes DR *et al*. Humoral immune response to alpha-methylacyl-coa racemase and prostate cancer. *Journal of the National Cancer Institute* 96(11), 834-843 (2004).
129. Reis BS, Jungbluth AA, Frosina D *et al*. Prostate cancer progression correlates with increased humoral immune response to a human endogenous retrovirus gag protein. *Clinical cancer research : an official journal of the American Association for Cancer Research* 19(22), 6112-6125 (2013).
130. Pontes ER, Matos LC, Da Silva EA *et al*. Auto-antibodies in prostate cancer: Humoral immune response to antigenic determinants coded by the differentially expressed transcripts flj23438 and vamp3. *The Prostate* 66(14), 1463-1473 (2006).
131. Todorova K, Ignatova I, Tchakarov S *et al*. Humoral immune response in prostate cancer patients after immunization with gene-based vaccines that encode for a protein that is proteasomally degraded. *Cancer immunity* 5, 1 (2005).
132. Knaul FM, Atun R, Farmer P, Frenk J. Seizing the opportunity to close the cancer divide. *Lancet* 381(9885), 2238-2239 (2013).
133. Roberts R. Bridging the urological divide. *Infectious agents and cancer* 6 Suppl 2, S4 (2011).
134. Zeigler-Johnson CM, Spangler E, Jalloh M, Gueye SM, Rennert H, Rebbeck TR. Genetic susceptibility to prostate cancer in men of african descent: Implications for global disparities in incidence and outcomes. *The Canadian journal of urology* 15(1), 3872-3882 (2008).
135. Odedina FT, Dagne G, Pressey S *et al*. Prostate cancer health and cultural beliefs of black men: The florida prostate cancer disparity project. *Infectious agents and cancer* 6 Suppl 2, S10 (2011).
136. Gakunga R, Parkin DM, African Cancer Registry N. Cancer registries in africa 2014: A survey of operational features and uses in cancer control planning. *International journal of cancer. Journal international du cancer* 137(9), 2045-2052 (2015).
137. Agalliu I, Adebiyi AO, Lounsbury DW *et al*. The feasibility of epidemiological research on prostate cancer in african men in ibadan, nigeria. *BMC public health* 15, 425 (2015).
138. Olapade-Olaopa EO, Obamuyide HA, Yisa GT. Management of advanced prostate cancer in africa. *The Canadian journal of urology* 15(1), 3890-3898 (2008).

139. Sfanos KS, De Marzo AM. Prostate cancer and inflammation: The evidence. *Histopathology* 60(1), 199-215 (2012).
140. Huang WY, Hayes R, Pfeiffer R *et al*. Sexually transmissible infections and prostate cancer risk. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 17(9), 2374-2381 (2008).
141. Adewole I, Martin DN, Williams MJ *et al*. Building capacity for sustainable research programmes for cancer in africa. *Nature reviews. Clinical oncology* 11(5), 251-259 (2014).
142. Dirks PH, Berger LR, Roberts EM *et al*. Geological and taphonomic context for the new hominin species homo naledi from the dinaledi chamber, south africa. *eLife* 4, (2015).
143. Val A, Dirks PH, Backwell LR, D'errico F, Berger LR. Taphonomic analysis of the faunal assemblage associated with the hominins (australopithecus sediba) from the early pleistocene cave deposits of malapa, south africa. *PloS one* 10(6), e0126904 (2015).
144. Crous PW, Groenewald JZ, Lombard L, Wingfield MJ. Homortomyces gen. Nov., a new dothidealean pycnidial fungus from the cradle of humankind. *IMA fungus* 3(2), 109-115 (2012).
145. Berger LR, Hawks J, De Ruiter DJ *et al*. Homo naledi, a new species of the genus homo from the dinaledi chamber, south africa. *eLife* 4, (2015).
146. Bailey DH, Hamilton MJ, Walker RS. Latitude, population size, and the language-farming dispersal hypothesis. *Evol Ecol Res* 14(8), 1057-1067 (2012).
147. Diamond J, Bellwood P. Farmers and their languages: The first expansions. *Science* 300(5619), 597-603 (2003).
148. Odedina FT, Akinremi TO, Chinegwundoh F *et al*. Prostate cancer disparities in black men of african descent: A comparative literature review of prostate cancer burden among black men in the united states, caribbean, united kingdom, and west africa. *Infectious agents and cancer* 4 Suppl 1, S2 (2009).
149. Zakharia F, Basu A, Absher D *et al*. Characterizing the admixed african ancestry of african americans. *Genome biology* 10(12), R141 (2009).
150. Heyns CF, Fisher M, Lecuona A, Van Der Merwe A. Prostate cancer among different racial groups in the western cape: Presenting features and management. *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde* 101(4), 267-270 (2011).
151. Tindall EA, Monare LR, Petersen DC *et al*. Clinical presentation of prostate cancer in black south africans. *The Prostate* 74(8), 880-891 (2014).
152. Gonzalez-Santos M, Montinaro F, Oosthuizen O *et al*. Genome-wide snp analysis of southern african populations provides new insights into the dispersal of bantu-speaking groups. *Genome biology and evolution* 7(9), 2560-2568 (2015).
153. Tishkoff SA, Reed FA, Friedlaender FR *et al*. The genetic structure and history of africans and african americans. *Science* 324(5930), 1035-1044 (2009).
154. Van Den Berghe PL. *Human family systems. An evolutionary view*. Elsevier, New York. (1979).
155. Pebley AR, And Mbugua, W. . Polygyny and fertility in sub-saharan africa. In: *Reproduction and social organization in sub-saharan africa*, Lesthaeghe RJ (Ed.^(Eds). University of California Press, Berkeley 338-364 (1989).
156. Darwin C, Wallace, Ar. "On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection". *Journal of the Proceedings of the Linnean Society of London. Zoology* 3, 46–50 (1858).
157. Ama PF, Simoneau JA, Boulay MR, Serresse O, Theriault G, Bouchard C. Skeletal muscle characteristics in sedentary black and caucasian males. *Journal of applied physiology* 61(5), 1758-1761 (1986).

158. Ettinger B, Sidney S, Cummings SR *et al*. Racial differences in bone density between young adult black and white subjects persist after adjustment for anthropometric, lifestyle, and biochemical differences. *The Journal of clinical endocrinology and metabolism* 82(2), 429-434 (1997).
159. Hui SL, Dimeglio LA, Longcope C *et al*. Difference in bone mass between black and white american children: Attributable to body build, sex hormone levels, or bone turnover? *J Clin Endocr Metab* 88(2), 642-649 (2003).
160. Wagner DR, Heyward VH. Measures of body composition in blacks and whites: A comparative review. *The American journal of clinical nutrition* 71(6), 1392-1402 (2000).
161. Schillaci MA. Sexual selection and the evolution of brain size in primates. *PLoS One* 1, e62 (2006).
162. Stanyon R, Bigoni F. Sexual selection and the evolution of behavior, morphology, neuroanatomy and genes in humans and other primates. *Neurosci Biobehav Rev* 46P4, 579-590 (2014).
163. Ball GF, Balthazart J, Mccarthy MM. Is it useful to view the brain as a secondary sexual characteristic? *Neurosci Biobehav R* 46, 628-638 (2014).
164. Verweij KJH, Burri AV, Zietsch BP. Testing the prediction from sexual selection of a positive genetic correlation between human mate preferences and corresponding traits. *Evol Hum Behav* 35(6), 497-501 (2014).
165. Berg V, Lummaa V, Lahdenpera M, Rotkirch A, Jokela M. Personality and long-term reproductive success measured by the number of grandchildren. *Evol Hum Behav* 35(6), 533-539 (2014).
166. Adamo SA, Spiteri RJ. He's healthy, but will he survive the plague? Possible constraints on mate choice for disease resistance. *Anim Behav* 77(1), 67-78 (2009).
167. Wu DWL, Bischof WF, Kingstone A. Natural gaze signaling in a social context. *Evol Hum Behav* 35(3), 211-218 (2014).
168. Hahn AC, Perrett DI. Neural and behavioral responses to attractiveness in adult and infant faces. *Neurosci Biobehav R* 46, 591-603 (2014).
169. Gillies GE, Mcarthur S. Estrogen actions in the brain and the basis for differential action in men and women: A case for sex-specific medicines. *Pharmacol Rev* 62(2), 155-198 (2010).
170. De Vries GJ, Sodersten P. Sex differences in the brain: The relation between structure and function. *Horm Behav* 55(5), 589-596 (2009).
171. Hines M. Sex-related variation in human behavior and the brain. *Trends in Cognitive Sciences* 14(10), 448-456 (2010).
172. Winkler EM, Christiansen K. Sex-hormone levels and body hair-growth in kung-san and kavango men from namibia. *Am J Phys Anthropol* 92(2), 155-164 (1993).
173. Bribiescas RG. Testosterone levels among ache hunter-gatherer men - a functional interpretation of population variation among adult males. *Hum Nature-Int Bios* 7(2), 163-188 (1996).
174. Trumble BC, Cummings DK, O'connor KA *et al*. Age-independent increases in male salivary testosterone during among tsimane forager-farmers. *Evol Hum Behav* 34(5), 350-357 (2013).
175. Alvarado LC. Do evolutionary life-history trade-offs influence prostate cancer risk? A review of population variation in testosterone levels and prostate cancer disparities. *Evol Appl* 6(1), 117-133 (2013).
176. Pettaway CA. Racial differences in the androgen/androgen receptor pathway in prostate cancer. *Journal of the National Medical Association* 91(12), 653-660 (1999).
177. Ross RK, Bernstein L, Lobo RA *et al*. 5-alpha-reductase activity and risk of prostate-cancer among japanese and united-states white and black-males. *Lancet* 339(8798), 887-889 (1992).
178. Kittles RA, Young D, Weinrich S *et al*. Extent of linkage disequilibrium between the androgen receptor gene cag and ggc repeats in human populations: Implications for prostate cancer risk. *Human genetics* 109(3), 253-261 (2001).

179. Autio KA, Morris MJ. Targeting bone physiology for the treatment of metastatic prostate cancer. *Clin Adv Hematol Oncol* 11(3), 134-143 (2013).
180. Shapiro LJ. Human genome project. *The Western journal of medicine* 158(2), 181 (1993).
181. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA *et al*. The cancer genome atlas pan-cancer analysis project. *Nature genetics* 45(10), 1113-1120 (2013).
182. Legrain P, Aebersold R, Archakov A *et al*. The human proteome project: Current state and future direction. *Molecular & cellular proteomics : MCP* 10(7), M111 009993 (2011).
183. Ponomarenko E, Baranova A, Lisitsa A, Albar JP, Archakov A. The chromosome-centric human proteome project at febs congress. *Proteomics* 14(2-3), 147-152 (2014).
184. Vaught J, Rogers J, Myers K *et al*. An nci perspective on creating sustainable biospecimen resources. *Journal of the National Cancer Institute. Monographs* 2011(42), 1-7 (2011).
185. Bledsoe MJ, Grizzle WE. Use of human specimens in research: The evolving united states regulatory, policy, and scientific landscape. *Diagnostic histopathology* 19(9), 322-330 (2013).
186. Vaught JB, Henderson MK, Compton CC. Biospecimens and biorepositories: From afterthought to science. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 21(2), 253-255 (2012).
187. Ollier W, Sprosen T, Peakman T. Uk biobank: From concept to reality. *Pharmacogenomics* 6(6), 639-646 (2005).
188. Rush A, Christiansen JH, Farrell JP *et al*. Biobank classification in an australian setting. *Biopreservation and biobanking* 13(3), 212-218 (2015).
189. Gasmelseed N, Elsir AA, Deblasio P, Biunno I. Sub-saharan centralized biorepository for genetic and genomic research. *The Science of the total environment* 423, 210-213 (2012).
190. Staunton C, Moodley K. Challenges in biobank governance in sub-saharan africa. *BMC medical ethics* 14, 35 (2013).
191. De Vries J, Abayomi A, Brandful J *et al*. A perpetual source of DNA or something really different: Ethical issues in the creation of cell lines for african genomics research. *BMC medical ethics* 15, 60 (2014).
192. Abayomi A, Christoffels A, Grewal R *et al*. Challenges of biobanking in south africa to facilitate indigenous research in an environment burdened with human immunodeficiency virus, tuberculosis, and emerging noncommunicable diseases. *Biopreservation and biobanking* 11(6), 347-354 (2013).
193. Powell IJ. The precise role of ethnicity and family history on aggressive prostate cancer: A review analysis. *Arch Esp Urol* 64(8), 711-719 (2011).
194. Freeman VL, Durazo-Arvizu R, Keys LC, Johnson MP, Schafernak K, Patel VK. Racial differences in survival among men with prostate cancer and comorbidity at time of diagnosis. *Am J Public Health* 94(5), 803-808 (2004).
195. Roberts R. From bench to bedside: The realities of reducing global prostate cancer disparity in black men. *Ecancermedicalscience* 8, 458 (2014).
196. World Medical A. World medical association declaration of helsinki: Ethical principles for medical research involving human subjects. *JAMA : the journal of the American Medical Association* 310(20), 2191-2194 (2013).
197. Williams JR. The declaration of helsinki and public health. *Bulletin of the World Health Organization* 86(8), 650-652 (2008).
198. Carlson RV, Boyd KM, Webb DJ. The revision of the declaration of helsinki: Past, present and future. *Br J Clin Pharmacol* 57(6), 695-713 (2004).
199. Keslar KS, Lin M, Zmijewska AA *et al*. Multicenter evaluation of a standardized protocol for noninvasive gene expression profiling. *Am J Transplant* 13(7), 1891-1897 (2013).

200. Overbye A, Skotland T, Koehler CJ *et al*. Identification of prostate cancer biomarkers in urinary exosomes. *Oncotarget* 6(30), 30357-30376 (2015).
201. Crichton PG, Harding M, Ruprecht JJ, Lee Y, Kunji ER. Lipid, detergent, and coomassie blue g-250 affect the migration of small membrane proteins in blue native gels: Mitochondrial carriers migrate as monomers not dimers. *J Biol Chem* 288(30), 22163-22173 (2013).
202. Zhang Q, Ames JM, Smith RD, Baynes JW, Metz TO. A perspective on the maillard reaction and the analysis of protein glycation by mass spectrometry: Probing the pathogenesis of chronic disease. *J Proteome Res* 8(2), 754-769 (2009).
203. Berkelman T. Removal of interfering substances in samples prepared for two-dimensional (2-d) electrophoresis. *Methods Mol Biol* 424, 51-62 (2008).
204. Zienkiewicz A, Rejon JD, De Dios Alche J, Rodriguez-Garcia MI, Castro AJ. A protocol for protein extraction from lipid-rich plant tissues suitable for electrophoresis. *Methods Mol Biol* 1072, 85-91 (2014).
205. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* 72, 248-254 (1976).
206. Compton SJ, Jones CG. Mechanism of dye response and interference in the bradford protein assay. *Analytical biochemistry* 151(2), 369-374 (1985).
207. Sedmak JJ, Grossberg SE. A rapid, sensitive, and versatile assay for protein using coomassie brilliant blue g250. *Analytical biochemistry* 79(1-2), 544-552 (1977).
208. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nature methods* 6(5), 359-362 (2009).
209. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using stagetips. *Nat Protoc* 2(8), 1896-1906 (2007).
210. A GM, G TB, Chen T *et al*. Mass resolution and mass accuracy: How much is enough? *Mass spectrometry* 2(Spec Iss), S0009 (2013).
211. Jorabchi K, Smith LM. Single droplet separations and surface partition coefficient measurements using laser ablation mass spectrometry. *Analytical chemistry* 81(23), 9682-9688 (2009).
212. Wang PH, Wilson SR. Mass spectrometry-based protein identification by integrating de novo sequencing with database searching. *Bmc Bioinformatics* 14, (2013).
213. Frank AM. A ranking-based scoring function for peptide-spectrum matches. *Journal of Proteome Research* 8(5), 2241-2252 (2009).
214. Mchugh L, Arthur JW. Computational methods for protein identification from mass spectrometry data. *Plos Comput Biol* 4(2), (2008).
215. Kall L, Storey JD, Maccoss MJ, Noble WS. Posterior error probabilities and false discovery rates: Two sides of the same coin. *J Proteome Res* 7(1), 40-44 (2008).
216. Kall L, Storey JD, Noble WS. Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinformatics* 24(16), i42-48 (2008).
217. Choi H, Nesvizhskii AI. False discovery rates and related statistical concepts in mass spectrometry-based proteomics. *J Proteome Res* 7(1), 47-50 (2008).
218. Noble WS. How does multiple testing correction work? *Nature Biotechnology* 27(12), 1135-1137 (2009).
219. Gupta N, Pevzner PA. False discovery rates of protein identifications: A strike against the two-peptide rule. *Journal of Proteome Research* 8(9), 4173-4181 (2009).
220. Cooks RG, Mueller T. Through a glass darkly: Glimpses into the future of mass spectrometry. *Mass spectrometry* 2(Spec Iss), S0001 (2013).

221. Gies AP, Vergne MJ, Orndorff RL, Hercules DM. Maldi-tof/tof cid study of 4-alkyl-substituted polystyrene fragmentation reactions. *Analytical and bioanalytical chemistry* 392(4), 609-626 (2008).
222. Boersema PJ, Mohammed S, Heck AJ. Phosphopeptide fragmentation and analysis by mass spectrometry. *Journal of mass spectrometry : JMS* 44(6), 861-878 (2009).
223. Li W, Song C, Bailey DJ, Tseng GC, Coon JJ, Wysocki VH. Statistical analysis of electron transfer dissociation pairwise fragmentation patterns. *Analytical chemistry* 83(24), 9540-9545 (2011).
224. Washburn MP, Wolters D, Yates JR, 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19(3), 242-247 (2001).
225. Wolters DA, Washburn MP, Yates JR, 3rd. An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* 73(23), 5683-5690 (2001).
226. Kelleher NL. Top-down proteomics. *Anal Chem* 76(11), 197A-203A (2004).
227. Ong SE, Blagoev B, Kratchmarova I *et al*. Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1(5), 376-386 (2002).
228. Ross PL, Huang YN, Marchese JN *et al*. Multiplexed protein quantitation in saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3(12), 1154-1169 (2004).
229. Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJ. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* 4(4), 484-494 (2009).
230. Thompson A, Schafer J, Kuhn K *et al*. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by ms/ms. *Anal Chem* 75(8), 1895-1904 (2003).
231. Michalski A, Damoc E, Hauschild JP *et al*. Mass spectrometry-based proteomics using q exactive, a high-performance benchtop quadrupole orbitrap mass spectrometer. *Mol Cell Proteomics* 10(9), M111 011015 (2011).
232. Liebler DC, Zimmerman LJ. Targeted quantitation of proteins by mass spectrometry. *Biochemistry* 52(22), 3797-3806 (2013).
233. Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: A tutorial. *Mol Syst Biol* 4, 222 (2008).
234. Nahnsen S, Kohlbacher O. In silico design of targeted srm-based experiments. *BMC bioinformatics* 13 Suppl 16, S8 (2012).
235. Deutsch E.W. CDS, Picotti P., Mendoza L., Sun Z., Farrah T., Bletz J., Kusebauch U., Brusniak M., Lam H., Aebersold R., Moritzr.L. Srmatlas: Generating targeted proteomics transition atlases for complete proteomes. *In Preparation*,
236. Farrah T, Deutsch EW, Kreisberg R *et al*. Passel: The peptideatlas srmexperiment library. *Proteomics* 12(8), 1170-1175 (2012).
237. Brusniak MY, Kwok ST, Christiansen M *et al*. Ataqs: A computational software tool for high throughput transition optimization and validation for selected reaction monitoring mass spectrometry. *BMC Bioinformatics* 12, 78 (2011).
238. Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Molecular & cellular proteomics : MCP* 11(11), 1475-1488 (2012).
239. Maclean B, Tomazela DM, Shulman N *et al*. Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26(7), 966-968 (2010).
240. Egertson JD, Kuehn A, Merrihew GE *et al*. Multiplexed ms/ms for improved data-independent acquisition. *Nat Methods* 10(8), 744-746 (2013).

241. Diamandis EP, Christopoulos TK. The biotin-(strept)avidin system: Principles and applications in biotechnology. *Clinical chemistry* 37(5), 625-636 (1991).
242. Holmberg A, Blomstergren A, Nord O, Lukacs M, Lundeberg J, Uhlen M. The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis* 26(3), 501-510 (2005).
243. Tong X, Smith LM. Solid-phase method for the purification of DNA sequencing reactions. *Analytical chemistry* 64(22), 2672-2677 (1992).
244. Del Rosario RB, Baron LA, Lawton RG, Wahl RL. Streptavidin-biotinylated igg conjugates: A simple procedure for reducing polymer formation. *International journal of radiation applications and instrumentation. Part B, Nuclear medicine and biology* 19(3), 417-421 (1992).
245. Chivers CE, Koner AL, Lowe ED, Howarth M. How the biotin-streptavidin interaction was made even stronger: Investigation via crystallography and a chimaeric tetramer. *The Biochemical journal* 435(1), 55-63 (2011).
246. Gonzalez M, Bagatolli LA, Echabe I *et al*. Interaction of biotin with streptavidin. Thermostability and conformational changes upon binding. *The Journal of biological chemistry* 272(17), 11288-11294 (1997).
247. Kodadek T. Protein microarrays: Prospects and problems. *Chemistry & biology* 8(2), 105-115 (2001).
248. Zhu H, Qian J. Applications of functional protein microarrays in basic and clinical research. *Advances in genetics* 79, 123-155 (2012).
249. O'kane SL, O'brien JK, Cahill DJ. Optimized autoantibody profiling on protein arrays. *Methods in molecular biology* 785, 331-341 (2011).
250. Sutandy FX, Qian J, Chen CS, Zhu H. Overview of protein microarrays. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* Chapter 27, Unit 27 21 (2013).
251. Sjoberg R, Mattsson C, Andersson E *et al*. Exploration of high-density protein microarrays for antibody validation and autoimmunity profiling. *New biotechnology*, (2015).
252. Gilbert I, Schiffmann S, Rubenwolf S *et al*. Double chip protein arrays using recombinant single-chain fv antibody fragments. *Proteomics* 4(5), 1417-1420 (2004).
253. Im H, Snyder M. Preparation of recombinant protein spotted arrays for proteome-wide identification of kinase targets. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* Chapter 27, Unit 27 24 (2013).
254. Miyaji T, Hewitt SM, Liotta LA, Star RA. Frozen protein arrays: A new method for arraying and detecting recombinant and native tissue proteins. *Proteomics* 2(11), 1489-1493 (2002).
255. Caiazzo RJ, Jr., O'rourke DJ, Barder TJ, Nelson BP, Liu BC. Native antigen fractionation protein microarrays for biomarker discovery. *Methods in molecular biology* 723, 129-148 (2011).
256. Ehrlich JR, Qin S, Liu BC. The 'reverse capture' autoantibody microarray: A native antigen-based platform for autoantibody profiling. *Nature protocols* 1(1), 452-460 (2006).
257. Liu BC, Dijohnson DA, O'rourke DJ. Antibody profiling with protein antigen microarrays in early stage cancer. *Expert opinion on medical diagnostics* 6(3), 187-196 (2012).
258. Eickhoff H, Konthur Z, Lueking A *et al*. Protein array technology: The tool to bridge genomics and proteomics. *Advances in biochemical engineering/biotechnology* 77, 103-112 (2002).
259. Bussow K, Quedenau C, Sievert V *et al*. A catalog of human cdna expression clones and its application to structural genomics. *Genome biology* 5(9), R71 (2004).
260. Holt LJ, Bussow K, Walter G, Tomlinson IM. By-passing selection: Direct screening for antibody-antigen interactions using protein arrays. *Nucleic acids research* 28(15), E72 (2000).
261. Walter G, Bussow K, Lueking A, Glokler J. High-throughput protein arrays: Prospects for molecular diagnostics. *Trends in molecular medicine* 8(6), 250-253 (2002).

262. Holz C, Lueking A, Bovekamp L *et al*. A human cdna expression library in yeast enriched for open reading frames. *Genome research* 11(10), 1730-1735 (2001).
263. Gutjahr C, Murphy D, Lueking A *et al*. Mouse protein arrays from a th1 cell cdna library for antibody screening and serum profiling. *Genomics* 85(3), 285-296 (2005).
264. Bussow K, Nordhoff E, Lubbert C, Lehrach H, Walter G. A human cdna library for high-throughput protein expression screening. *Genomics* 65(1), 1-8 (2000).
265. Weiner H, Faupel T, Bussow K. Protein arrays from cdna expression libraries. *Methods in molecular biology* 264, 1-13 (2004).
266. Qiu J, Labaer J. Nucleic acid programmable protein array a just-in-time multiplexed protein expression and purification platform. *Methods in enzymology* 500, 151-163 (2011).
267. Link AJ, Labaer J. Construction of nucleic acid programmable protein arrays (nappa) 1: Coating glass slides with amino silane. *CSH protocols* 2008, pdb prot5056 (2008).
268. Sahin U, Tureci O. Antigen identification using serex. *Methods in molecular biology* 1061, 59-77 (2013).
269. Nand A, Gautam A, Perez JB, Merino A, Zhu J. Emerging technology of in situ cell free expression protein microarrays. *Protein & cell* 3(2), 84-88 (2012).
270. He M, Taussig MJ. Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (pisa method). *Nucleic acids research* 29(15), E73-73 (2001).
271. He M, Stoevesandt O, Taussig MJ. In situ synthesis of protein arrays. *Current opinion in biotechnology* 19(1), 4-9 (2008).
272. Angenendt P, Kreutzberger J, Glokler J, Hoheisel JD. Generation of high density protein microarrays by cell-free in situ expression of unpurified pcr products. *Molecular & cellular proteomics : MCP* 5(9), 1658-1666 (2006).
273. Kwon K, Grose C, Pieper R, Pandya GA, Fleischmann RD, Peterson SN. High quality protein microarray using in situ protein purification. *BMC biotechnology* 9, 72 (2009).
274. Zarate X, Galbraith DW. A cell-free expression platform for production of protein microarrays. *Methods in molecular biology* 1118, 297-307 (2014).
275. Kitts PA, Ayres MD, Possee RD. Linearization of baculovirus DNA enhances the recovery of recombinant virus expression vectors. *Nucleic acids research* 18(19), 5667-5672 (1990).
276. Zhao Y, Chapman DA, Jones IM. Improving baculovirus recombination. *Nucleic acids research* 31(2), E6-6 (2003).
277. Beeton-Kempen N, Duarte J, Shoko A *et al*. Development of a novel, quantitative protein microarray platform for the multiplexed serological analysis of autoantibodies to cancer-testis antigens. *International journal of cancer. Journal international du cancer* 135(8), 1842-1851 (2014).
278. Duarte J, Serufuri, J.,Bmulder, N., Blackburn, J. Protein function microarrays: Design, use and bioinformatic analysis in cancer biomarker discovery and quantitation. In: *Bioinformatics of human proteomics*, Wang X (Ed.^(Eds). Springer, Dordrecht 39-74 (2013).
279. Espina V, Mehta AI, Winters ME *et al*. Protein microarrays: Molecular profiling technologies for clinical specimens. *Proteomics* 3(11), 2091-2100 (2003).
280. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19(2), 185-193 (2003).
281. Quackenbush J. Microarray data normalization and transformation. *Nature genetics* 32 Suppl, 496-501 (2002).
282. Warde-Farley D, Donaldson SL, Comes O *et al*. The genemania prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 38(Web Server issue), W214-220 (2010).

283. Szklarczyk D, Franceschini A, Kuhn M *et al*. The string database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 39(Database issue), D561-568 (2011).
284. Von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. String: A database of predicted functional associations between proteins. *Nucleic acids research* 31(1), 258-261 (2003).
285. Franceschini A, Szklarczyk D, Frankild S *et al*. String v9.1: Protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 41(Database issue), D808-815 (2013).
286. Szklarczyk D, Franceschini A, Wyder S *et al*. String v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43(Database issue), D447-452 (2015).
287. Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. *Science* 278(5338), 631-637 (1997).
288. Cox J, Mann M. Maxquant enables high peptide identification rates, individualized p.P.B.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* 26(12), 1367-1372 (2008).
289. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: A peptide search engine integrated into the maxquant environment. *Journal of proteome research* 10(4), 1794-1805 (2011).
290. Fertig EJ, Ding J, Favorov AV, Parmigiani G, Ochs MF. Cogaps: An r/c++ package to identify patterns and biological process activity in transcriptomic data. *Bioinformatics* 26(21), 2792-2793 (2010).
291. Stephens MA. Edf statistics for goodness of fit and some comparisons. *J Am Stat Assoc* 69(347), 730-737 (1974).
292. Batterham AM, Atkinson G. How big does my sample need to be? A primer on the murky world of sample size estimation. *Phys Ther Sport* 6(3), 153-163 (2005).
293. Faul F, Erdfelder E, Lang AG, Buchner A. G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39(2), 175-191 (2007).
294. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using g*power 3.1: Tests for correlation and regression analyses. *Behav Res Methods* 41(4), 1149-1160 (2009).
295. Oliveros J. Venny. An interactive tool for comparing lists with venn diagrams. *BioinfoGP, CNB-CSIC*,
296. Ferlay J, Soerjomataram I, Dikshit R *et al*. Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012. *International journal of cancer. Journal international du cancer*, (2014).
297. Wu I, Modlin CS. Disparities in prostate cancer in african american men: What primary care physicians can do. *Cleveland Clinic journal of medicine* 79(5), 313-320 (2012).
298. Mcintosh H. Why do african-american men suffer more prostate cancer? *Journal of the National Cancer Institute* 89(3), 188-189 (1997).
299. Culig Z. Distinguishing indolent from aggressive prostate cancer. *Recent results in cancer research. Fortschritte der Krebsforschung. Progres dans les recherches sur le cancer* 202, 141-147 (2014).
300. Schiffer E. Biomarkers for prostate cancer. *World journal of urology* 25(6), 557-562 (2007).
301. Theodorescu D, Schiffer E, Bauer HW *et al*. Discovery and validation of urinary biomarkers for prostate cancer. *Proteomics. Clinical applications* 2(4), 556-570 (2008).
302. You J, Cozzi P, Walsh B *et al*. Innovative biomarkers for prostate cancer early diagnosis and progression. *Critical reviews in oncology/hematology* 73(1), 10-22 (2010).
303. Han ZD, Zhang YQ, He HC *et al*. Identification of novel serological tumor markers for human prostate cancer using integrative transcriptome and proteome analysis. *Medical oncology* 29(4), 2877-2888 (2012).

304. Emiliozzi P, Corsetti A, Tassi B, Federico G, Martini M, Pansadoro V. Best approach for prostate cancer detection: A prospective study on transperineal versus transrectal six-core prostate biopsy. *Urology* 61(5), 961-966 (2003).
305. Fortner JG. Inadvertent spread of cancer at surgery. *Journal of surgical oncology* 53(3), 191-196 (1993).
306. Roddam AW, Price CP, Allen NE, Ward AM. Assessing the clinical impact of prostate-specific antigen assay variability and nonequimolarity: A simulation study based on the population of the united kingdom. *Clinical chemistry* 50(6), 1012-1016 (2004).
307. Maattanen L, Hakama M, Tammela TL *et al*. Specificity of serum prostate-specific antigen determination in the finnish prostate cancer screening trial. *British journal of cancer* 96(1), 56-60 (2007).
308. Downes MR, Byrne JC, Pennington SR, Dunn MJ, Fitzpatrick JM, Watson RW. Urinary markers for prostate cancer. *BJU international* 99(2), 263-268 (2007).
309. Steen H, Mann M. The abc's (and xyz's) of peptide sequencing. *Nature reviews. Molecular cell biology* 5(9), 699-711 (2004).
310. Meissner F, Mann M. Quantitative shotgun proteomics: Considerations for a high-quality workflow in immunology. *Nature immunology* 15(2), 112-117 (2014).
311. Schiffer E, Vlahou A, Petrolekas A *et al*. Prediction of muscle-invasive bladder cancer using urinary proteomics. *Clinical cancer research : an official journal of the American Association for Cancer Research* 15(15), 4935-4943 (2009).
312. Yang N, Feng S, Shedden K *et al*. Urinary glycoprotein biomarker discovery for bladder cancer detection using lc/ms-ms and label-free quantification. *Clinical cancer research : an official journal of the American Association for Cancer Research* 17(10), 3349-3359 (2011).
313. Young BL, Mlamla Z, Gqamana PP *et al*. The identification of tuberculosis biomarkers in human urine samples. *The European respiratory journal* 43(6), 1719-1729 (2014).
314. Huttenhain R, Soste M, Selevsek N *et al*. Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Science translational medicine* 4(142), 142ra194 (2012).
315. Vermassen T, Van Praet C, Vanderschaeghe D *et al*. Capillary electrophoresis of urinary prostate glycoproteins assists in the diagnosis of prostate cancer. *Electrophoresis* 35(7), 1017-1024 (2014).
316. Vermassen T, Van Praet C, Lumen N *et al*. Urinary prostate protein glycosylation profiling as a diagnostic biomarker for prostate cancer. *The Prostate* 75(3), 314-322 (2015).
317. Principe S, Kim Y, Fontana S *et al*. Identification of prostate-enriched proteins by in-depth proteomic analyses of expressed prostatic secretions in urine. *Journal of proteome research* 11(4), 2386-2396 (2012).
318. Brown MJ. Hypertension and ethnic group. *Bmj* 332(7545), 833-836 (2006).
319. Adeola HA, Soares NC, Paccez JD, Kaestner L, Blackburn JM, Zerbini LF. Discovery of novel candidate urinary protein biomarkers for prostate cancer in a multiethnic cohort of south african patients via label-free mass spectrometry. *Proteomics. Clinical applications* 9(5-6), 597-609 (2015).
320. Karpievitch YV, Dabney AR, Smith RD. Normalization and missing value imputation for label-free lc-ms analysis. *BMC bioinformatics* 13, (2012).
321. Adachi J, Kumar C, Zhang Y, Olsen JV, Mann M. The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome biology* 7(9), R80 (2006).
322. Marimuthu A, O'meally RN, Chaerkady R *et al*. A comprehensive map of the human urinary proteome. *Journal of proteome research* 10(6), 2734-2743 (2011).

323. Wilhelm M, Schlegl J, Hahne H *et al*. Mass-spectrometry-based draft of the human proteome. *Nature* 509(7502), 582-+ (2014).
324. Schwanhausser B, Busse D, Li N *et al*. Global quantification of mammalian gene expression control. *Nature* 473(7347), 337-342 (2011).
325. Guan X. Cancer metastases: Challenges and opportunities. *Acta Pharmaceutica Sinica B* 5(5), 402–418 (2015).
326. Ziobar BL, Kramer, R.H. Adhesion receptors in oral cancer invasion. In: *Head and neck cancer: Emerging perspectives*, Ensley JF, Gutkind J.S., Jacobs, J.R., Lippman, S.M. (Ed.^(Eds). Elsevier Science, USA 66-67 (2003).
327. Kentsis A. Challenges and opportunities for discovery of disease biomarkers using urine proteomics. *Pediatrics international : official journal of the Japan Pediatric Society* 53(1), 1-6 (2011).
328. States DJ, Omenn GS, Blackwell TW *et al*. Challenges in deriving high-confidence protein identifications from data gathered by a hupo plasma proteome collaborative study. *Nature biotechnology* 24(3), 333-338 (2006).
329. Oberg AL, Vitek O. Statistical design of quantitative mass spectrometry-based proteomic experiments. *Journal of proteome research* 8(5), 2144-2156 (2009).
330. Cairns DA, Barrett JH, Billingham LJ *et al*. Sample size determination in clinical proteomic profiling experiments using mass spectrometry for class comparison. *Proteomics* 9(1), 74-86 (2009).
331. Nagaraj N, Mann M. Quantitative analysis of the intra- and inter-individual variability of the normal urinary proteome. *Journal of proteome research* 10(2), 637-645 (2011).
332. Molina L, Salvetat N, Ben Ameur R *et al*. Analysis of the variability of human normal urine by 2d-ge reveals a "public" and a "private" proteome. *J Proteomics* 75(1), 70-80 (2011).
333. Khan A, Packer NH. Simple urinary sample preparation for proteomic analysis. *Journal of proteome research* 5(10), 2824-2838 (2006).
334. Thomas CE, Sexton W, Benson K, Sutphen R, Koomen J. Urine collection and processing for protein biomarker discovery and quantification. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 19(4), 953-959 (2010).
335. Karpievitch YV, Taverner T, Adkins JN *et al*. Normalization of peak intensities in bottom-up ms-based proteomics using singular value decomposition. *Bioinformatics* 25(19), 2573-2580 (2009).
336. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics* 8(1), 118-127 (2007).
337. Griffin NM, Yu JY, Long F *et al*. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nature biotechnology* 28(1), 83-U116 (2010).
338. Schirle M, Heurtier MA, Kuster B. Profiling core proteomes of human cell lines by one-dimensional page and liquid chromatography-tandem mass spectrometry. *Molecular & Cellular Proteomics* 2(12), 1297-1305 (2003).
339. Schilling B, Rardin MJ, Maclean BX *et al*. Platform-independent and label-free quantitation of proteomic data using ms1 extracted ion chromatograms in skyline application to protein acetylation and phosphorylation. *Molecular & Cellular Proteomics* 11(5), 202-214 (2012).
340. Titulaer MK. Candidate biomarker discovery for angiogenesis by automatic integration of orbitrap ms1 spectral- and x!Tandem ms2 sequencing information. *Genomics, proteomics & bioinformatics* 11(3), 182-194 (2013).
341. Krey JF, Wilmarth PA, Shin JB *et al*. Accurate label-free protein quantitation with high- and low-resolution mass spectrometers. *Journal of proteome research* 13(2), 1034-1044 (2014).

342. Zhou W, Liotta LA, Petricoin EF. The spectra count label-free quantitation in cancer proteomics. *Cancer genomics & proteomics* 9(3), 135-142 (2012).
343. Tu C, Li J, Sheng Q, Zhang M, Qu J. Systematic assessment of survey scan and ms2-based abundance strategies for label-free quantitative proteomics using high-resolution ms data. *Journal of proteome research* 13(4), 2069-2079 (2014).
344. Daly CE, Ng LL, Hakimi A, Willingale R, Jones DJ. Qualitative and quantitative characterization of plasma proteins when incorporating traveling wave ion mobility into a liquid chromatography-mass spectrometry workflow for biomarker discovery: Use of product ion quantitation as an alternative data analysis tool for label free quantitation. *Anal Chem* 86(4), 1972-1979 (2014).
345. Levin Y. The role of statistical power analysis in quantitative proteomics. *Proteomics* 11(12), 2565-2567 (2011).
346. Pieper R, Gatlin CL, Mcgrath AM *et al*. Characterization of the human urinary proteome: A method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots. *Proteomics* 4(4), 1159-1174 (2004).
347. Anderson NG, Anderson NL, Tollaksen SL. Proteins of human urine. I. Concentration and analysis by two-dimensional electrophoresis. *Clinical chemistry* 25(7), 1199-1210 (1979).
348. Marimuthu A, O'meally RN, Chaerkady R *et al*. A comprehensive map of the human urinary proteome. *Journal of proteome research* 10(6), 2734-2743 (2011).
349. Kiprijanovska S, Stavridis S, Stankov O *et al*. Mapping and identification of the urine proteome of prostate cancer patients by 2d page/ms. *International journal of proteomics* 2014, 594761 (2014).
350. Chen YT, Tsao CY, Li JM, Tsai CY, Chiu SF, Tseng TL. Large-scale protein identification of human urine proteome by multi-dimensional lc and ms/ms. *Proteomics. Clinical applications* 1(6), 577-587 (2007).
351. Liu X, Shao C, Wei L *et al*. An individual urinary proteome analysis in normal human beings to define the minimal sample number to represent the normal urinary proteome. *Proteome science* 10(1), 70 (2012).
352. Kim MS, Pinto SM, Getnet D *et al*. A draft map of the human proteome. *Nature* 509(7502), 575-581 (2014).
353. Sun W, Chen Y, Li FX *et al*. Dynamic urinary proteomic analysis reveals stable proteins to be potential biomarkers. *Proteom Clin Appl* 3(3), 370-382 (2009).
354. Haj-Ahmad TA, Abdalla MA, Haj-Ahmad Y. Potential urinary protein biomarker candidates for the accurate detection of prostate cancer among benign prostatic hyperplasia patients. *Journal of Cancer* 5(2), 103-114 (2014).
355. Silvera D, Formenti SC, Schneider RJ. Translational control in cancer. *Nature reviews. Cancer* 10(4), 254-266 (2010).
356. Rickman DS, Chen YB, Banerjee S *et al*. Erg cooperates with androgen receptor in regulating trefoil factor 3 in prostate cancer disease progression. *Neoplasia* 12(12), 1031-1040 (2010).
357. Li QR, Fan KX, Li RX *et al*. A comprehensive and non-prefractionation on the protein level approach for the human urinary proteome: Touching phosphorylation in urine. *Rapid communications in mass spectrometry : RCM* 24(6), 823-832 (2010).
358. Mahadevan NR, Rodvold J, Almanza G, Perez AF, Wheeler MC, Zanetti M. Er stress drives lipocalin 2 upregulation in prostate cancer cells in an nf-kappab-dependent manner. *BMC cancer* 11, 229 (2011).
359. Foley R, Hollywood D, Lawler M. Molecular pathology of prostate cancer: The key to identifying new biomarkers of disease. *Endocrine-related cancer* 11(3), 477-488 (2004).
360. Madu CO, Lu Y. Novel diagnostic biomarkers for prostate cancer. *Journal of Cancer* 1, 150-177 (2010).

361. Pratapa PN, Patz EF, Jr., Hartemink AJ. Finding diagnostic biomarkers in proteomic spectra. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 279-290 (2006).
362. Fuzery AK, Levin J, Chan MM, Chan DW. Translation of proteomic biomarkers into fda approved cancer diagnostics: Issues and challenges. *Clinical proteomics* 10(1), 13 (2013).
363. Frantzi M, Bhat A, Latosinska A. Clinical proteomic biomarkers: Relevant issues on study design & technical considerations in biomarker development. *Clinical and translational medicine* 3(1), 7 (2014).
364. Velonas VM, Woo HH, Remedios CG, Assinder SJ. Current status of biomarkers for prostate cancer. *International journal of molecular sciences* 14(6), 11034-11060 (2013).
365. Hayes JH, Barry MJ. Screening for prostate cancer with the prostate-specific antigen test: A review of current evidence. *JAMA : the journal of the American Medical Association* 311(11), 1143-1149 (2014).
366. Gao L, Alumkal J. Epigenetic regulation of androgen receptor signaling in prostate cancer. *Epigenetics : official journal of the DNA Methylation Society* 5(2), 100-104 (2010).
367. Grant CM, Kyprianou N. Epithelial mesenchymal transition (emt) in prostate growth and tumor progression. *Translational andrology and urology* 2(3), 202-211 (2013).
368. Collins AT, Berry PA, Hyde C, Stower MJ, Maitland NJ. Prospective identification of tumorigenic prostate cancer stem cells. *Cancer research* 65(23), 10946-10951 (2005).
369. Goldkorn A, Ely B, Tangen CM *et al*. Circulating tumor cell telomerase activity as a prognostic marker for overall survival in swog 0421: A phase iii metastatic castration resistant prostate cancer trial. *International journal of cancer. Journal international du cancer*, (2014).
370. Parkin DM, Bray F, Ferlay J, Jemal A. Cancer in africa 2012. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 23(6), 953-966 (2014).
371. Filella X, Foj L. Emerging biomarkers in the detection and prognosis of prostate cancer. *Clinical chemistry and laboratory medicine : CCLM / FESCC*, (2015).
372. Schalken J, Dijkstra S, Baskin-Bey E, Van Oort I. Potential utility of cancer-specific biomarkers for assessing response to hormonal treatments in metastatic prostate cancer. *Therapeutic advances in urology* 6(6), 245-252 (2014).
373. Anderson NL, Anderson NG. The human plasma proteome: History, character, and diagnostic prospects. *Molecular & cellular proteomics : MCP* 1(11), 845-867 (2002).
374. Gutman S, Kessler LG. The us food and drug administration perspective on cancer biomarker development. *Nature reviews. Cancer* 6(7), 565-571 (2006).
375. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: The long and uncertain path to clinical utility. *Nature biotechnology* 24(8), 971-983 (2006).
376. Taube SE, Clark GM, Dancey JE, Mcshane LM, Sigman CC, Gutman SI. A perspective on challenges and issues in biomarker development and drug and biomarker codevelopment. *Journal of the National Cancer Institute* 101(21), 1453-1463 (2009).
377. Zhang GL, Deluca DS, Brusic V. Database resources for proteomics-based analysis of cancer. *Methods in molecular biology* 723, 349-364 (2011).
378. Zhang Y, Zhang Y, Adachi J *et al*. Mapu: Max-planck unified database of organellar, cellular, tissue and body fluid proteomes. *Nucleic acids research* 35(Database issue), D771-779 (2007).
379. Ahmed J, Meinel T, Dunkel M *et al*. Cancerresource: A comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge. *Nucleic acids research* 39(Database issue), D960-967 (2011).
380. Mclaughlin T, Siepen JA, Selley J *et al*. Pepseeker: A database of proteome peptide identifications for investigating fragmentation patterns. *Nucleic acids research* 34(Database issue), D649-654 (2006).

381. Vizcaino JA, Foster JM, Martens L. Proteomics data repositories: Providing a safe haven for your data and acting as a springboard for further research. *J Proteomics* 73(11), 2136-2146 (2010).
382. Colangelo CM, Shifman M, Cheung KH *et al*. Yped: An integrated bioinformatics suite and database for mass spectrometry-based proteomics research. *Genomics, proteomics & bioinformatics* 13(1), 25-35 (2015).
383. Shadforth I, Xu W, Crowther D, Bessant C. Gapp: A fully automated software for the confident identification of human peptides from tandem mass spectra. *Journal of proteome research* 5(10), 2849-2852 (2006).
384. Gaudet P, Michel PA, Zahn-Zabal M *et al*. The nextprot knowledgebase on human proteins: Current status. *Nucleic acids research* 43(D1), D764-D770 (2015).
385. Smith BE, Hill JA, Gjukich MA, Andrews PC. Tranche distributed repository and proteomecommons.Org. *Methods in molecular biology* 696, 123-145 (2011).
386. Martens L. Proteomics databases and repositories. *Methods in molecular biology* 694, 213-227 (2011).
387. Perez-Riverol Y, Alpi E, Wang R, Hermjakob H, Vizcaino JA. Making proteomics data accessible and reusable: Current state of proteomics databases and repositories. *Proteomics* 15(5-6), 930-949 (2015).
388. Gromov PS, Ostergaard M, Gromova I, Celis JE. Human proteomic databases: A powerful resource for functional genomics in health and disease. *Progress in biophysics and molecular biology* 80(1-2), 3-22 (2002).
389. Riffle M, Eng JK. Proteomics data repositories. *Proteomics* 9(20), 4653-4663 (2009).
390. Kamath KS, Vasavada MS, Srivastava S. Proteomic databases and tools to decipher post-translational modifications. *J Proteomics* 75(1), 127-144 (2011).
391. Vizcaino JA, Mueller M, Hermjakob H, Martens L. Charting online omics resources: A navigational chart for clinical researchers. *Proteomics. Clinical applications* 3(1), 18-29 (2009).
392. Desiere F, Deutsch EW, King NL *et al*. The peptideatlas project. *Nucleic acids research* 34(Database issue), D655-658 (2006).
393. Deutsch EW. The peptideatlas project. *Methods in molecular biology* 604, 285-296 (2010).
394. Deutsch EW, Lam H, Aebersold R. Peptideatlas: A resource for target selection for emerging targeted proteomics workflows. *EMBO reports* 9(5), 429-434 (2008).
395. Rost H, Malmstrom L, Aebersold R. A computational tool to detect and avoid redundancy in selected reaction monitoring. *Molecular & cellular proteomics : MCP* 11(8), 540-549 (2012).
396. Huttenhain R, Surinova S, Ossola R *et al*. N-glycoprotein srmatlas: A resource of mass spectrometric assays for n-glycosites enabling consistent and multiplexed protein quantification for clinical applications. *Molecular & cellular proteomics : MCP* 12(4), 1005-1016 (2013).
397. Kusebauch U, Deutsch EW, Campbell DS, Sun Z, Farrah T, Moritz RL. Using peptideatlas, srmatlas, and passel: Comprehensive resources for discovery and targeted proteomics. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* 46, 13 25 11-13 25 28 (2014).
398. Uhlen M, Bjorling E, Agaton C *et al*. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Molecular & cellular proteomics : MCP* 4(12), 1920-1932 (2005).
399. Nilsson P, Paavilainen L, Larsson K *et al*. Towards a human proteome atlas: High-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics* 5(17), 4327-4337 (2005).
400. Doerr A. Targeting with prm. *Nature methods* 9(10), 950 (2012).
401. Ronsein GE, Pamir N, Von Haller PD *et al*. Parallel reaction monitoring (prm) and selected reaction monitoring (srm) exhibit comparable linearity, dynamic range and precision for targeted quantitative hdl proteomics. *J Proteomics* 113, 388-399 (2015).

402. Adeola HA, Calder B, Soares NC, Kaestner L, Blackburn JM, Zerbini LF. In silico verification and parallel reaction monitoring prevalidation of potential prostate cancer biomarkers. *Future oncology*, (2015).
403. Lin D, Alborn WE, Slebos RJ, Liebler DC. Comparison of protein immunoprecipitation-multiple reaction monitoring with elisa for assay of biomarker candidates in plasma. *J Proteome Res* 12(12), 5996-6003 (2013).
404. Wasinger VC, Zeng M, Yau Y. Current status and advances in quantitative proteomic mass spectrometry. *Int J Proteomics* 2013, 180605 (2013).
405. Fortin T, Salvador A, Charrier JP *et al*. Clinical quantitation of prostate-specific antigen biomarker in the low nanogram/milliliter range by conventional bore liquid chromatography-tandem mass spectrometry (multiple reaction monitoring) coupling and correlation with elisa tests. *Mol Cell Proteomics* 8(5), 1006-1015 (2009).
406. Ye X, Blonder J, Veenstra TD. Targeted proteomics for validation of biomarkers in clinical samples. *Briefings in functional genomics & proteomics* 8(2), 126-135 (2009).
407. Solier C, Langen H. Antibody-based proteomics and biomarker research - current status and limitations. *Proteomics* 14(6), 774-783 (2014).
408. Mischak H, Kolch W, Aivaliotis M *et al*. Comprehensive human urine standards for comparability and standardization in clinical proteome analysis. *Proteomics. Clinical applications* 4(4), 464-478 (2010).
409. Remo A, Simeone I, Pancione M *et al*. Systems biology analysis reveals nfat5 as a novel biomarker and master regulator of inflammatory breast cancer. *Journal of translational medicine* 13, 138 (2015).
410. Cheng TC, Chuang KH, Roffler SR *et al*. Discovery of specific inhibitors for intestinal e. Coli beta-glucuronidase through in silico virtual screening. *TheScientificWorldJournal* 2015, 740815 (2015).
411. Marani MM, Costa J, Mafra I, Oliveira MB, Camperi SA, De Souza Almeida Leite JR. In silico peptide prediction for antibody generation to recognize 5-enolpyruvylshikimate-3-phosphate synthase (epsps) in genetically modified organisms. *Biopolymers* 104(2), 91-100 (2015).
412. Sacan A, Ekins S, Kortagere S. Applications and limitations of in silico models in drug discovery. *Methods in molecular biology* 910, 87-124 (2012).
413. Ponten F, Jirstrom K, Uhlen M. The human protein atlas--a tool for pathology. *The Journal of pathology* 216(4), 387-393 (2008).
414. Ponten F, Schwenk JM, Asplund A, Edqvist PH. The human protein atlas as a proteomic resource for biomarker discovery. *Journal of internal medicine* 270(5), 428-446 (2011).
415. Berglund L, Bjorling E, Oksvold P *et al*. A genecentric human protein atlas for expression profiles based on antibodies. *Molecular & cellular proteomics : MCP* 7(10), 2019-2027 (2008).
416. Fagerberg L, Hallstrom BM, Oksvold P *et al*. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & cellular proteomics : MCP* 13(2), 397-406 (2014).
417. Rhodes KJ, Trimmer JS. Antibodies as valuable neuroscience research tools versus reagents of mass distraction. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26(31), 8017-8020 (2006).
418. Praekelt U, Kopp PM, Rehm K *et al*. New isoform-specific monoclonal antibodies reveal different sub-cellular localisations for talin1 and talin2. *European journal of cell biology* 91(3), 180-191 (2012).
419. Gallien S, Bourmaud A, Kim SY, Domon B. Technical considerations for large-scale parallel reaction monitoring analysis. *J Proteomics* 100, 147-159 (2014).
420. Ross RK, Coetzee GA, Reichardt J, Skinner E, Henderson BE. Does the racial-ethnic variation in prostate-cancer risk have a hormonal basis. *Cancer* 75(7), 1778-1782 (1995).

421. Freedman ML, Pearce CL, Penney KL *et al*. Systematic evaluation of genetic variation at the androgen receptor locus and risk of prostate cancer in a multiethnic cohort study. *American journal of human genetics* 76(1), 82-90 (2005).
422. Sartor O, Zheng Q, Eastham JA. Androgen receptor gene cag repeat length varies in a race-specific fashion in men without prostate cancer. *Urology* 53(2), 378-380 (1999).
423. Hayes RB, Ziegler RG, Gridley G *et al*. Dietary factors and risks for prostate cancer among blacks and whites in the united states. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 8(1), 25-34 (1999).
424. Hollis BW, Marshall DT, Savage SJ, Garrett-Mayer E, Kindy MS, Gattoni-Celli S. Vitamin d3 supplementation, low-risk prostate cancer, and health disparities. *The Journal of steroid biochemistry and molecular biology* 136, 233-237 (2013).
425. Martin DN, Starks AM, Ambs S. Biological determinants of health disparities in prostate cancer. *Current opinion in oncology* 25(3), 235-241 (2013).
426. Powell IJ, Dyson G, Land S *et al*. Genes associated with prostate cancer are differentially expressed in african american and european american men. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 22(5), 891-897 (2013).
427. Dias SJ, Zhou X, Ivanovic M *et al*. Nuclear mta1 overexpression is associated with aggressive prostate cancer, recurrence and metastasis in african americans. *Scientific reports* 3, 2331 (2013).
428. Filella X, Foj L. Emerging biomarkers in the detection and prognosis of prostate cancer. *Clinical chemistry and laboratory medicine : CCLM / FESCC* 53(7), 963-973 (2015).
429. Liong ML, Lim CR, Yang H *et al*. Blood-based biomarkers of aggressive prostate cancer. *PloS one* 7(9), e45802 (2012).
430. Hernandez J, Thompson IM. Prostate-specific antigen: A review of the validation of the most commonly used cancer biomarker. *Cancer* 101(5), 894-904 (2004).
431. Quintero IB, Araujo CL, Pulkka AE *et al*. Prostatic acid phosphatase is not a prostate specific target. *Cancer research* 67(14), 6549-6554 (2007).
432. Stephan C, Ralla B, Jung K. Prostate-specific antigen and other serum and urine markers in prostate cancer. *Biochimica et biophysica acta* 1846(1), 99-112 (2014).
433. Bratt O, Lilja H. Serum markers in prostate cancer detection. *Current opinion in urology* 25(1), 59-64 (2015).
434. Behesnilian AS, Reiter RE. Risk stratification of prostate cancer in the modern era. *Current opinion in urology* 25(3), 246-251 (2015).
435. Zhang DY, Ye F, Gao L *et al*. Proteomics, pathway array and signaling network-based medicine in cancer. *Cell division* 4, 20 (2009).
436. Rho JH, Lampe PD. High-throughput screening for native autoantigen-autoantibody complexes using antibody microarrays. *Journal of proteome research* 12(5), 2311-2320 (2013).
437. Gnjatic S, Ritter E, Buchler MW *et al*. Seromic profiling of ovarian and pancreatic cancer. *Proceedings of the National Academy of Sciences of the United States of America* 107(11), 5088-5093 (2010).
438. Blackburn JM, Shoko A. Protein function microarrays for customised systems-oriented proteome analysis. *Methods in molecular biology* 785, 305-330 (2011).
439. Alhamdani MS, Schroder C, Hoheisel JD. Oncoproteomic profiling with antibody microarrays. *Genome medicine* 1(7), 68 (2009).
440. Haab BB. Methods and applications of antibody microarrays in cancer research. *Proteomics* 3(11), 2116-2122 (2003).

441. Haab BB. Antibody arrays in cancer research. *Molecular & cellular proteomics : MCP* 4(4), 377-383 (2005).
442. Cho-Chung YS. Autoantibody biomarkers in the detection of cancer. *Biochimica et biophysica acta* 1762(6), 587-591 (2006).
443. Stempfer R, Syed P, Vierlinger K *et al*. Tumour auto-antibody screening: Performance of protein microarrays using serex derived antigens. *BMC cancer* 10, 627 (2010).
444. Simpson AJ, Caballero OL, Jungbluth A, Chen YT, Old LJ. Cancer/testis antigens, gametogenesis and cancer. *Nature reviews. Cancer* 5(8), 615-625 (2005).
445. Kalejs M, Erenpreisa J. Cancer/testis antigens and gametogenesis: A review and "brain-storming" session. *Cancer cell international* 5(1), 4 (2005).
446. Balafoutas D, Zur Hausen A, Mayer S *et al*. Cancer testis antigens and ny-br-1 expression in primary breast cancer: Prognostic and therapeutic implications. *BMC cancer* 13, 271 (2013).
447. Fratta E, Coral S, Covre A *et al*. The biology of cancer testis antigens: Putative function, regulation and therapeutic potential. *Molecular oncology* 5(2), 164-182 (2011).
448. Kulkarni P, Shiraishi T, Rajagopalan K, Kim R, Mooney SM, Getzenberg RH. Cancer/testis antigens and urological malignancies. *Nat Rev Urol* 9(7), 386-396 (2012).
449. Gnjatic S, Wheeler C, Ebner M *et al*. Seromic analysis of antibody responses in non-small cell lung cancer patients and healthy donors using conformational protein arrays. *Journal of immunological methods* 341(1-2), 50-58 (2009).
450. Takahashi S, Shiraishi T, Miles N, Trock BJ, Kulkarni P, Getzenberg RH. Nanowire analysis of cancer-testis antigens as biomarkers of aggressive prostate cancer. *Urology* 85(3), 704 e701-707 (2015).
451. Shiraishi T, Getzenberg RH, Kulkarni P. Cancer/testis antigens: Novel tools for discerning aggressive and non-aggressive prostate cancer. *Asian journal of andrology* 14(3), 400-404 (2012).
452. Suyama T, Shiraishi T, Zeng Y *et al*. Expression of cancer/testis antigens in prostate cancer is associated with disease progression. *The Prostate* 70(16), 1778-1787 (2010).
453. Cheema Z, Hari-Gupta Y, Kita GX *et al*. Expression of the cancer-testis antigen boris correlates with prostate cancer. *The Prostate* 74(2), 164-176 (2014).
454. Shiraishi T, Terada N, Zeng Y *et al*. Cancer/testis antigens as potential predictors of biochemical recurrence of prostate cancer following radical prostatectomy. *Journal of translational medicine* 9, 153 (2011).
455. Adeola HA, Smith M, Kaestner L, Blackburn JM, Zerbini LF. Novel potential serological prostate cancer biomarkers using ct100+ cancer antigen microarray platform in a multi-cultural south african cohort. *Oncotarget*, (2016).
456. Kingsmore SF. Multiplexed protein measurement: Technologies and applications of protein and antibody arrays. *Nature reviews. Drug discovery* 5(4), 310-320 (2006).
457. Burbelo PD, Ching KH, Bush ER, Han BL, Iadarola MJ. Antibody-profiling technologies for studying humoral responses to infectious agents. *Expert review of vaccines* 9(6), 567-578 (2010).
458. Tabakman SM, Lau L, Robinson JT *et al*. Plasmonic substrates for multiplexed protein microarrays with femtomolar sensitivity and broad dynamic range. *Nature communications* 2, 466 (2011).
459. Dudas SP, Chatterjee M, Tainsky MA. Usage of cancer associated autoantibodies in the detection of disease. *Cancer biomarkers : section A of Disease markers* 6(5-6), 257-270 (2010).
460. Chatterjee M, Draghici S, Tainsky MA. Immunotheranostics: Breaking tolerance in immunotherapy using tumor autoantigens identified on protein microarrays. *Current opinion in drug discovery & development* 9(3), 380-385 (2006).

461. Tojo A, Kinugasa S. Mechanisms of glomerular albumin filtration and tubular reabsorption. *International journal of nephrology* 2012, 481520 (2012).
462. Couturier MR, Graf EH, Griffin AT. Urine antigen tests for the diagnosis of respiratory infections: Legionellosis, histoplasmosis, pneumococcal pneumonia. *Clinics in laboratory medicine* 34(2), 219-236 (2014).
463. Sinclair A, Xie X, Teltscher M, Dendukuri N. Systematic review and meta-analysis of a urine-based pneumococcal antigen test for diagnosis of community-acquired pneumonia caused by streptococcus pneumoniae. *Journal of clinical microbiology* 51(7), 2303-2310 (2013).
464. Choudhry V, Saxena RK. Detection of mycobacterium tuberculosis antigens in urinary proteins of tuberculosis patients. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology* 21(1), 1-5 (2002).
465. Corral RS, Altcheh J, Alexandre SR, Grinstein S, Freilij H, Katzin AM. Detection and characterization of antigens in urine of patients with acute, congenital, and chronic chagas' disease. *Journal of clinical microbiology* 34(8), 1957-1962 (1996).
466. Yzerman EP, Den Boer JW, Lettinga KD, Schellekens J, Dankert J, Peeters M. Sensitivity of three urinary antigen tests associated with clinical severity in a large outbreak of legionnaires' disease in the netherlands. *Journal of clinical microbiology* 40(9), 3232-3236 (2002).
467. Kauffman CA. Histoplasmosis: A clinical and laboratory update. *Clinical microbiology reviews* 20(1), 115-132 (2007).
468. Chapin-Robertson K, Bechtel C, Waycott S, Kontnick C, Edberg SC. Cryptococcal antigen detection from the urine of aids patients. *Diagnostic microbiology and infectious disease* 17(3), 197-201 (1993).
469. Umezawa ES, Shikanai-Yasuda MA, Da Silveira JF, Cotrim PC, Paranhos G, Katzin AM. Trypanosoma cruzi: Detection of a circulating antigen in urine of chagasic patients sharing common epitopes with an immunodominant repetitive antigen. *Experimental parasitology* 76(4), 352-357 (1993).
470. Attallah AM, Ismail H, El Masry SA *et al*. Rapid detection of a schistosoma mansoni circulating antigen excreted in urine of infected individuals by using a monoclonal antibody. *Journal of clinical microbiology* 37(2), 354-357 (1999).
471. Machnicka B, Prokopowicz D, Dziemian E, Kolodziej-Sobociniska M. Detection of trichinella spiralis antigens in urine of men and animals. *Wiadomosci parazytologiczne* 47(2), 217-225 (2001).
472. Holmquist P, Sjoblad S, Torffvit O. Pore size and charge selectivity of the glomerular membrane at the time of diagnosis of diabetes. *Pediatr Nephrol* 19(12), 1361-1366 (2004).
473. Haraldsson B, Nystrom J, Deen WM. Properties of the glomerular barrier and mechanisms of proteinuria. *Physiological reviews* 88(2), 451-487 (2008).
474. Kurts C, Panzer U, Anders HJ, Rees AJ. The immune system and kidney disease: Basic concepts and clinical implications. *Nature reviews. Immunology* 13(10), 738-753 (2013).
475. Gottschalk C, Damuzzo V, Gotot J *et al*. Batf3-dependent dendritic cells in the renal lymph node induce tolerance against circulating antigens. *Journal of the American Society of Nephrology : JASN* 24(4), 543-549 (2013).
476. Pindjakova J, Griffin MD. The renal lymph node and immune tolerance to filtered antigens. *Journal of the American Society of Nephrology : JASN* 24(4), 519-521 (2013).
477. Mishra DK, Chen Z, Wu Y, Sarkissyan M, Koeffler HP, Vadgama JV. Global methylation pattern of genes in androgen-sensitive and androgen-independent prostate cancer cells. *Molecular cancer therapeutics* 9(1), 33-45 (2010).
478. Atanackovic D, Luetkens T, Kloth B *et al*. Cancer-testis antigen expression and its epigenetic modulation in acute myeloid leukemia. *American journal of hematology* 86(11), 918-922 (2011).

479. Kouprina N, Noskov VN, Solomon G *et al*. Mutational analysis of spanx genes in families with x-linked prostate cancer. *The Prostate* 67(8), 820-828 (2007).
480. Yao S, Ireland SJ, Bee A *et al*. Splice variant prkc-zeta(-prc) is a novel biomarker of human prostate cancer. *British journal of cancer* 107(2), 388-399 (2012).
481. Zabransky DJ, Smith HA, Thoburn CJ *et al*. Lenalidomide modulates il-8 and anti-prostate antibody levels in men with biochemically recurrent prostate cancer. *The Prostate* 72(5), 487-498 (2012).
482. Hakem A, Bohgaki M, Lemmers B *et al*. Role of pirh2 in mediating the regulation of p53 and c-myc. *PLoS genetics* 7(11), e1002360 (2011).
483. Dai C, Gu W. P53 post-translational modification: Deregulated in tumorigenesis. *Trends in molecular medicine* 16(11), 528-536 (2010).
484. Lee SJ, Lee SH, Yoon MH, Park BJ. A new p53 target gene, rkip, is essential for DNA damage-induced cellular senescence and suppression of erk activation. *Neoplasia* 15(7), 727-737 (2013).
485. Suppiah A, Greenman J. Clinical utility of anti-p53 auto-antibody: Systematic review and focus on colorectal cancer. *World journal of gastroenterology : WJG* 19(29), 4651-4670 (2013).
486. Amini S, Fathi F, Mobalegi J, Sofimajidpour H, Ghadimi T. The expressions of stem cell markers: Oct4, nanog, sox2, nucleostemin, bmi, zfx, tcl1, tbx3, dppa4, and esrrb in bladder, colon, and prostate cancer, and certain cancer cell lines. *Anatomy & cell biology* 47(1), 1-11 (2014).
487. Duan Z, Duan Y, Lamendola DE *et al*. Overexpression of mage/gage genes in paclitaxel/doxorubicin-resistant human cancer cell lines. *Clinical cancer research : an official journal of the American Association for Cancer Research* 9(7), 2778-2785 (2003).
488. Wang H, Zhou W, Zheng Z *et al*. The hdac inhibitor depsipeptide transactivates the p53/p21 pathway by inducing DNA damage. *DNA repair* 11(2), 146-156 (2012).
489. Chun HK, Chung KS, Kim HC *et al*. Oip5 is a highly expressed potential therapeutic target for colorectal and gastric cancers. *BMB reports* 43(5), 349-354 (2010).
490. Zhang JY. Mini-array of multiple tumor-associated antigens to enhance autoantibody detection for immunodiagnosis of hepatocellular carcinoma. *Autoimmunity reviews* 6(3), 143-148 (2007).
491. Li L, Wang K, Dai L, Wang P, Peng XX, Zhang JY. Detection of autoantibodies to multiple tumor-associated antigens in the immunodiagnosis of ovarian cancer. *Molecular medicine reports* 1(4), 589-594 (2008).
492. Ye H, Sun C, Ren P *et al*. Mini-array of multiple tumor-associated antigens (taas) in the immunodiagnosis of breast cancer. *Oncology letters* 5(2), 663-668 (2013).
493. Bunger S, Haug U, Kelly M *et al*. A novel multiplex-protein array for serum diagnostics of colon cancer: A case-control study. *BMC cancer* 12, 393 (2012).
494. De Martel C, Ferlay J, Franceschi S *et al*. Global burden of cancers attributable to infections in 2008: A review and synthetic analysis. *The lancet oncology* 13(6), 607-615 (2012).
495. Viljoen KS, Dakshinamurthy A, Goldberg P, Blackburn JM. Quantitative profiling of colorectal cancer-associated bacteria reveals associations between fusobacterium spp., enterotoxigenic bacteroides fragilis (etbf) and clinicopathological features of colorectal cancer. *PloS one* 10(3), e0119462 (2015).
496. Kyi C, Postow MA. Checkpoint blocking antibodies in cancer immunotherapy. *FEBS letters* 588(2), 368-376 (2014).
497. Park TS, Rosenberg SA, Morgan RA. Treating cancer with genetically engineered t cells. *Trends in biotechnology* 29(11), 550-557 (2011).
498. Joniau S, Abrahamsson PA, Bellmunt J *et al*. Current vaccination strategies for prostate cancer. *European urology* 61(2), 290-306 (2012).
499. Chang RT, Kirby R, Challacombe BJ. Is there a link between bph and prostate cancer? *The Practitioner* 256(1750), 13-16, 12 (2012).

500. Orsted DD, Bojesen SE, Nielsen SF, Nordestgaard BG. Association of clinical benign prostate hyperplasia with prostate cancer incidence and mortality revisited: A nationwide cohort study of 3,009,258 men. *European urology* 60(4), 691-698 (2011).
501. De Nunzio C, Kramer G, Marberger M *et al*. The controversial relationship between benign prostatic hyperplasia and prostate cancer: The role of inflammation. *European urology* 60(1), 106-117 (2011).
502. Miah S, Catto J. Bph and prostate cancer risk. *Indian journal of urology : IJU : journal of the Urological Society of India* 30(2), 214-218 (2014).
503. Kampf C, Bergman J, Oksvold P *et al*. A tool to facilitate clinical biomarker studies--a tissue dictionary based on the human protein atlas. *BMC medicine* 10, 103 (2012).
504. Liu Y, Huttenhain R, Collins B, Aebersold R. Mass spectrometric protein maps for biomarker discovery and clinical research. *Expert review of molecular diagnostics* 13(8), 811-825 (2013).
505. Huttenhain R, Malmstrom J, Picotti P, Aebersold R. Perspectives of targeted mass spectrometry for protein biomarker verification. *Current opinion in chemical biology* 13(5-6), 518-525 (2009).
506. Surinova S, Radova L, Choi M *et al*. Non-invasive prognostic protein biomarker signatures associated with colorectal cancer. *EMBO molecular medicine* 7(9), 1153-1165 (2015).
507. Qin S, Zhou Y, Lok AS *et al*. Srm targeted proteomics in search for biomarkers of hcv-induced progression of fibrosis to cirrhosis in halt-c patients. *Proteomics* 12(8), 1244-1252 (2012).
508. Harlan R, Zhang H. Targeted proteomics: A bridge between discovery and validation. *Expert Rev Proteomic* 11(6), 657-661 (2014).
509. Hashim D, Boffetta P. Occupational and environmental exposures and cancers in developing countries. *Ann Glob Health* 80(5), 393-411 (2014).
510. Thun MJ, Delancey JO, Center MM, Jemal A, Ward EM. The global burden of cancer: Priorities for prevention. *Carcinogenesis* 31(1), 100-110 (2010).
511. De Martel C, Ferlay J, Franceschi S *et al*. Global burden of cancers attributable to infections in 2008: A review and synthetic analysis. *Lancet Oncology* 13(6), (2012).
512. Danaei G. Global burden of infection-related cancer revisited. *Lancet Oncology* 13(6), 564-565 (2012).
513. Parkin DM. The global health burden of infection -associated cancers in the year 2002. *International Journal of Cancer* 118(12), 3030-3044 (2006).
514. Oh JK, Weiderpass E. Infection and cancer: Global distribution and burden of diseases. *Ann Glob Health* 80(5), 384-392 (2014).
515. Klein EA, Silverman R. Inflammation, infection, and prostate cancer. *Current opinion in urology* 18(3), 315-319 (2008).
516. Sfanos KS, Isaacs WB, De Marzo AM. Infections and inflammation in prostate cancer. *American journal of clinical and experimental urology* 1(1), 3-11 (2013).
517. Sutcliffe S. Sexually transmitted infections and risk of prostate cancer: Review of historical and emerging hypotheses. *Future oncology* 6(8), 1289-1311 (2010).
518. Litman HJ, Bhasin S, Link CL, Araujo AB, Mckinlay JB. Serum androgen levels in black, hispanic, and white men. J Clin Endocrinol Metab 91(11), 4326-4334 (2006).

CHAPTER 8: ANNEXURE

8.1. Annexure I- Total annotated urinary proteome coverage of prostate cancer cohort

Protein names	Gene names	PEP Score
10 kDa heat shock protein, mitochondrial	HSPE1	7.0185E-60
116 kDa U5 small nuclear ribonucleoprotein component	EFTUD2	6.0199E-80
14 kDa phosphohistidine phosphatase	PHPT1	3.5189E-25
14-3-3 protein beta/alpha;14-3-3 protein beta/alpha, N-terminally processed	YWHAB	5.722E-07
14-3-3 protein epsilon	YWHAE	0
14-3-3 protein eta	YWHAH	5.576E-123
14-3-3 protein gamma;14-3-3 protein gamma, N-terminally processed	YWHAG	0
14-3-3 protein sigma	SFN	6.6163E-09
14-3-3 protein theta;14-3-3 protein sigma	YWHAQ;SFN	4.6834E-30
14-3-3 protein zeta/delta	YWHAZ	1.2883E-35
14-3-3 protein zeta/delta	YWHAZ	7.4808E-27
2,4-dienoyl-CoA reductase, mitochondrial	DECR1	0.000034155
26S protease regulatory subunit 10B	PSMC6	9.7136E-164
26S proteasome non-ATPase regulatory subunit 12	PSMD12	7.0184E-168
26S proteasome non-ATPase regulatory subunit 2	PSMD2	5.3442E-16
26S proteasome non-ATPase regulatory subunit 6	PSMD6	5.9382E-12
26S proteasome non-ATPase regulatory subunit 9	PSMD9	5.5384E-08
3-mercaptopyruvate sulfurtransferase;Sulfurtransferase	MPST	0
40S ribosomal protein S12	RPS12	1.3927E-27
4F2 cell-surface antigen heavy chain	SLC3A2	1.6768E-07
60 kDa heat shock protein, mitochondrial	HSPD1	0
60 kDa SS-A/Ro ribonucleoprotein	TROVE2	2.6641E-146

60S acidic ribosomal protein P0	RPLP0	2.0811E-10
60S acidic ribosomal protein P1	RPLP1	8.1958E-07
60S acidic ribosomal protein P2	RPLP2	4.2184E-40
60S ribosomal protein L13	RPL13	2.5827E-07
6-phosphogluconate dehydrogenase, decarboxylating	PGD	2.4568E-33
6-phosphogluconolactonase	PGLS	3.802E-17
7,8-dihydro-8-oxoguanine triphosphatase	NUDT1	0.0012421
78 kDa glucose-regulated protein	HSPA5	3.1518E-30
A disintegrin and metalloproteinase with thrombospondin motifs 1	ADAMTS1	6.0132E-09
Abhydrolase domain-containing protein 14B	ABHD14B	1.9826E-78
Abl interactor 1	ABI1	0.0011621
Acid ceramidase;Acid ceramidase subunit alpha;Acid ceramidase subunit beta	ASAH1	0
Acidic leucine-rich nuclear phosphoprotein 32 family member A	ANP32A	0
Acidic leucine-rich nuclear phosphoprotein 32 family member B;Acidic leucine-rich nuclear phosphoprotein 32 family member A	ANP32B;ANP32A	4.3618E-15
Actin, alpha cardiac muscle 1;Actin, alpha skeletal muscle;Actin, aortic smooth muscle;Actin, gamma-enteric smooth muscle	ACTC1;ACTA1;ACTA2; ACTG2	0.00034984
Actin, cytoplasmic 1;Actin, cytoplasmic 1, N-terminally processed	ACTB	8.8665E-263
Actin, cytoplasmic 2;Actin, cytoplasmic 2, N-terminally processed	ACTG1	4.547E-11
Actin-like protein 6A	ACTL6A	3.3554E-192
Actin-related protein 2	ACTR2	0
Actin-related protein 2/3 complex subunit 1B	ARPC1B	0
Actin-related protein 2/3 complex subunit 2	ARPC2	6.2852E-293
Actin-related protein 2/3 complex subunit 3	ARPC3	0
Actin-related protein 2/3 complex subunit 4	ARPC4-TTLL3;ARPC4	0
Actin-related protein 2/3 complex subunit 5	ARPC5	0
Actin-related protein 3;Actin-related protein 3B	ACTR3;ACTR3B	0
Activated RNA polymerase II transcriptional coactivator p15	SUB1	0
Activin receptor type-1	ACVR1	0
Activin receptor type-1B	ACVR1B	8.7997E-277
Acylamino-acid-releasing enzyme	APEH	0

Acyl-CoA-binding protein	DBI	0
Acyl-CoA-binding protein	DBI	3.7021E-107
Acylphosphatase;Acylphosphatase-1	ACYP1	0
ADAM DEC1	ADAMDEC1	0
ADAMTS-like protein 4	ADAMTSL4	0
Adapter molecule crk	CRK	0
Adenosine kinase	ADK	6.2999E-191
Adenylate kinase 2, mitochondrial	AK2	0
Adenylyl cyclase-associated protein 1	CAP1	0
Adiponectin	ADIPOQ	0
Adipose most abundant gene transcript 2 protein	APM2	0
ADM;Adrenomedullin;Proadrenomedullin N-20 terminal peptide	ADM	4.7254E-91
ADP-ribosyl cyclase 2	BST1	1.8618E-138
ADP-ribosylation factor 1;ADP-ribosylation factor 3;ADP-ribosylation factor 5	ARF1;ARF3;ARF5	1.0621E-194
ADP-ribosylation factor 3;ADP-ribosylation factor 1;ADP-ribosylation factor 5	ARF3;ARF1;ARF5	3.9574E-95
ADP-ribosylation factor-binding protein GGA3;ADP-ribosylation factor-binding protein GGA1;ADP-ribosylation factor-binding protein GGA2	GGA3;GGA1;GGA2	6.3522E-74
ADP-ribosylation factor-like protein 8B;ADP-ribosylation factor-like protein 8A	ARL8B;ARL8A	4.4158E-138
ADP-sugar pyrophosphatase	NUDT5	0
Advanced glycosylation end product-specific receptor	AGER	0
Afamin	AFM	1.079E-250
Aflatoxin B1 aldehyde reductase member 2;Aflatoxin B1 aldehyde reductase member 4	AKR7A2;AKR7L	0
Aggrecan core protein;Aggrecan core protein 2	ACAN	3.6658E-146
Agrin	AGRN	2.3072E-301
Agrin	AGRN	0
Alcohol dehydrogenase [NADP(+)]	AKR1A1	0
Aldehyde dehydrogenase, dimeric NADP-preferring	ALDH3A1	0
Allograft inflammatory factor 1	AIF1	3.422E-303
Alpha-1-acid glycoprotein 1	ORM1	4.4852E-76
Alpha-1-acid glycoprotein 2	ORM2	0

Alpha-1-antichymotrypsin;Alpha-1-antichymotrypsin His-Pro-less	SERPINA3	0
Alpha-1-antitrypsin;Short peptide from AAT	SERPINA1	6.737E-122
Alpha-1B-glycoprotein	A1BG	1.1203E-32
Alpha-2-antiplasmin	SERPINF2	0
Alpha-2-HS-glycoprotein;Alpha-2-HS-glycoprotein chain A;Alpha-2-HS-glycoprotein chain B	AHSG	0
Alpha-2-macroglobulin	A2M	0
Alpha-actinin-1	ACTN1	1.5173E-181
Alpha-actinin-4	ACTN4	3.7378E-52
Alpha-amylase 1	AMY1A	0
Alpha-amylase 2B	AMY2B	9.9925E-239
Alpha-enolase	ENO1	0
Alpha-galactosidase A	GLA	0
Alpha-hemoglobin-stabilizing protein	AHSP	0
Alpha-N-acetylgalactosaminidase	NAGA	0
Alpha-N-acetylglucosaminidase;Alpha-N-acetylglucosaminidase 82 kDa form;Alpha-N-acetylglucosaminidase 77 kDa form	NAGLU	7.4222E-27
Alpha-synuclein	SNCA	0
Aminoacylase-1	ACY1	4.5883E-15
Aminopeptidase N	ANPEP	3.2586E-279
Amphiphysin	AMPH	3.3222E-22
Amyloid beta A4 precursor protein-binding family B member 1-interacting protein	APBB1IP	5.1119E-22
Amyloid beta A4 protein;N-APP;Soluble APP-alpha;Soluble APP-beta;C99;Beta-amyloid protein 42;Beta-amyloid protein 40;C83;P3(42);P3(40);C80;Gamma-secretase C-terminal fragment 59;Gamma-secretase C-terminal fragment 57;Gamma-secretase C-terminal fragment 50;C31	APP	5.8527E-40
Amyloid-like protein 1;C30	APLP1	5.9976E-22
Angiopoietin-related protein 2	ANGPTL2	1.6371E-54
Angiotensinogen;Angiotensin-1;Angiotensin-2;Angiotensin-3	AGT	3.7088E-88
Annexin A1;Annexin	ANXA1	0
Annexin A11;Annexin	ANXA11	0

Annexin A2;Putative annexin A2-like protein;Annexin	ANXA2;ANXA2P2	1.8197E-14
Annexin A3;Annexin	ANXA3	0
Annexin A4;Annexin	ANXA4	2.7176E-229
Annexin A5;Annexin	ANXA5	1.2691E-292
Annexin A6;Annexin	ANXA6	4.2485E-42
Anthrax toxin receptor 1	ANTXR1	4.214E-20
Antigen KI-67	MKI67	0
Antithrombin-III	SERPINC1	8.9024E-132
AP-1 complex subunit beta-1	AP1B1	0
AP-1 complex subunit beta-1	AP1B1	2.7676E-26
AP-2 complex subunit alpha-2	AP2A2	3.1541E-67
AP-2 complex subunit beta	AP2B1	2.3011E-121
Apolipoprotein A-I;Truncated apolipoprotein A-I	APOA1	0
Apolipoprotein A-II;Truncated apolipoprotein A-II	APOA2	0
Apolipoprotein A-IV	APOA4	0
Apolipoprotein B receptor	APOBR	1.1688E-11
Apolipoprotein B-100;Apolipoprotein B-48	APOB	4.4871E-88
Apolipoprotein C-I;Truncated apolipoprotein C-I	APOC1	0
Apolipoprotein C-II	APOC2	0
Apolipoprotein C-III	APOC3	0
Apolipoprotein D	APOD	2.0536E-26
Apolipoprotein E	APOE	6.7925E-49
Apolipoprotein F	APOF	4.194E-150
Apolipoprotein L1	APOL1	5.0593E-177
Apolipoprotein M	APOM	1.2317E-82
Apolipoprotein(a)	LPA	8.853E-301
Apoptosis regulator BAX	BAX	0
Apoptosis-associated speck-like protein containing a CARD	PYCARD	7.192E-219
Aquaporin-1	AQP1	5.959E-69
Aquaporin-2	AQP2	0

ARF GTPase-activating protein GIT2	GIT2	3.3989E-60
Arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein 2	ACAP2	0
Arginase-1	ARG1	0
Argininosuccinate synthase	ASS1	1.1083E-45
Arylsulfatase A;Arylsulfatase A component B;Arylsulfatase A component C	ARSA	8.6758E-210
Arylsulfatase B	ARSB	1.0521E-240
Asialoglycoprotein receptor 1	ASGR1	1.9263E-15
Asialoglycoprotein receptor 2	ASGR2	2.9298E-35
Aspartate aminotransferase, cytoplasmic;Aspartate aminotransferase	GOT1	8.242E-52
Aspartate--tRNA ligase, cytoplasmic	DARS;DKFZp781B1120 2	0
ATP synthase subunit alpha, mitochondrial	ATP5A1	0
ATPase family AAA domain-containing protein 2	ATAD2	0
ATP-citrate synthase	ACLY	3.3413E-55
ATP-dependent RNA helicase A	DHX9	1.1765E-281
ATP-dependent RNA helicase DDX39A;Spliceosome RNA helicase DDX39B	DDX39A;DDX39B	3.0238E-171
Attractin	ATRN	2.737E-218
Azurocidin	AZU1	0
B- and T-lymphocyte attenuator	BTLA	1.374E-09
Bactericidal permeability-increasing protein	BPI	9.2744E-12
BAG family molecular chaperone regulator 1	BAG1	2.9355E-61
Barrier-to-autointegration factor	BANF1	4.4582E-09
Basal cell adhesion molecule	BCAM	9.0526E-18
Basement membrane-specific heparan sulfate proteoglycan core protein;Endorepellin;LG3 peptide	HSPG2	0
Basigin	BSG	2.3149E-221
Bcl-2-like protein 15	BCL2L15	2.7898E-215
Beta-1,3-N-acetylglucosaminyltransferase lunatic fringe	LFNG	7.4367E-19
Beta-2-glycoprotein 1	APOH	0
Beta-2-microglobulin;Beta-2-microglobulin form pI 5.3	B2M	5.1684E-220
Beta-actin-like protein 2	ACTBL2	8.6037E-32

Beta-defensin 1	DEFB1	5.7604E-14
Beta-galactosidase	GLB1	1.4405E-66
Beta-galactoside alpha-2,6-sialyltransferase 2	ST6GAL2	1.833E-25
Beta-glucuronidase	GUSB	1.2402E-11
Beta-hexosaminidase subunit alpha	HEXA	0
Beta-hexosaminidase subunit beta;Beta-hexosaminidase subunit beta chain B;Beta-hexosaminidase subunit beta chain A	HEXB	0
Betaine--homocysteine S-methyltransferase 1;S-methylmethionine--homocysteine S-methyltransferase BHMT2	BHMT;BHMT2	1.6039E-18
Beta-mannosidase	MANBA	1.1571E-09
Beta-microseminoprotein	MSMB	8.433E-20
BH3-interacting domain death agonist;BH3-interacting domain death agonist p15;BH3-interacting domain death agonist p13;BH3-interacting domain death agonist p11	BID	0
Bifunctional glutamate/proline--tRNA ligase;Glutamate--tRNA ligase;Proline--tRNA ligase	EPRS	1.7488E-16
Bifunctional protein NCOAT;Protein O-GlcNAcase;Histone acetyltransferase	MGEA5	8.7527E-107
Biglycan	BGN	1.7946E-173
Bile salt-activated lipase	CEL	1.5165E-25
Biotinidase	BTD	6.6943E-75
Bisphosphoglycerate mutase	BPGM	0
BMP and activin membrane-bound inhibitor homolog	BAMBI	8.9032E-57
BolA-like protein 2	BOLA2	0
Bone marrow proteoglycan;Eosinophil granule major basic protein	PRG2	1.1486E-17
BPI fold-containing family B member 1	BPIFB1	4.8951E-36
BPI fold-containing family B member 2	BPIFB2	1.8961E-97
Brain acid soluble protein 1	BASP1	1.2942E-248
Brevican core protein	BCAN	3.7003E-269
Bridging integrator 2	BIN2	2.838E-09
BTB/POZ domain-containing protein KCTD12	KCTD12	0.00043273
Butyrophilin subfamily 2 member A1;Butyrophilin subfamily 2 member A2	BTN2A1;BTN2A2	2.5181E-61
C4b-binding protein alpha chain	C4BPA	4.1417E-50
C4b-binding protein beta chain	C4BPB	0.0013202

Cadherin-1;E-Cad/CTF1;E-Cad/CTF2;E-Cad/CTF3	CDH1	1.064E-37
Cadherin-11	CDH11	0
Cadherin-13	CDH13	2.1563E-42
Cadherin-13	CDH13	4.8715E-73
Cadherin-15	CDH15	1.0457E-10
Cadherin-17	CDH17	9.7501E-35
Cadherin-2	CDH2	1.5573E-22
Cadherin-3	CDH3	1.272E-10
Cadherin-5	CDH5	0.00028705
Cadherin-6	CDH6	9.9298E-101
Cadherin-related family member 2	CDHR2	1.3061E-108
Cadherin-related family member 5	CDHR5	3.4565E-120
Calbindin	CALB1	6.0388E-40
Calcineurin subunit B type 1	PPP3R1	0.00015823
Calcineurin-like phosphoesterase domain-containing protein 1	CPPED1	1.4696E-18
Calcitonin gene-related peptide 2	CALCB	1.2714E-06
Calcium/calmodulin-dependent protein kinase type II subunit gamma	CAMK2G	1.7103E-184
Calcium-binding protein 39	CAB39	7.3828E-21
Calcium-binding protein p22	CHP	2.7578E-141
Calcyclin-binding protein	CACYBP	2.3067E-81
Calmodulin	CALM2;CALM1	0
Calmodulin-like protein 5	CALML5	2.4381E-68
Calnexin	CANX	0
Calpain small subunit 1;Calpain small subunit 2	CAPNS1;CAPNS2	0
Calpain-1 catalytic subunit	CAPN1	1.9591E-09
Calpastatin	CAST	0
Calponin-2	CNN2	4.2962E-09
Calreticulin	CALR	2.1587E-24
Calsyntenin-1;Soluble Alc-alpha;CTF1-alpha	CLSTN1	4.0193E-223
Calsyntenin-3	CLSTN3	5.2535E-27

cAMP-dependent protein kinase type I-alpha regulatory subunit	PRKAR1A	0
cAMP-dependent protein kinase type II-alpha regulatory subunit;cAMP-dependent protein kinase type II-beta regulatory subunit	PRKAR2A;PRKAR2B	0
CAP-Gly domain-containing linker protein 1	CLIP1	1.1719E-61
Caprin-1	CAPRIN1	0
CapZ-interacting protein	RCSD1	4.6809E-169
Carbonic anhydrase 1	CA1	0
Carbonic anhydrase 2	CA2	0.000085474
Carboxypeptidase E	CPE	3.4579E-06
Carboxypeptidase M	CPM	1.4718E-11
Carboxypeptidase N catalytic chain	CPN1	0
Carboxypeptidase N subunit 2	CPN2	3.8235E-06
Carboxypeptidase Q	CPQ;PGCP	7.0641E-298
Carboxypeptidase Z	CPZ	0.0011261
Carcinoembryonic antigen-related cell adhesion molecule 1	CEACAM1	0.00088215
Carcinoembryonic antigen-related cell adhesion molecule 6	CEACAM6	2.6401E-304
Carcinoembryonic antigen-related cell adhesion molecule 7	CEACAM7	4.3192E-64
Carcinoembryonic antigen-related cell adhesion molecule 8	CEACAM8	4.7379E-52
Cartilage intermediate layer protein 1;Cartilage intermediate layer protein 1 C1;Cartilage intermediate layer protein 1 C2	CILP	1.8497E-23
Cartilage intermediate layer protein 2;Cartilage intermediate layer protein 2 C1;Cartilage intermediate layer protein 2 C2	CILP2	0.0013008
Cartilage oligomeric matrix protein	COMP	1.8764E-48
Caspase-14;Caspase-14 subunit p19;Caspase-14 subunit p10	CASP14	0
Caspase-7;Caspase-7 subunit p20;Caspase-7 subunit p11	CASP7	4.5229E-210
Caspase-8;Caspase-8 subunit p18;Caspase-8 subunit p10	CASP8	4.6827E-241
Catalase	CAT	2.146E-17
Cathelicidin antimicrobial peptide;Antibacterial protein FALL-39;Antibacterial protein LL-37	CAMP	5.7869E-95
Cathepsin B;Cathepsin B light chain;Cathepsin B heavy chain	CTSB	3.5126E-232
Cathepsin D;Cathepsin D light chain;Cathepsin D heavy chain	CTSD	4.1633E-66

Cathepsin G	CTSG	6.654E-07
Cathepsin L1;Cathepsin L1 heavy chain;Cathepsin L1 light chain;Cathepsin L2;Cathepsin K;Putative cathepsin L-like protein 3	CTSL1;CTSL2;CTSK;CTSL3	2.7608E-44
Cathepsin S	CTSS	1.8146E-19
Cathepsin Z	CTSZ	7.4799E-10
Cation-independent mannose-6-phosphate receptor	IGF2R	1.3476E-22
CB1 cannabinoid receptor-interacting protein 1	CNRIP1	0
CD166 antigen	ALCAM	0.000021672
CD177 antigen	CD177	0.0070671
CD209 antigen	CD209	0
CD27 antigen	CD27	5.9286E-116
CD276 antigen	CD276	0
CD320 antigen	CD320	4.535E-157
CD44 antigen	CD44	1.1526E-264
CD5 antigen-like	CD5L	4.3109E-139
CD59 glycoprotein	CD59	0.0020285
CD63 antigen	CD63	1.0685E-71
CD9 antigen	CD9	0
CD97 antigen;CD97 antigen subunit alpha;CD97 antigen subunit beta	CD97	0.0023201
CD99 antigen	CD99	0
CD99 antigen-like protein 2	CD99L2	9.455E-239
Cell adhesion molecule 1	CADM1	4.3589E-10
Cell adhesion molecule 1	CADM1	9.586E-12
Cell adhesion molecule 2	CADM2	0.00086539
Cell adhesion molecule 3	CADM3	1.4065E-17
Cell adhesion molecule 4	CADM4	6.8362E-214
Cell division control protein 42 homolog;Rho-related GTP-binding protein RhoQ	CDC42;RHOQ	1.1453E-87
Cell division cycle 5-like protein	CDC5L	1.6184E-49
Cell surface glycoprotein MUC18	MCAM	1.0091E-247
Ceramide kinase	CERK	5.8181E-31

Ceroid-lipofuscinosis neuronal protein 5	CLN5	1.5031E-31
Ceruloplasmin	CP	2.4758E-09
Charged multivesicular body protein 1b	CHMP1B	1.0362E-29
Charged multivesicular body protein 5	CHMP5	2.5136E-29
Chitinase-3-like protein 1	CHI3L1	0.00001131
Chitotriosidase-1	CHIT1	0.0001223
Chloride intracellular channel protein 1	CLIC1	1.292E-08
Cholesteryl ester transfer protein	CETP	2.735E-07
Choline transporter-like protein 4	SLC44A4	4.2186E-69
Chondroitin sulfate proteoglycan 4	CSPG4	3.3028E-06
Chondrolectin	CHODL	5.7116E-09
Chromobox protein homolog 1	CBX1	0.00071335
Chromobox protein homolog 3	CBX3	0.00043571
Chromodomain-helicase-DNA-binding protein 6;Chromodomain-helicase-DNA-binding protein 7;Chromodomain-helicase-DNA-binding protein 9;Chromodomain-helicase-DNA-binding protein 8	CHD6;CHD7;CHD9;CHD8	0.000040649
Chromogranin-A;Vasostatin-1;Vasostatin-2;EA-92;ES-43;Pancreastatin;SS-18;WA-8;WE-14;LF-19;AL-11;GV-19;GR-44;ER-37	CHGA	0
Ciliary neurotrophic factor receptor subunit alpha	CNTFR	7.6355E-122
Clathrin interactor 1	CLINT1	0.000043735
Cleavage and polyadenylation specificity factor subunit 5	NUDT21	3.5335E-162
Cleavage and polyadenylation specificity factor subunit 6	CPSF6	0
Clusterin;Clusterin beta chain;Clusterin alpha chain	CLU	1.2011E-06
CMRF35-like molecule 1	CD300LF	4.3015E-185
CMRF35-like molecule 2	CD300E	3.8929E-16
CMRF35-like molecule 8	CD300A	3.4892E-163
CMRF35-like molecule 9	CD300LG	3.1298E-101
C-Myc-binding protein	MYCBP	5.5892E-40
Coactosin-like protein	COTL1	6.6282E-27
Coagulation factor IX;Coagulation factor IXa light chain;Coagulation factor IXa heavy chain	F9	0
Coagulation factor V;Coagulation factor V heavy chain;Coagulation factor V light chain	F5	7.7082E-23

Coagulation factor X;Factor X light chain;Factor X heavy chain;Activated factor Xa heavy chain	F10	1.7324E-18
Coagulation factor XII;Coagulation factor XIIa heavy chain;Beta-factor XIIa part 1;Beta-factor XIIa part 2;Coagulation factor XIIa light chain	F12	0
Coagulation factor XIII B chain	F13B	5.3092E-13
Coatomer subunit epsilon	COPE	4.4625E-134
Cochlin	COCH	0
Cofilin-1	CFL1	3.2369E-47
Coiled-coil and C2 domain-containing protein 1A	CC2D1A	2.4259E-80
Coiled-coil domain-containing protein 58	CCDC58	2.1582E-42
Coiled-coil domain-containing protein 88B	CCDC88B	2.8582E-122
Colipase	CLPS	4.6633E-105
Collagen alpha-1(I) chain	COL1A1	1.9198E-31
Collagen alpha-1(II) chain;Collagen alpha-1(II) chain;Chondrocalcin	COL2A1	0
Collagen alpha-1(III) chain	COL3A1	3.0607E-105
Collagen alpha-1(VI) chain	COL6A1	0
Collagen alpha-1(VII) chain	COL7A1	7.6842E-182
Collagen alpha-1(XII) chain	COL12A1	3.9263E-07
Collagen alpha-1(XIV) chain	COL14A1	0
Collagen alpha-1(XV) chain;Endostatin	COL15A1	4.5735E-11
Collagen alpha-1(XVII) chain;120 kDa linear IgA disease antigen;97 kDa linear IgA disease antigen	COL17A1	4.7633E-153
Collagen alpha-1(XVIII) chain;Endostatin	COL18A1	1.5413E-17
Collagen alpha-1(XXVIII) chain	COL28A1	5.4903E-276
Collagen alpha-2(I) chain	COL1A2	0
Collagen alpha-2(IV) chain;Canstatin	COL4A2	0
Collagen alpha-2(VI) chain	COL6A2	0
Collagen alpha-3(V) chain	COL5A3	4.2522E-97
Collagen alpha-3(VI) chain	COL6A3	6.8642E-52
Collagen alpha-6(VI) chain	COL6A6	9.985E-47
Collectin-12	COLEC12	6.8032E-123

Collectrin	TMEM27	1.7333E-23
COMM domain-containing protein 9	COMMD9	2.1484E-85
Complement C1q subcomponent subunit B	C1QB	6.7298E-56
Complement C1q subcomponent subunit C	C1QC	0
Complement C1r subcomponent;Complement C1r subcomponent heavy chain;Complement C1r subcomponent light chain	C1R	1.2751E-272
Complement C1r subcomponent-like protein	C1RL	6.9164E-79
Complement C1s subcomponent;Complement C1s subcomponent heavy chain;Complement C1s subcomponent light chain	C1S	8.1433E-160
Complement C3;Complement C3 beta chain;Complement C3 alpha chain;C3a anaphylatoxin;Complement C3b alpha chain;Complement C3c alpha chain fragment 1;Complement C3dg fragment;Complement C3g fragment;Complement C3d fragment;Complement C3f fragment;Complement C3c alpha chain fragment 2	C3	6.6394E-15
Complement C4-A;Complement C4 beta chain;Complement C4-A alpha chain;C4a anaphylatoxin;C4b-A;C4d-A;Complement C4 gamma chain	C4A	3.127E-64
Complement C4-B;Complement C4 beta chain;Complement C4-B alpha chain;C4a anaphylatoxin;C4b-B;C4d-B;Complement C4 gamma chain	C4B	0
Complement C5;Complement C5 beta chain;Complement C5 alpha chain;C5a anaphylatoxin;Complement C5 alpha chain	C5	1.7808E-109
Complement component C6	C6	8.316E-152
Complement component C7	C7	0
Complement component C8 alpha chain	C8A	3.77E-13
Complement component C8 beta chain	C8B	1.005E-203
Complement component C8 gamma chain	C8G	5.1984E-06
Complement component C9;Complement component C9a;Complement component C9b	C9	0
Complement decay-accelerating factor	CD55	4.3108E-06
Complement decay-accelerating factor	CD55;DAF	4.0769E-61
Complement factor B;Complement factor B Ba fragment;Complement factor B Bb fragment	CFB	0.00051934
Complement factor D	CFD	0.000089278
Complement factor H	CFH	1.641E-16
Complement factor H-related protein 1	CFHR1	0.00030524

Complement factor H-related protein 2	CFHR2	9.9452E-79
Complement factor I;Complement factor I heavy chain;Complement factor I light chain	CFI	3.0742E-29
Complement receptor type 1	CR1	2.5341E-15
Connective tissue growth factor	CTGF	7.5469E-185
Conserved oligomeric Golgi complex subunit 5	COG5	0
Constitutive coactivator of PPAR-gamma-like protein 1	FAM120A	2.3929E-56
Contactin-1	CNTN1	0.0025079
COP9 signalosome complex subunit 2	COPS2	0.0071316
COP9 signalosome complex subunit 4	COPS4	0.00005602
Copine-1	CPNE1	0.0012962
Copine-3;Copine-7;Copine-5;Copine-8;Copine-6;Copine-4;Copine-9;Copine-2	CPNE3;CPNE7;CPNE6; CPNE5;CPNE8;CPNE4; CPNE9;CPNE2	0.000019374
Copper transport protein ATOX1	ATOX1	0
Core histone macro-H2A.1;Histone H2A	H2AFY	1.3509E-127
Corneodesmosin	CDSN	4.8434E-23
Cornifin-A	SPRR1A	8.6769E-18
Cornifin-B	SPRR1B	1.244E-11
Cornulin	CRNN	0.0019581
Coronin-1A	CORO1A	0.000025574
Coronin-1C	CORO1C	7.6442E-65
Corticosteroid-binding globulin	SERPINA6	1.527E-07
Costars family protein ABRACL	ABRACL;RP11- 501K14.2	0.00060914
Coxsackievirus and adenovirus receptor	CXADR	1.1561E-13
Creatine kinase B-type	CKB	0.000089488
Crk-like protein	CRKL	1.2661E-07
C-type lectin domain family 14 member A	CLEC14A	1.1969E-06
C-type lectin domain family 4 member G	CLEC4G	0.00092566
C-type lectin domain family 5 member A	CLEC5A	2.545E-11
C-type lectin domain family 7 member A	CLEC7A	4.39E-265

C-type mannose receptor 2	MRC2	4.5244E-21
Cubilin	CUBN	0.00069021
CUGBP Elav-like family member 2	CELF2;CUGBP2	0.0011696
CXADR-like membrane protein	CLMP	0.0014063
C-X-C chemokine receptor type 2	CXCR2	6.2351E-06
C-X-C motif chemokine 16	CXCL16	4.9523E-07
Cyclic AMP-responsive element-binding protein 3-like protein 3;Processed cyclic AMP-responsive element-binding protein 3-like protein 3	CREB3L3	2.229E-23
Cystatin-A	CSTA	0
Cystatin-B	CSTB	0.00068683
Cystatin-C	CST3	5.7733E-11
Cystatin-M	CST6	3.0505E-111
Cystatin-S	CST4	1.156E-269
Cystatin-SA	CST2	0.00021603
Cysteine and histidine-rich domain-containing protein 1	CHORDC1	0.000080462
Cysteine-rich secretory protein 3	CRISP3	0.0011341
Cysteine-rich secretory protein LCCL domain-containing 2;Peptidase inhibitor 15	CRISPLD2;PI15	4.4793E-11
Cytidine deaminase	CDA	0.0054199
Cytochrome b-245 heavy chain	CYBB	0.00068035
Cytochrome b5 reductase 4	CYB5R4	9.7416E-95
Cytochrome c	CYCS	0.000013692
Cytochrome c oxidase subunit 6B1	COX6B1	0.00016103
Cytohesin-1	CYTH1	5.5791E-25
Cytoplasmic dynein 1 intermediate chain 2	DYNC1I2	1.0218E-13
Cytosolic non-specific dipeptidase	CNDP2	2.3156E-20
Cytosolic purine 5-nucleotidase	NT5C2	0.00019437
DCC-interacting protein 13-alpha	APPL1	0.000051303
DCN1-like protein 1	DCUN1D1	1.4951E-23
D-dopachrome decarboxylase;D-dopachrome decarboxylase-like protein	DDT;DDTL	6.0306E-12
Decorin	DCN	6.8606E-08

Deleted in malignant brain tumors 1 protein	DMBT1	5.2302E-08
Delta and Notch-like epidermal growth factor-related receptor	DNER	4.2461E-09
Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial	ECH1	0.0018278
Delta-aminolevulinic acid dehydratase	ALAD	2.6913E-194
Dematin	EPB49	0
Deoxyribonuclease-1	DNASE1	3.9561E-11
Deoxyribonuclease-2-alpha	DNASE2	6.0707E-07
Dermatopontin	DPT	0.00028807
Dermcidin;Survival-promoting peptide;DCD-1	DCD	6.7564E-18
Dermokine	DMKN	1.3815E-57
Desmocollin-1	DSC1	1.3412E-17
Desmocollin-2	DSC2	8.323E-20
Desmoglein-1	DSG1	0
Desmoglein-2	DSG2	8.826E-08
Deubiquitinating protein VCIP135	VCPIP1	3.5621E-06
Differentially expressed in FDCP 6 homolog	DEF6	2.3209E-283
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	1.8341E-10
Dihydropteridine reductase	QDPR	1.1513E-53
Di-N-acetylchitobiase	CTBS	3.2481E-09
Dipeptidase 1	DPEP1	0.00056324
Dipeptidyl peptidase 1;Dipeptidyl peptidase 1 exclusion domain chain;Dipeptidyl peptidase 1 heavy chain;Dipeptidyl peptidase 1 light chain	CTSC	3.3747E-18
Dipeptidyl peptidase 2	DPP7	3.8566E-08
Dipeptidyl peptidase 4;Dipeptidyl peptidase 4 membrane form;Dipeptidyl peptidase 4 soluble form	DPP4	1.9468E-06
Diphosphoinositol polyphosphate phosphohydrolase 1	NUDT3	1.8662E-206
Discoidin domain-containing receptor 2	DDR2	6.8617E-41
Disintegrin and metalloproteinase domain-containing protein 8	ADAM8	9.505E-09
DNA damage-binding protein 1	DDB1	1.077E-07
DNA primase large subunit	PRIM2	5.134E-84

DNA-(apurinic or apyrimidinic site) lyase;DNA-(apurinic or apyrimidinic site) lyase, mitochondrial	APEX1	2.908E-12
DNA-binding protein A;Y-box-binding protein 2;Nuclease-sensitive element-binding protein 1	CSDA;YBX2;YBX1	3.0843E-43
DnaJ homolog subfamily A member 2	DNAJA2	0.00024346
DnaJ homolog subfamily C member 8	DNAJC8	0.000016088
Docking protein 3	DOK3	5.5165E-28
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	RPN2	1.1716E-21
Double-strand-break repair protein rad21 homolog	RAD21	2.2661E-33
Drebrin-like protein	DBNL	2.0403E-33
Dual specificity mitogen-activated protein kinase kinase 1	MAP2K1	0.000049618
Dynactin subunit 1	DCTN1;DKFZp686E0752	0.000096516
Dynactin subunit 2	DCTN2	8.5197E-09
Dynamin-2	DNM2	3.9227E-06
Dynein light chain 2, cytoplasmic	DYNLL2	1.4768E-65
Dysferlin	DYSF	0.0013984
Dyslexia-associated protein KIAA0319-like protein	KIAA0319L	4.1036E-33
Dystroglycan;Alpha-dystroglycan;Beta-dystroglycan	DAG1	1.9513E-07
E3 ubiquitin-protein ligase RNF149	RNF149	0.0023443
E3 ubiquitin-protein ligase RNF167	RNF167	0.00037937
Early endosome antigen 1	EEA1	0.0011153
Ecto-ADP-ribosyltransferase 3	ART3	7.8037E-20
Ectonucleotide pyrophosphatase/phosphodiesterase family member 2	ENPP2	7.4086E-09
EF-hand domain-containing protein D2	EFHD2	8.3892E-10
EF-hand domain-containing protein KIAA0494	KIAA0494	1.3659E-17
EGF-containing fibulin-like extracellular matrix protein 1	EFEMP1	1.1816E-26
EGF-containing fibulin-like extracellular matrix protein 2	EFEMP2	0.000057632
EH domain-containing protein 1	EHD1	1.6787E-246
Elafin	PI3	6.5493E-24
ELAV-like protein 1	ELAVL1	1.758E-30

Elongation factor 1-alpha 1;Putative elongation factor 1-alpha-like 3;Elongation factor 1-alpha 2	EEF1A1;EEF1A1P5;EEF1A2	7.4048E-21
Elongation factor 1-beta	EEF1B2	0.0011448
Elongation factor 1-delta	EEF1D	2.5001E-78
Elongation factor 1-gamma	EEF1G	0.0018233
Elongation factor 2	EEF2	1.6566E-90
Endoglin	ENG	1.8159E-09
Endonuclease domain-containing 1 protein	ENDOD1	2.3242E-16
Endoplasmic reticulum aminopeptidase 1	ERAP1	2.634E-63
Endoplasmic reticulum resident protein 29	ERP29	0.00051416
Endoplasmic reticulum resident protein 44	ERP44	1.3832E-10
Endoplasmin	HSP90B1	0.0011527
Endosialin	CD248	0.0011129
Endothelial cell-selective adhesion molecule	ESAM	3.696E-13
Endothelial differentiation-related factor 1	EDF1	1.7643E-16
Endothelial protein C receptor	PROCR	2.9527E-23
Endothelin-3	EDN3	1.3705E-22
Enhancer of rudimentary homolog	ERH	5.3626E-06
Enolase-phosphatase E1	ENOPH1	6.4257E-151
Enoyl-CoA delta isomerase 1, mitochondrial	ECI1;DCI	1.7394E-10
Enoyl-CoA hydratase, mitochondrial	ECHS1	2.5679E-26
Eosinophil cationic protein	RNASE3	3.2734E-45
Eosinophil lysophospholipase	CLC	0.000027303
Ephrin type-A receptor 1	EPHA1	1.0512E-08
Ephrin type-A receptor 2	EPHA2	5.1309E-07
Ephrin type-A receptor 7	EPHA7	0.00062473
Ephrin type-B receptor 2	EPHB2	0.00034409
Ephrin type-B receptor 3;Ephrin type-B receptor 1	EPHB3;EPHB1	0.00068302
Ephrin type-B receptor 4	EPHB4	0.00019137
Ephrin type-B receptor 6	EPHB6	5.9068E-16

Ephrin-A1;Ephrin-A1, secreted form	EFNA1	6.5681E-07
Ephrin-A5	EFNA5	0.000017125
Ephrin-B1	EFNB1	9.487E-27
Epidermal growth factor receptor substrate 15	EPS15	7.6207E-06
Epidermal growth factor receptor substrate 15-like 1	EPS15L1	2.2696E-25
Epididymal secretory protein E1	NPC2	4.4077E-143
Epididymal secretory protein E1	NPC2	6.9E-16
Epididymis-specific alpha-mannosidase	MAN2B2	1.1004E-164
Epithelial cell adhesion molecule	EPCAM	0.00017355
Epithelial discoidin domain-containing receptor 1	DDR1	1.2955E-15
ERO1-like protein alpha	ERO1L	0.0013659
Ester hydrolase C11orf54	C11orf54	0.00068416
Eukaryotic translation initiation factor 2 subunit 1	EIF2S1	0.00063521
Eukaryotic translation initiation factor 2 subunit 2	EIF2S2	0.000031373
Eukaryotic translation initiation factor 4B	EIF4B	6.2208E-44
Eukaryotic translation initiation factor 5	EIF5	7.0326E-06
Eukaryotic translation initiation factor 6	EIF6	3.223E-28
Exportin-2	CSE1L	0.0011425
Extended synaptotagmin-1	ESYT1	4.6436E-08
Extended synaptotagmin-2	ESYT2	1.5564E-125
Extracellular matrix protein 1	ECM1	0.00074098
Extracellular sulfatase Sulf-2	SULF2	2.1854E-08
Extracellular superoxide dismutase [Cu-Zn]	SOD3	1.1242E-27
Ezrin	EZR	4.0318E-07
F-actin-capping protein subunit alpha-1;F-actin-capping protein subunit alpha-2	CAPZA1;CAPZA2	3.4769E-09
F-actin-capping protein subunit beta	CAPZB	0.000069605
F-actin-capping protein subunit beta	CAPZB	0.00080035
Far upstream element-binding protein 2	KHSRP	0.0018717
Farnesyl pyrophosphate synthase	FDPS	2.556E-97
Fatty acid synthase;[Acyl-carrier-protein] S-acetyltransferase;[Acyl-carrier-protein] S-	FASN	2.4152E-07

malonyltransferase;3-oxoacyl-[acyl-carrier-protein] synthase;3-oxoacyl-[acyl-carrier-protein] reductase;3-hydroxypalmitoyl-[acyl-carrier-protein] dehydratase;Enoyl-[acyl-carrier-protein] reductase;Oleoyl-[acyl-carrier-protein] hydrolase		
Fatty acid-binding protein, adipocyte	FABP4	5.6372E-52
Fatty acid-binding protein, epidermal	FABP5	6.6574E-25
Fatty acid-binding protein, heart	FABP3	0.000016988
Fatty acid-binding protein, liver	FABP1	9.2134E-08
F-box only protein 10	FBXO10	2.6443E-78
Fermitin family homolog 3	FERMT3	0.000001802
Ferritin light chain	FTL	6.9974E-16
Fetuin-B	FETUB	9.7221E-07
Fibrillin-1	FBN1	0.00010972
Fibrillin-2	FBN2	0.000024509
Fibrinogen alpha chain;Fibrinopeptide A;Fibrinogen alpha chain	FGA	3.4457E-10
Fibrinogen beta chain;Fibrinopeptide B;Fibrinogen beta chain	FGB	0.0012524
Fibrinogen gamma chain	FGG	3.5632E-189
Fibroblast growth factor receptor 1;Fibroblast growth factor receptor	FGFR1	2.0659E-145
Fibroblast growth factor receptor 2	FGFR2	2.5345E-07
Fibroblast growth factor-binding protein 2	FGFBP2	0.00025338
Fibrocystin	PKHD1	0.0000734
Fibronectin;Anastellin;Ugl-Y1;Ugl-Y2;Ugl-Y3	FN1	0.00022904
Fibulin-1	FBLN1	0.000057538
Fibulin-1	FBLN1	5.9976E-06
Fibulin-2	FBLN2	1.5528E-142
Fibulin-5	FBLN5	8.5842E-37
Fibulin-7	FBLN7	1.2778E-06
Ficolin-1	FCN1	3.0781E-76
Ficolin-2	FCN2	1.0444E-06
Ficolin-3	FCN3	4.3331E-76
Filaggrin	FLG	6.8142E-10
Filamin-A	FLNA	1.66E-51

Filamin-A	FLNA	6.5673E-14
Filamin-B	FLNB	1.1551E-237
Filamin-C	FLNC	1.1719E-73
FK506-binding protein 15	FKBP15	2.2704E-31
Flap endonuclease 1	FEN1	6.7153E-30
Flavin reductase (NADPH)	BLVRB	1.2169E-56
Flotillin-2	FLOT2	1.0813E-213
Folate receptor alpha	FOLR1	0
Folate receptor beta	FOLR2	0.00077716
Folate receptor gamma	FOLR3	0
Folate receptor gamma;Folate receptor beta	FOLR3;FOLR2	1.0877E-22
Follistatin-related protein 1	FSTL1	1.5852E-17
Forkhead box protein P2	FOXP2	8.4181E-09
Formin-binding protein 1	FNBP1	0.00039136
Formin-like protein 1	FMNL1	9.3161E-74
FRAS1-related extracellular matrix protein 2	FREM2	5.8766E-16
Frizzled-4	FZD4	1.2473E-34
Frizzled-7;Frizzled-2	FZD7;FZD2	6.8357E-06
Fructose-1,6-bisphosphatase 1	FBP1	3.8469E-10
Fructose-bisphosphate aldolase A;Fructose-bisphosphate aldolase	ALDOA	5.439E-11
Fructose-bisphosphate aldolase A;Fructose-bisphosphate aldolase	ALDOA	0.0014753
Fructose-bisphosphate aldolase B;Fructose-bisphosphate aldolase	ALDOB	1.3704E-10
Fructose-bisphosphate aldolase C;Fructose-bisphosphate aldolase	ALDOC	2.1837E-40
Fumarylacetoacetase	FAH;DKFZp686F13224	4.2264E-156
Furin	FURIN	1.4451E-07
G antigen 3	GAGE3	0.000001021
Galectin-1	LGALS1	0.00004905
Galectin-3	LGALS3	7.718E-07
Galectin-3-binding protein	LGALS3BP	2.2418E-12
Galectin-9B;Galectin-9C;Galectin-9	LGALS9B;LGALS9C;LGA	4.5569E-16

	LS9	
Gamma-glutamyl hydrolase	GGH	0.00028357
Gamma-glutamylaminecyclotransferase	A2LD1	1.0522E-10
Gamma-glutamylcyclotransferase	GGCT	8.388E-36
Gamma-glutamyltranspeptidase 1;Gamma-glutamyltranspeptidase 1 heavy chain;Gamma-glutamyltranspeptidase 1 light chain;Gamma-glutamyltranspeptidase 2;Gamma-glutamyltranspeptidase 2 heavy chain;Gamma-glutamyltranspeptidase 2 light chain;Putative gamma-glutamyltranspeptidase 3;Putative gamma-glutamyltranspeptidase 3 heavy chain;Putative gamma-glutamyltranspeptidase 3 light chain	GGT1;GGT2;GGT3P	6.0629E-12
Gamma-interferon-inducible lysosomal thiol reductase	IFI30	0.0010132
Gamma-synuclein	SNCG	3.8879E-188
Gamma-tubulin complex component 2	TUBGCP2	0.0014057
Ganglioside GM2 activator;Ganglioside GM2 activator isoform short	GM2A	0.00025262
Gastrotropin	FABP6	3.3714E-11
GDNF family receptor alpha-1	GFRA1	0.00044583
Gelsolin	GSN	3.0704E-91
Glia maturation factor beta	GMFB	0.000029512
Glia maturation factor gamma;Glia maturation factor beta	GMFG;GMFB	0.000000249
Glucose-6-phosphate 1-dehydrogenase	G6PD	3.6098E-15
Glucose-6-phosphate isomerase	GPI	0.00030386
Glucosidase 2 subunit beta	PRKCSH	0.0006004
Glutamine-dependent NAD(+) synthetase	NADSYN1	1.1006E-06
Glutaminyl-peptide cyclotransferase	QPCT	2.1311E-12
Glutamyl aminopeptidase	ENPEP	4.0268E-103
Glutaredoxin-1	GLRX	3.936E-10
Glutathione peroxidase 3	GPX3	3.5077E-35
Glutathione reductase, mitochondrial	GSR	4.0563E-06
Glutathione S-transferase Mu 2	GSTM2	0.000021762
Glutathione S-transferase omega-1	GSTO1	0.0012883
Glutathione S-transferase P	GSTP1	0
Glutathione synthetase	GSS	5.5571E-75

Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	0.00051292
Glyceraldehyde-3-phosphate dehydrogenase, testis-specific	GAPDHS	2.617E-14
Glycerol-3-phosphate dehydrogenase, mitochondrial	GPD2	2.0436E-81
Glycogen debranching enzyme;4-alpha-glucanotransferase;Amylo-alpha-1,6-glucosidase	AGL	0
Glycogen phosphorylase, liver form;Phosphorylase	PYGL	0.00057162
Glycogenin-1	GYG1	5.572E-26
Glycolipid transfer protein	GLTP	9.2444E-11
Glycylpeptide N-tetradecanoyltransferase 1;Glycylpeptide N-tetradecanoyltransferase;Glycylpeptide N-tetradecanoyltransferase 2	NMT1;NMT2	2.9167E-07
Glyoxalase domain-containing protein 4	GLOD4	3.0715E-09
Glypican-1;Secreted glypican-1	GPC1	0.000008593
Golgi membrane protein 1	GOLM1	2.4433E-198
Golgi-associated plant pathogenesis-related protein 1	GLIPR2	6.0192E-09
G-protein coupled receptor family C group 5 member B	GPRC5B	0.000045408
G-protein coupled receptor family C group 5 member C	GPRC5C	7.2703E-15
G-protein-signaling modulator 3	GPSM3	5.5239E-09
Grancalcin	GCA	1.5059E-06
Granulins;Acrogranin;Paragranulin;Granulin-1;Granulin-2;Granulin-3;Granulin-4;Granulin-5;Granulin-6;Granulin-7	GRN	1.5079E-29
Group XV phospholipase A2	PLA2G15	6.2132E-19
Growth arrest-specific protein 1	GAS1	2.5962E-18
Growth factor receptor-bound protein 2	GRB2	2.8358E-25
Growth/differentiation factor 15	GDF15	0.00015389
Growth/differentiation factor 7	GDF7	0.00099808
GTP cyclohydrolase 1 feedback regulatory protein	GCHFR	0.00010076
GTP-binding protein SAR1a	SAR1A	6.8838E-17
Guanine nucleotide-binding protein G(i) subunit alpha-2;Guanine nucleotide-binding protein G(i) subunit alpha-1	GNAI2;GNAI1	0
Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12	GNG12	4.2428E-24
Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-2;Guanine nucleotide-binding protein subunit gamma	GNG2	5.3605E-13

Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	GNB1	0
Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2;Guanine nucleotide-binding protein subunit beta-4	GNB2;GNB4	1.2457E-50
Guanine nucleotide-binding protein G(k) subunit alpha;Guanine nucleotide-binding protein G(s) subunit alpha isoforms XLas;Guanine nucleotide-binding protein G(s) subunit alpha isoforms short;Guanine nucleotide-binding protein G(olf) subunit alpha;Guanine nucleotide-binding protein G(i) subunit alpha-2;Guanine nucleotide-binding protein G(t) subunit alpha-3;Guanine nucleotide-binding protein G(o) subunit alpha;Guanine nucleotide-binding protein G(t) subunit alpha-2;Guanine nucleotide-binding protein G(i) subunit alpha-1;Guanine nucleotide-binding protein G(t) subunit alpha-1	GNAI3;GNAS;GNAL;GNAI2;GNAT3;GNAO1;GNAT2;GNAI1;GNAT1	3.1844E-09
Guanine nucleotide-binding protein G(s) subunit alpha isoforms XLas;Guanine nucleotide-binding protein G(s) subunit alpha isoforms short	GNAS	6.3923E-15
Guanylate cyclase activator 2B;Guanylate cyclase C-activating peptide 2;Uroguanylin	GUCA2B	1.4728E-114
Guanylin;HMW-guanylin;Guanylin	GUCA2A	4.8251E-06
Haptoglobin;Haptoglobin alpha chain;Haptoglobin beta chain;Haptoglobin-related protein	HP;HPR	2.3726E-40
Haptoglobin-related protein	HPR	0.0010563
HEAT repeat-containing protein 7B2	HEATR7B2	0.00082649
Heat shock 70 kDa protein 1A/1B	HSPA1A;HSPA1B	7.3899E-24
Heat shock 70 kDa protein 1-like	HSPA1L	1.2489E-12
Heat shock 70 kDa protein 6;Putative heat shock 70 kDa protein 7	HSPA6;HSPA7	9.1659E-08
Heat shock cognate 71 kDa protein	HSPA8	6.5944E-06
Heat shock factor-binding protein 1	HSBP1	1.5799E-10
Heat shock protein beta-1	HSPB1	9.5577E-73
Heat shock protein beta-11	HSPB11	2.4176E-10
Heat shock protein HSP 90-alpha	HSP90AA1	4.4141E-07
Heat shock protein HSP 90-beta	HSP90AB1	3.0286E-14
Hematological and neurological expressed 1 protein	HN1	0.0016425
Hematological and neurological expressed 1 protein	HN1	4.8946E-74
Hematopoietic lineage cell-specific protein	HCLS1	1.064E-14
Heme-binding protein 2	HEBP2;RP3-422G23.1	2.8932E-17
Hemicentin-1	HMCN1	4.1496E-30

Hemicentin-2	HMCN2	1.3599E-16
Hemoglobin subunit alpha	HBA1	1.1863E-20
Hemoglobin subunit beta;LVV-hemorphin-7	HBB	0.00067848
Hemoglobin subunit delta	HBD	4.672E-21
Hemoglobin subunit gamma-1;Hemoglobin subunit gamma-2	HBG1;HBG2	0.000016308
Hemopexin	HPX	0.00024676
Heparin cofactor 2	SERPIND1	4.1746E-137
Hepatitis A virus cellular receptor 2	HAVCR2	1.3141E-19
Hepatocyte growth factor activator;Hepatocyte growth factor activator short chain;Hepatocyte growth factor activator long chain	HGFAC	0.00028698
Hepatocyte growth factor-like protein;Hepatocyte growth factor-like protein alpha chain;Hepatocyte growth factor-like protein beta chain	MST1	3.3914E-82
Hepatoma-derived growth factor	HDGF	0.00064193
Hepatoma-derived growth factor-related protein 2	HDGFRP2	0.000087172
Hepcidin;Hepcidin-25;Hepcidin-20	HAMP	4.3296E-07
Heterogeneous nuclear ribonucleoprotein A/B	HNRNPAB	0.0019671
Heterogeneous nuclear ribonucleoprotein A1;Heterogeneous nuclear ribonucleoprotein A1-like 2	HNRNPA1;HNRNPA1L2	0.000028058
Heterogeneous nuclear ribonucleoprotein A3	HNRNPA3	7.1878E-09
Heterogeneous nuclear ribonucleoprotein D0	HNRNPD	6.893E-19
Heterogeneous nuclear ribonucleoprotein F;Heterogeneous nuclear ribonucleoprotein F, N-terminally processed	HNRNPF	5.2577E-07
Heterogeneous nuclear ribonucleoprotein H;Heterogeneous nuclear ribonucleoprotein H, N-terminally processed;Heterogeneous nuclear ribonucleoprotein F;Heterogeneous nuclear ribonucleoprotein F, N-terminally processed;Heterogeneous nuclear ribonucleoprotein H2	HNRNPH1;HNRNPF;HNRNPH2	3.7527E-30
Heterogeneous nuclear ribonucleoprotein H2;Heterogeneous nuclear ribonucleoprotein H;Heterogeneous nuclear ribonucleoprotein H, N-terminally processed	HNRNPH2;HNRNPH1	0.00071967
Heterogeneous nuclear ribonucleoprotein H3	HNRNPH3	1.468E-20
Heterogeneous nuclear ribonucleoprotein K	HNRNPK	0.000047075
Heterogeneous nuclear ribonucleoprotein K	HNRNPK	1.575E-50
Heterogeneous nuclear ribonucleoprotein L	HNRNPL	3.2482E-07

Heterogeneous nuclear ribonucleoprotein M	HNRNPM	1.3665E-09
Heterogeneous nuclear ribonucleoprotein Q	SYNCRIP	2.935E-88
Heterogeneous nuclear ribonucleoprotein R	HNRNPR;HNRPR	1.6506E-11
Heterogeneous nuclear ribonucleoprotein U	HNRNPU	1.6148E-92
Heterogeneous nuclear ribonucleoprotein U-like protein 2	HNRNPUL2;hCG_2044799	1.2201E-15
Heterogeneous nuclear ribonucleoproteins A2/B1	HNRNPA2B1	0.00065236
Heterogeneous nuclear ribonucleoproteins C1/C2;Heterogeneous nuclear ribonucleoprotein C-like 1	HNRNPC;HNRNPCL1	9.3449E-20
Hexokinase-3	HK3	6.9279E-10
High mobility group nucleosome-binding domain-containing protein 3;Non-histone chromosomal protein HMG-17	HMGN3;HMGN2	0.00048981
High mobility group nucleosome-binding domain-containing protein 4	HMGN4	2.2322E-07
High mobility group protein B1;Putative high mobility group protein B1-like 1	HMGB1;HMGB1P1	1.1399E-11
High mobility group protein B2	HMGB2	2.225E-08
High mobility group protein B3;Putative high mobility group protein B3-like protein	HMGB3	5.1539E-40
Hippocalcin-like protein 1;Neurocalcin-delta;Neuron-specific calcium-binding protein hippocalcin	HPCAL1;NCALD;HPCA	6.5985E-41
Histidine-rich glycoprotein	HRG	1.501E-21
Histone H1.0	H1F0	2.0161E-154
Histone H1.2	HIST1H1C	5.6616E-149
Histone H1.3	HIST1H1D	2.8596E-12
Histone H1.4	HIST1H1E	7.4212E-14
Histone H1.5	HIST1H1B	1.2881E-52
Histone H1x	H1FX	0.0018252
Histone H2A.x;Histone H2A type 1-A;Histone H2A type 1-B/E;Histone H2A type 1;Histone H2A type 1-D;Histone H2A type 2-A;Histone H2A type 3;Histone H2A type 1-C;Histone H2A type 2-C;Histone H2A.J;Histone H2A type 1-H;Histone H2A type 1-J;Histone H2A type 2-B;Histone H2A.Z;Histone H2A.V;Histone H2A	H2AFX;HIST1H2AA;HIST1H2AB;HIST1H2AG;HIST1H2AD;HIST2H2AA3;HIST3H2A;HIST1H2AC;HIST2H2AC;H2AFJ;HIST1H2AH;HIST1H2AJ;HIST2H2AB;H2AFZ;H2AF	2.7522E-19

	V	
Histone H2B;Histone H2B type 1-K;Histone H2B type F-S;Histone H2B type 1-D;Histone H2B type 1-C/E/F/G/I;Histone H2B type 2-F;Histone H2B type 1-H;Histone H2B type 1-N;Histone H2B type 1-M;Histone H2B type 1-L;Histone H2B type 1-J;Histone H2B type 1-O;Histone H2B type 1-B;Histone H2B type 2-E;Histone H2B type 1-A;Histone H2B type 3-B	HIST2H2BF;HIST1H2BK;H2BFS;HIST1H2BD;HIST1H2BC;HIST1H2BH;HIST1H2BN;HIST1H2BM;HIST1H2BL;HIST1H2BJ;HIST1H2BO;HIST1H2BB;HIST2H2BE;HIST1H2BA;HIST3H2BB	0.00022836
Histone H3.1;Histone H3.3;Histone H3.1t;Histone H3.2;Histone H3	HIST1H3A;H3F3A;HIST3H3;HIST2H3A	0.000030865
Histone H3.1;Histone H3.3;Histone H3.1t;Histone H3.2;Histone H3;Histone H3.3C	HIST1H3A;H3F3A;HIST3H3;HIST2H3A;H3F3C	5.0837E-15
Histone H3;Histone H3.1;Histone H3.3;Histone H3.1t;Histone H3.2;Histone H3.3C	HIST2H3PS2;HIST1H3A;H3F3A;HIST3H3;HIST2H3A;H3F3C	4.8978E-09
Histone H4	HIST1H4A	0.000024515
HLA class I histocompatibility antigen, A-24 alpha chain;HLA class I histocompatibility antigen, A-23 alpha chain	HLA-A	0.000055965
HLA class I histocompatibility antigen, A-3 alpha chain;HLA class I histocompatibility antigen, A-11 alpha chain;HLA class I histocompatibility antigen, A-30 alpha chain;HLA class I histocompatibility antigen, A-68 alpha chain;HLA class I histocompatibility antigen, A-2 alpha chain;HLA class I histocompatibility antigen, A-69 alpha chain;HLA class I histocompatibility antigen, A-31 alpha chain;HLA class I histocompatibility antigen, A-33 alpha chain;HLA class I histocompatibility antigen, A-34 alpha chain;HLA class I histocompatibility antigen, A-66 alpha chain;HLA class I histocompatibility antigen, A-74 alpha chain	HLA-A	1.6511E-10
HLA class I histocompatibility antigen, A-68 alpha chain;HLA class I histocompatibility antigen, A-69 alpha chain;HLA class I histocompatibility antigen, A-33 alpha chain;HLA class I histocompatibility antigen, A-34 alpha chain;HLA class I histocompatibility antigen, A-66 alpha chain;HLA class I histocompatibility antigen, A-25 alpha chain	HLA-A;HLA-B	2.2288E-10
HLA class I histocompatibility antigen, Cw-12 alpha chain;HLA class I histocompatibility antigen, Cw-14 alpha chain;HLA class I histocompatibility antigen, Cw-16 alpha chain;HLA class I histocompatibility antigen, B-54 alpha chain;HLA class I histocompatibility antigen,	HLA-C;HLA-B;HLA-A	3.5922E-166

Cw-17 alpha chain;HLA class I histocompatibility antigen, Cw-4 alpha chain;HLA class I histocompatibility antigen, A-24 alpha chain;HLA class I histocompatibility antigen, A-1 alpha chain;HLA class I histocompatibility antigen, A-23 alpha chain;HLA class I histocompatibility antigen, A-36 alpha chain;HLA class I histocompatibility antigen, B-55 alpha chain;HLA class I histocompatibility antigen, B-59 alpha chain		
HLA class I histocompatibility antigen, Cw-12 alpha chain;HLA class I histocompatibility antigen, Cw-14 alpha chain;HLA class I histocompatibility antigen, Cw-16 alpha chain;HLA class I histocompatibility antigen, Cw-3 alpha chain;HLA class I histocompatibility antigen, Cw-15 alpha chain;HLA class I histocompatibility antigen, B-54 alpha chain;HLA class I histocompatibility antigen, Cw-17 alpha chain;HLA class I histocompatibility antigen, Cw-8 alpha chain;HLA class I histocompatibility antigen, Cw-18 alpha chain;HLA class I histocompatibility antigen, Cw-6 alpha chain;HLA class I histocompatibility antigen, Cw-5 alpha chain;HLA class I histocompatibility antigen, Cw-2 alpha chain;HLA class I histocompatibility antigen, Cw-4 alpha chain	HLA-C;HLA-B	1.4974E-65
HLA class I histocompatibility antigen, Cw-17 alpha chain;HLA class I histocompatibility antigen, Cw-3 alpha chain;HLA class I histocompatibility antigen, Cw-7 alpha chain;HLA class I histocompatibility antigen, Cw-1 alpha chain;HLA class I histocompatibility antigen, Cw-8 alpha chain;HLA class I histocompatibility antigen, Cw-12 alpha chain;HLA class I histocompatibility antigen, Cw-14 alpha chain;HLA class I histocompatibility antigen, Cw-15 alpha chain;HLA class I histocompatibility antigen, Cw-18 alpha chain;HLA class I histocompatibility antigen, Cw-16 alpha chain;HLA class I histocompatibility antigen, Cw-6 alpha chain;HLA class I histocompatibility antigen, Cw-5 alpha chain;HLA class I histocompatibility antigen, B-54 alpha chain	HLA-C;HLA-B	0.0016822
HLA class II histocompatibility antigen gamma chain	CD74	1.0591E-16
Hsp90 co-chaperone Cdc37	CDC37	0.0010088
Hyaluronan-binding protein 2;Hyaluronan-binding protein 2 50 kDa heavy chain;Hyaluronan-binding protein 2 50 kDa heavy chain alternate form;Hyaluronan-binding protein 2 27 kDa light chain;Hyaluronan-binding protein 2 27 kDa light chain alternate form	HABP2	1.0413E-07
Hyaluronidase-1	HYAL1	1.9334E-07
Hypoxanthine-guanine phosphoribosyltransferase	HPRT1	0.0001875
Hypoxia up-regulated protein 1	HYOU1	0.000023816
ICOS ligand	ICOSLG	0.00058409

Ig alpha-1 chain C region	IGHA1	2.0499E-08
Ig alpha-2 chain C region	IGHA2	5.014E-15
Ig delta chain C region	IGHD	9.8223E-06
Ig gamma-1 chain C region	IGHG1	2.1812E-102
Ig gamma-2 chain C region	IGHG2	7.2922E-07
Ig gamma-3 chain C region	IGHG3	9.0958E-22
Ig gamma-4 chain C region	IGHG4	1.1542E-15
Ig heavy chain V-I region HG3		0.000070879
Ig heavy chain V-I region Mot;Ig heavy chain V-I region SIE;Ig heavy chain V-I region EU		9.0902E-69
Ig heavy chain V-I region V35		3.2835E-09
Ig heavy chain V-II region ARH-77		0.000025099
Ig heavy chain V-II region NEWM		0.00064647
Ig heavy chain V-II region SESS;Ig heavy chain V-II region OU;Ig heavy chain V-II region COR		5.322E-66
Ig heavy chain V-II region WAH		0.0004774
Ig heavy chain V-III region BRO;Ig heavy chain V-III region TEI;Ig heavy chain V-III region WEA		2.0867E-06
Ig heavy chain V-III region BUR		0
Ig heavy chain V-III region BUT		0.00093521
Ig heavy chain V-III region GAL		2.0926E-20
Ig heavy chain V-III region GAR		0
Ig heavy chain V-III region HIL		2.8216E-15
Ig heavy chain V-III region JON		8.7045E-33
Ig heavy chain V-III region KOL		5.2269E-10
Ig heavy chain V-III region LAY		0.00014788
Ig heavy chain V-III region NIE		2.2422E-36
Ig heavy chain V-III region TIL;Ig heavy chain V-III region POM;Ig heavy chain V-III region WAS		1.363E-26
Ig heavy chain V-III region TUR;Ig heavy chain V-III region NIE		0.00017231
Ig heavy chain V-III region VH26		0.0016753
Ig heavy chain V-III region WEA		8.2377E-126

Ig heavy chain V-III region ZAP		2.4681E-82
Ig kappa chain C region	IGKC	7.0704E-06
Ig kappa chain V-I region AU		8.2665E-14
Ig kappa chain V-I region BAN		7.9784E-11
Ig kappa chain V-I region DEE		3.816E-162
Ig kappa chain V-I region EU;Ig kappa chain V-I region CAR		7.3836E-07
Ig kappa chain V-I region HK101		5.8752E-13
Ig kappa chain V-I region HK102	IGKV1-5	4.0031E-38
Ig kappa chain V-I region Ka		0.000017446
Ig kappa chain V-I region Kue		0.000013007
Ig kappa chain V-I region Lay		2.9493E-11
Ig kappa chain V-I region Mev		2.6297E-17
Ig kappa chain V-I region Ni		9.6444E-07
Ig kappa chain V-I region Rei		0.00029562
Ig kappa chain V-I region Roy;Ig kappa chain V-I region AG;Ig kappa chain V-I region Gal;Ig kappa chain V-I region Hau;Ig kappa chain V-I region Rei;Ig kappa chain V-I region WAT		2.4083E-09
Ig kappa chain V-I region Scw		0.00040137
Ig kappa chain V-I region Walker		2.1382E-21
Ig kappa chain V-I region WEA		2.2633E-07
Ig kappa chain V-I region Wes		0.00017646
Ig kappa chain V-II region FR		0.0013869
Ig kappa chain V-II region GM607;Ig kappa chain V-II region TEW;Ig kappa chain V-II region RPMI 6410;Ig kappa chain V-II region Cum		5.9229E-36
Ig kappa chain V-II region MIL		4.4378E-12
Ig kappa chain V-II region TEW		0.000019597
Ig kappa chain V-III region B6		1.0327E-69
Ig kappa chain V-III region CLL		6.4894E-19
Ig kappa chain V-III region GOL		0.00011351
Ig kappa chain V-III region HAH		5.2514E-34
Ig kappa chain V-III region HAH;Ig kappa chain V-III region HIC		4.2044E-21
Ig kappa chain V-III region IARC/BL41		3.5447E-53

Ig kappa chain V-III region NG9		0.000059585
Ig kappa chain V-III region POM		0.0013726
Ig kappa chain V-III region SIE;Ig kappa chain V-III region WOL		0.000037909
Ig kappa chain V-III region VG		0.000016591
Ig kappa chain V-IV region B17		0.0016206
Ig kappa chain V-IV region JI;Ig kappa chain V-IV region;Ig kappa chain V-IV region B17;Ig kappa chain V-IV region STH	IGKV4-1	1.5474E-23
Ig kappa chain V-IV region Len		0.00044717
Ig lambda chain V region 4A		0.0013422
Ig lambda chain V-I region BL2;Ig lambda chain V-I region NIG-64;Ig lambda chain V-I region NEW;Ig lambda chain V-I region EPS		7.2305E-19
Ig lambda chain V-I region HA		2.3288E-54
Ig lambda chain V-I region NEW		1.2044E-08
Ig lambda chain V-I region NEWM		1.0556E-06
Ig lambda chain V-I region VOR		2.1562E-18
Ig lambda chain V-I region WAH		2.5316E-35
Ig lambda chain V-II region BUR;Ig lambda chain V-II region BOH;Ig lambda chain V-II region NIG-58		0.000029902
Ig lambda chain V-II region NIG-84		1.5197E-06
Ig lambda chain V-III region LOI		0.00026577
Ig lambda chain V-III region SH		3.5857E-22
Ig lambda chain V-IV region Bau		3.2879E-08
Ig lambda chain V-IV region Hil		2.6819E-06
Ig lambda chain V-IV region MOL		0.0002695
Ig lambda chain V-V region DEL;Ig lambda chain V-VII region MOT		0.00016122
Ig lambda chain V-VI region EB4;Ig lambda chain V-VI region WLT		2.3876E-07
Ig lambda-2 chain C regions;Ig lambda-3 chain C regions	IGLC2;IGLC3	8.4073E-07
Ig lambda-2 chain C regions;Ig lambda-3 chain C regions;Ig lambda-6 chain C region	IGLC2;IGLC3;IGLC6	1.7763E-20
Ig lambda-6 chain C region	IGLC6	0
Ig lambda-7 chain C region	IGLC7	0.0012219
Ig mu chain C region;Ig mu heavy chain disease protein	IGHM	0.0018035

Ig mu heavy chain disease protein		1.8838E-72
IGF-like family receptor 1	IGFLR1	0.000050049
IgGFc-binding protein	FCGBP	0.000033592
Immunoglobulin J chain	IGJ	0.0001247
Immunoglobulin lambda-like polypeptide 1	IGLL1	8.5806E-06
Immunoglobulin lambda-like polypeptide 5;Ig lambda-1 chain C regions	IGLL5;IGLC1	0.0042421
Immunoglobulin superfamily containing leucine-rich repeat protein	ISLR	5.5798E-67
Immunoglobulin superfamily member 21	IGSF21	0.00047254
Immunoglobulin superfamily member 8	IGSF8	8.4164E-06
Importin subunit beta-1	KPNB1	4.0994E-14
Inorganic pyrophosphatase 2, mitochondrial	PPA2	1.3661E-28
Inositol monophosphatase 1	IMPA1	1.5031E-08
Insulin-like growth factor II;Insulin-like growth factor II;Insulin-like growth factor II Ala-25 Del;Preptin	IGF2	1.7038E-38
Insulin-like growth factor II;Insulin-like growth factor II;Insulin-like growth factor II Ala-25 Del;Preptin	IGF2	5.4873E-10
Insulin-like growth factor-binding protein 2	IGFBP2	0
Insulin-like growth factor-binding protein 3	IGFBP3	2.0429E-19
Insulin-like growth factor-binding protein 3	IGFBP3	4.6778E-44
Insulin-like growth factor-binding protein 4	IGFBP4	8.4992E-28
Insulin-like growth factor-binding protein 5	IGFBP5	5.066E-37
Insulin-like growth factor-binding protein 6	IGFBP6	8.0924E-07
Insulin-like growth factor-binding protein 7	IGFBP7	8.8268E-08
Insulin-like growth factor-binding protein complex acid labile subunit	IGFALS	1.4443E-41
Integral membrane protein DGCR2/IDD	DGCR2;DKFZp686I1730	1.7388E-244
Integral membrane protein GPR180	GPR180	2.7124E-06
Integrin alpha-7;Integrin alpha-7 heavy chain;Integrin alpha-7 light chain;Integrin alpha-7 70 kDa form	ITGA7	6.9948E-123
Integrin alpha-M	ITGAM	2.0354E-117
Integrin alpha-M	ITGAM	0.00036153

Integrin beta-1	ITGB1	0.000045717
Integrin beta-2;Integrin beta	ITGB2	3.3228E-45
Intelectin-1	ITLN1	2.5427E-07
Inter-alpha-trypsin inhibitor heavy chain H1	ITIH1	0.0036674
Inter-alpha-trypsin inhibitor heavy chain H2	ITIH2	5.1376E-96
Inter-alpha-trypsin inhibitor heavy chain H3	ITIH3	8.9842E-95
Inter-alpha-trypsin inhibitor heavy chain H4;70 kDa inter-alpha-trypsin inhibitor heavy chain H4;35 kDa inter-alpha-trypsin inhibitor heavy chain H4	ITIH4	1.0341E-06
Intercellular adhesion molecule 1	ICAM1	3.9471E-20
Intercellular adhesion molecule 2	ICAM2	1.1158E-15
Interferon alpha/beta receptor 2	IFNAR2	0.000013405
Interferon-induced transmembrane protein 3;Interferon-induced transmembrane protein 2;Interferon-induced transmembrane protein 1	IFITM3;IFITM2;IFITM1	0.00076179
Interleukin enhancer-binding factor 2	ILF2	2.2793E-11
Interleukin enhancer-binding factor 3;Spermatid perinuclear RNA-binding protein	ILF3;STRBP	4.6081E-09
Interleukin-1 receptor accessory protein	IL1RAP	1.8155E-34
Interleukin-1 receptor antagonist protein	IL1RN	1.7075E-46
Interleukin-10 receptor subunit beta	IL10RB	2.4532E-28
Interleukin-18-binding protein	IL18BP	1.3945E-17
Interleukin-18-binding protein	IL18BP	1.1361E-20
Interleukin-6 receptor subunit beta	IL6ST	1.1541E-11
Involucrin	IVL	2.8245E-21
Isoamyl acetate-hydrolyzing esterase 1 homolog	IAH1	0
IST1 homolog	IST1	1.2732E-23
IST1 homolog	IST1	6.9131E-50
Janus kinase and microtubule-interacting protein 1	JAKMIP1	0.00059792
Junctional adhesion molecule A	F11R	9.0608E-08
Junctional adhesion molecule A	F11R	0.00079068
Junctional adhesion molecule C	JAM3	1.1822E-130
Kallikrein-1	KLK1	1.0879E-45
Kallikrein-11;Kallikrein-11 inactive chain 1;Kallikrein-11 inactive chain 2	KLK11	0.00018172

Kallistatin	SERPINA4	6.5259E-78
Kelch repeat and BTB domain-containing protein 11	KBTBD11	0.000044848
Keratin, type I cytoskeletal 10	KRT10	9.7592E-17
Keratin, type I cytoskeletal 13	KRT13	6.1935E-06
Keratin, type I cytoskeletal 16	KRT16	0.0012819
Keratin, type I cytoskeletal 17	KRT17;JUP	3.6553E-29
Keratin, type I cytoskeletal 19	KRT19	9.3773E-44
Keratin, type I cytoskeletal 9	KRT9	3.2163E-41
Keratin, type II cytoskeletal 1	KRT1	0.00043531
Keratin, type II cytoskeletal 2 epidermal	KRT2	2.7257E-44
Keratin, type II cytoskeletal 4	KRT4	1.3862E-07
Keratin, type II cytoskeletal 5	KRT5	2.2276E-37
Keratin, type II cytoskeletal 6B	KRT6B	0.00052328
Keratin, type II cytoskeletal 7	KRT7	0.000039252
Keratin, type II cytoskeletal 8	KRT8;KRT8P11	9.7236E-59
Keratin, type II cytoskeletal 8	KRT8	3.8165E-13
Keratinocyte-associated transmembrane protein 2	KCT2	0.00012941
KH domain-containing, RNA-binding, signal transduction-associated protein 1	KHDRBS1	1.9284E-57
Kin of IRRE-like protein 1	KIRREL	1.2768E-151
Kinesin-1 heavy chain	KIF5B	1.2392E-07
Kininogen-1;Kininogen-1 heavy chain;T-kinin;Bradykinin;Lysyl-bradykinin;Kininogen-1 light chain;Low molecular weight growth-promoting factor	KNG1	0.000066384
Kunitz-type protease inhibitor 1	SPINT1	3.5589E-07
Kunitz-type protease inhibitor 2	SPINT2	0.0018601
Lactotransferrin;Kaliocin-1;Lactoferroxin-A;Lactoferroxin-B;Lactoferroxin-C	LTF	1.609E-13
Lactoylglutathione lyase	GLO1	2.3021E-295
Lamina-associated polypeptide 2, isoform alpha;Thymopoietin;Thymopentin;Lamina-associated polypeptide 2, isoforms beta/gamma;Thymopoietin;Thymopentin	TMPO	1.0378E-08
Lamin-B receptor	LBR	0.000029296
Lamin-B1	LMNB1	2.6961E-41
Lamin-B2	LMNB2	4.2977E-21

Laminin subunit alpha-2	LAMA2	3.1617E-21
Laminin subunit alpha-3	LAMA3	8.1092E-08
Laminin subunit alpha-4	LAMA4	2.249E-47
Laminin subunit alpha-5	LAMA5	3.7241E-06
Laminin subunit beta-1	LAMB1	1.6213E-20
Laminin subunit beta-2	LAMB2	5.9904E-16
Laminin subunit gamma-1	LAMC1	7.3139E-72
L-asparaginase	ASRGL1	0.000049881
Latent-transforming growth factor beta-binding protein 1	LTBP1	2.8402E-06
Latent-transforming growth factor beta-binding protein 2	LTBP2	4.6464E-20
Latent-transforming growth factor beta-binding protein 4	LTBP4	6.1252E-240
Latent-transforming growth factor beta-binding protein 4	LTBP4	0
Latrophilin-1	LPHN1	3.0001E-25
Leucine-rich alpha-2-glycoprotein	LRG1	3.3605E-34
Leucine-rich repeat flightless-interacting protein 1	LRRFIP1	7.5756E-63
Leucine-rich repeat flightless-interacting protein 2	LRRFIP2	9.473E-09
Leucine-rich repeat neuronal protein 1	LRRN1	7.7832E-38
Leucine-rich repeat-containing protein 19	LRRC19	8.7823E-28
Leukocyte elastase inhibitor	SERPINB1	0.000055609
Leukocyte immunoglobulin-like receptor subfamily A member 5	LILRA5	1.1337E-08
Leukocyte immunoglobulin-like receptor subfamily B member 4	LILRB4	1.8318E-239
Leukocyte-associated immunoglobulin-like receptor 1	LAIR1	2.6787E-06
Leukocyte-associated immunoglobulin-like receptor 2	LAIR2	3.8841E-216
Leukotriene A-4 hydrolase	LTA4H	2.2E-16
LIM and SH3 domain protein 1	LASP1	0.00094027
Lipocalin-1;Putative lipocalin 1-like protein 1	LCN1;LCN1P1	0.00040873
Lipopolysaccharide-binding protein	LBP	1.8381E-81
Lithostathine-1-alpha	REG1A	5.6623E-14
Lithostathine-1-beta	REG1B	3.9182E-07
Liver carboxylesterase 1	CES1	5.3098E-10

Liver-expressed antimicrobial peptide 2	LEAP2	1.7265E-19
L-lactate dehydrogenase A chain	LDHA	1.8467E-09
L-lactate dehydrogenase B chain;L-lactate dehydrogenase	LDHB	2.0678E-13
Low affinity immunoglobulin gamma Fc region receptor III-A	FCGR3A	0
Low affinity immunoglobulin gamma Fc region receptor III-B	FCGR3B	0.0016861
Low molecular weight phosphotyrosine protein phosphatase	ACP1	4.5444E-28
Low-density lipoprotein receptor	LDLR	0
Low-density lipoprotein receptor-related protein 2	LRP2	0.000074232
L-selectin	SELL	0
Lumican	LUM	2.5261E-07
Ly-6/neurotoxin-like protein 1	LYNX1	0
Ly6/PLAUR domain-containing protein 2	LYPD2	1.2915E-07
Ly6/PLAUR domain-containing protein 6B	LYPD6B	2.0257E-77
Lymphatic vessel endothelial hyaluronic acid receptor 1	LYVE1	4.4199E-24
Lymphocyte antigen 6D	LY6D	1.0586E-74
Lymphocyte antigen 6H	LY6H	3.2622E-28
Lymphocyte antigen 6H	LY6H	5.2199E-11
Lymphocyte cytosolic protein 2	LCP2	9.5068E-24
Lymphocyte-specific protein 1	LSP1	0
Lysophosphatidylcholine acyltransferase 2	LPCAT2	7.8387E-40
Lysosomal acid phosphatase	ACP2	0.000020284
Lysosomal alpha-glucosidase;76 kDa lysosomal alpha-glucosidase;70 kDa lysosomal alpha-glucosidase	GAA	3.9125E-19
Lysosomal alpha-mannosidase;Lysosomal alpha-mannosidase A peptide;Lysosomal alpha-mannosidase B peptide;Lysosomal alpha-mannosidase C peptide;Lysosomal alpha-mannosidase D peptide;Lysosomal alpha-mannosidase E peptide	MAN2B1	3.2704E-07
Lysosomal protective protein;Lysosomal protective protein 32 kDa chain;Lysosomal protective protein 20 kDa chain	CTSA	2.234E-36
Lysosomal Pro-X carboxypeptidase	PRCP	5.1358E-07
Lysosome-associated membrane glycoprotein 1	LAMP1	2.1799E-130
Lysosome-associated membrane glycoprotein 2	LAMP2	0.000027932

Lysozyme C	LYZ	1.1931E-10
Lysyl oxidase homolog 1	LOXL1	6.3107E-131
m7GpppX diphosphatase	DCPS	4.4352E-14
Macrophage colony-stimulating factor 1 receptor	CSF1R	6.7181E-06
Macrophage colony-stimulating factor 1;Processed macrophage colony-stimulating factor 1	CSF1	0
Macrophage mannose receptor 1	MRC1	0.00042295
Macrophage migration inhibitory factor	MIF	1.273E-36
Macrophage-capping protein	CAPG	1.9844E-23
Magnesium-dependent phosphatase 1;NEDD8	NEDD8-MDP1;MDP1;NEDD8	2.8084E-06
Major prion protein	PRNP	4.332E-116
Major vault protein	MVP	7.2354E-15
Malate dehydrogenase, mitochondrial;Malate dehydrogenase	MDH2	1.718E-139
Malate dehydrogenase;Malate dehydrogenase, cytoplasmic	MDH1	2.6163E-42
Malcavernin	CCM2	2.3529E-07
Maltase-glucoamylase, intestinal;Maltase;Glucoamylase	MGAM	6.2653E-18
Mannan-binding lectin serine protease 1;Mannan-binding lectin serine protease 1 heavy chain;Mannan-binding lectin serine protease 1 light chain	MASP1	1.4057E-75
Mannan-binding lectin serine protease 2;Mannan-binding lectin serine protease 2 A chain;Mannan-binding lectin serine protease 2 B chain	MASP2	1.1699E-09
Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA	MAN1A1	1.4521E-17
Mannosyl-oligosaccharide 1,2-alpha-mannosidase IC	MAN1C1	7.2694E-131
MANSC domain-containing protein 1	MANSC1	2.0127E-73
MAPK-interacting and spindle-stabilizing protein-like	MAPK1IP1L	9.1255E-20
MARCKS-related protein	MARCKSL1	4.4252E-40
Matrilin-4	MATN4	4.7878E-06
Matrilysin	MMP7	8.7057E-08
Matrin-3	MATR3	2.5953E-16
Matrix metalloproteinase-9;67 kDa matrix metalloproteinase-9;82 kDa matrix metalloproteinase-9	MMP9	3.2127E-32

Matrix-remodeling-associated protein 8	MXRA8	0.0017156
Medium-chain specific acyl-CoA dehydrogenase, mitochondrial	ACADM	2.0787E-09
Melanoma inhibitory activity protein 3	MIA3	2.8757E-10
Melanotransferrin	MFI2	2.77E-26
Membrane protein FAM174A	FAM174A	1.3258E-53
Meprin A subunit alpha	MEP1A	9.45E-29
Mesencephalic astrocyte-derived neurotrophic factor	MANF	3.8264E-06
Mesothelin;Megakaryocyte-potentiating factor;Mesothelin, cleaved form	MSLN	5.3203E-163
Metalloproteinase inhibitor 1	TIMP1	7.6322E-08
Metalloproteinase inhibitor 2	TIMP2	0.00061075
Metallothionein-1G;Metallothionein-2;Metallothionein-1E;Metallothionein-1X;Metallothionein-1M	MT1G;MT2A;MT1E;MT1X;MT1M	0.00020917
Metallothionein-1H	MT1H	9.637E-27
Microtubule-associated protein RP/EB family member 1	MAPRE1	3.7425E-52
Microtubule-associated protein RP/EB family member 2	MAPRE2	0
Microtubule-associated proteins 1A/1B light chain 3 beta 2;Microtubule-associated proteins 1A/1B light chain 3B	MAP1LC3B2;MAP1LC3B	0
Migration and invasion enhancer 1	MIEN1	8.1084E-22
Mimecan	OGN	0.0016521
Minor histocompatibility protein HA-1;Minor histocompatibility antigen HA-1	HMHA1	1.1569E-32
Mitochondrial fission 1 protein	FIS1	0
Mitochondrial folate transporter/carrier	SLC25A32	8.6642E-36
Mitochondrial import receptor subunit TOM22 homolog	TOMM22	2.7928E-62
Mitochondrial import receptor subunit TOM70	TOMM70A	0.00028792
Moesin	MSN	0
Molybdopterin synthase sulfur carrier subunit	MOCS2	4.7497E-14
Monocyte differentiation antigen CD14;Monocyte differentiation antigen CD14, urinary form;Monocyte differentiation antigen CD14, membrane-bound form	CD14	3.9928E-16
Motile sperm domain-containing protein 2	MOSPD2	0
Mucin-1;Mucin-1 subunit alpha;Mucin-1 subunit beta	MUC1	2.4876E-282
Mucin-4;Mucin-4 alpha chain;Mucin-4 beta chain	MUC4	4.6126E-23

Mucin-5AC	MUC5AC	8.9306E-22
Mucin-5B	MUC5B	0.000011416
Mucin-6	MUC6	0
Mucin-like protein 1	MUCL1	0.00018055
Mucosal addressin cell adhesion molecule 1	MADCAM1	0.00046627
Multifunctional protein ADE2;Phosphoribosylaminoimidazole-succinocarboxamide synthase;Phosphoribosylaminoimidazole carboxylase	PAICS	2.5485E-56
Multimerin-1;Platelet glycoprotein Ia*;155 kDa platelet multimerin	MMRN1	4.972E-86
Multimerin-2	MMRN2	1.9363E-179
Myelin protein zero-like protein 1	MPZL1	0.0014737
Myeloblastin	PRTN3	0.00028755
Myeloid cell nuclear differentiation antigen	MNDA	1.472E-33
Myeloid cell surface antigen CD33	CD33	5.0518E-17
Myeloperoxidase;Myeloperoxidase;89 kDa myeloperoxidase;84 kDa myeloperoxidase;Myeloperoxidase light chain;Myeloperoxidase heavy chain	MPO	3.1016E-64
Myocilin	MYOC	3.2169E-26
Myoglobin	MB	5.4596E-07
Myomegalin;CDK5 regulatory subunit-associated protein 2	PDE4DIP;CDK5RAP2	3.0504E-13
Myosin light chain 3;Myosin light chain 1/3, skeletal muscle isoform	MYL3;MYL1	1.5963E-09
Myosin light chain kinase, smooth muscle;Myosin light chain kinase, smooth muscle, deglutamylated form	MYLK	6.4139E-34
Myosin light polypeptide 6	MYL6	2.6103E-07
Myosin regulatory light chain 12B;Myosin regulatory light chain 12A;Myosin regulatory light polypeptide 9	MYL12B;MYL12A;MYL9	5.3336E-12
Myosin-2	MYH2	8.4818E-08
Myosin-9	MYH9	7.3061E-160
Myotrophin	MTPN	2.7987E-06
Myotubularin-related protein 13	SBF2	7.428E-118
Myristoylated alanine-rich C-kinase substrate	MARCKS	5.2782E-15
N(4)-(beta-N-acetylglucosaminyl)-L-asparaginase;Glycosylasparaginase alpha chain;Glycosylasparaginase beta chain	AGA	1.0605E-131

N(G),N(G)-dimethylarginine dimethylaminohydrolase 2	DDAH2	0.00028119
Na(+)/H(+) exchange regulatory cofactor NHE-RF1	SLC9A3R1	2.7446E-14
Na(+)/H(+) exchange regulatory cofactor NHE-RF3;Putative PDZ domain-containing protein 1P	PDZK1;PDZK1P1	0.00077906
N-acetyl-D-glucosamine kinase	NAGK	7.8174E-11
N-acetyl-D-glucosamine kinase	NAGK	5.3812E-79
N-acetylgalactosamine-6-sulfatase	GALNS	0
N-acetylglucosamine-6-sulfatase	GNS	6.757E-21
N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase	B3GNT1	2.6315E-09
N-acetylmuramoyl-L-alanine amidase	PGLYRP2	6.6027E-10
N-alpha-acetyltransferase 38, NatC auxiliary subunit	NAA38	8.6411E-06
Napsin-A	NAPSA	5.1129E-39
Nascent polypeptide-associated complex subunit alpha	NACA	2.8696E-22
Natural cytotoxicity triggering receptor 3 ligand 1	NCR3LG1	1.5819E-54
Neogenin	NEO1	2.8916E-10
Neprilysin	MME	3.932E-22
Nesprin-1	SYNE1	8.6197E-25
Neudesin	NENF	1.6433E-28
Neural cell adhesion molecule 1	NCAM1	4.2033E-52
Neural cell adhesion molecule 2	NCAM2	1.7102E-27
Neural cell adhesion molecule L1-like protein;Processed neural cell adhesion molecule L1-like protein	CHL1	2.3347E-15
Neurexin-1-alpha;Neurexin-1-beta	NRXN1	7.9154E-16
Neurexin-3-alpha	NRXN3	2.2056E-17
Neuroblast differentiation-associated protein AHNAK	AHNAK	2.4021E-14
Neuroblastoma suppressor of tumorigenicity 1	NBL1	0.000001326
Neuroendocrine protein 7B2;N-terminal peptide;C-terminal peptide	SCG5	1.4416E-09
Neurogenic locus notch homolog protein 1;Notch 1 extracellular truncation;Notch 1 intracellular domain	NOTCH1	4.8959E-22
Neurogenic locus notch homolog protein 2;Notch 2 extracellular truncation;Notch 2 intracellular domain;Notch homolog 2 N-terminal-like protein	NOTCH2;NOTCH2NL	6.3678E-19

Neurogenic locus notch homolog protein 3;Notch 3 extracellular truncation;Notch 3 intracellular domain	NOTCH3	0.00075882
Neuronal growth regulator 1	NEGR1	2.1082E-106
Neuropathy target esterase	PNPLA6	1.0098E-18
Neuroplastin	NPTN;DKFZp566H1924	3.8637E-30
Neurosecretory protein VGF;Neuroendocrine regulatory peptide-1;Neuroendocrine regulatory peptide-2	VGF	9.0451E-140
Neuroserpin	SERPINI1	9.7841E-15
Neurotrimin	NTM	1.0248E-25
Neutral alpha-glucosidase AB	GANAB	7.0249E-17
Neutrophil collagenase	MMP8	0
Neutrophil cytosol factor 1;Putative neutrophil cytosol factor 1B;Putative neutrophil cytosol factor 1C	NCF1;NCF1B;NCF1C	7.1036E-47
Neutrophil cytosol factor 2	NCF2	1.5794E-184
Neutrophil cytosol factor 4	NCF4	6.5278E-47
Neutrophil defensin 1;HP 1-56;Neutrophil defensin 2;Neutrophil defensin 3;HP 3-56;Neutrophil defensin 2	DEFA1;DEFA3	0
Neutrophil defensin 4	DEFA4	4.4171E-11
Neutrophil elastase	ELANE	2.1703E-20
Neutrophil gelatinase-associated lipocalin	LCN2	1.8056E-20
NF-kappa-B essential modulator	IKBKG	2.3657E-14
NHL repeat-containing protein 3	NHLRC3	5.352E-42
Nicotinamide phosphoribosyltransferase	NAMPT;RP11-92J19.4	6.1068E-13
Nidogen-1	NID1	1.1245E-26
Non-histone chromosomal protein HMG-14	HMGN1	0.00040597
Non-POU domain-containing octamer-binding protein	NONO	5.6908E-61
Non-secretory ribonuclease	RNASE2	1.4883E-36
NSFL1 cofactor p47	NSFL1C	4.3524E-06
N-sulphoglucosamine sulphohydrolase	SGSH	2.0365E-20
Nuclear mitotic apparatus protein 1	NUMA1	6.4311E-20
Nuclear transport factor 2	NUTF2	0.000027708

Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	NUCKS1	2.7916E-11
Nucleobindin-1	NUCB1	1.6792E-259
Nucleobindin-2	NUCB2	0.000018273
Nucleolin	NCL	2.2206E-101
Nucleolysin TIA-1 isoform p40;Nucleolysin TIAR	TIA1;TIAL1	3.7988E-40
Nucleoprotein TPR	TPR	0
Nucleoside diphosphate kinase A	NME1	3.2241E-11
Nucleoside diphosphate kinase;Nucleoside diphosphate kinase B;Nucleoside diphosphate kinase A;Putative nucleoside diphosphate kinase	NME1-NME2;NME2;NME1;NME2P1	1.343E-36
Nucleosome assembly protein 1-like 4	NAP1L4b;NAP1L4	7.6896E-94
Olfactomedin-4	OLFM4	7.8163E-112
Opioid growth factor receptor	OGFR	6.723E-19
Opioid-binding protein/cell adhesion molecule	OPCML	1.8769E-271
Osteoclast-associated immunoglobulin-like receptor	OSCAR	1.5602E-09
Osteoclast-stimulating factor 1	OSTF1	9.0067E-25
Osteopontin	SPP1	2.1337E-23
OX-2 membrane glycoprotein	CD200	0.000017265
Oxidized low-density lipoprotein receptor 1;Oxidized low-density lipoprotein receptor 1, soluble form	OLR1	7.2621E-285
Palmitoyl-protein thioesterase 1	PPT1	7.4552E-12
Pancreatic alpha-amylase	AMY2A	4.2343E-24
Pancreatic secretory granule membrane major glycoprotein GP2	GP2	1.1999E-189
Pancreatic secretory trypsin inhibitor	SPINK1	0.00029695
Pantetheinase	VNN1	5.1661E-86
Pappalysin-2	PAPPA2	3.3172E-11
Parvalbumin alpha	PVALB	0.0003115
PDZ and LIM domain protein 2	PDLIM2	2.1427E-83
PDZ and LIM domain protein 5	PDLIM5	4.2033E-210
PDZ and LIM domain protein 7	PDLIM7	9.5025E-52
Peflin	PEF1	4.2146E-23

Pepsin A-3;Pepsin A-4;Pepsin A-5	PGA3;PGA4;PGA5	0.00024016
Pepsin A-4;Pepsin A-3;Pepsin A-5	PGA4;PGA3;PGA5	2.3483E-21
Peptidase inhibitor 16	PI16	2.4641E-08
Peptidoglycan recognition protein 1	PGLYRP1	0
Peptidyl-prolyl cis-trans isomerase A;Peptidyl-prolyl cis-trans isomerase	PPIA	4.5778E-07
Peptidyl-prolyl cis-trans isomerase B	PPIB	3.3756E-08
Peptidyl-prolyl cis-trans isomerase C	PPIC	6.428E-22
Peptidyl-prolyl cis-trans isomerase FKBP1A;Peptidyl-prolyl cis-trans isomerase	FKBP1A;FKBP12-Exip2	7.7857E-23
Peptidyl-prolyl cis-trans isomerase FKBP3	FKBP3	2.4066E-244
Perilipin-3	PLIN3	5.8408E-25
Periplakin	PPL	4.3399E-148
Peroxiredoxin-1	PRDX1	0
Peroxiredoxin-2	PRDX2	2.1383E-07
Peroxiredoxin-5, mitochondrial	PRDX5	1.0164E-07
Peroxiredoxin-6	PRDX6	0
Peroxisomal biogenesis factor 19	PEX19	0.00011599
Peroxisomal multifunctional enzyme type 2;(3R)-hydroxyacyl-CoA dehydrogenase;Enoyl-CoA hydratase 2	HSD17B4	3.6672E-58
PERQ amino acid-rich with GYF domain-containing protein 2	GIGYF2	0
Phenylalanine--tRNA ligase alpha subunit	FARSA	8.3739E-22
Phosphatidylcholine-sterol acyltransferase	LCAT	1.2731E-25
Phosphatidylethanolamine-binding protein 1;Hippocampal cholinergic neurostimulating peptide	PEBP1	0.00005255
Phosphatidylethanolamine-binding protein 4	PEBP4	9.5606E-259
Phosphatidylinositol 3-kinase regulatory subunit alpha	PIK3R1	6.9144E-08
Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit alpha	PIK3C2A	3.3539E-10
Phosphatidylinositol-binding clathrin assembly protein;Clathrin coat assembly protein AP180	PICALM;SNAP91	1.5433E-09
Phosphoglucomutase-1	PGM1	6.4922E-21
Phosphoglycerate kinase 1;Phosphoglycerate kinase	PGK1	0
Phosphoglycerate mutase 1;Phosphoglycerate mutase 2;Probable phosphoglycerate	PGAM1;PGAM2;PGAM	8.9729E-112

mutase 4	4	
Phosphoinositide-3-kinase-interacting protein 1	PIK3IP1	8.5109E-21
Phospholipase B-like 1	PLBD1	0
Phospholipase D3	PLD3	8.1543E-88
Phospholipid hydroperoxide glutathione peroxidase, mitochondrial	GPX4	1.094E-103
Phospholysine phosphohistidine inorganic pyrophosphate phosphatase	LHPP	0.00027849
Phosphoserine aminotransferase	PSAT1	2.7254E-296
Phosphoserine phosphatase	PSPH	2.1442E-37
Pigment epithelium-derived factor	SERPINF1	1.2662E-61
Pituitary tumor-transforming gene 1 protein-interacting protein	PTTG1IP	1.2386E-88
PKHD domain-containing transmembrane protein C17orf101	C17orf101	1.6032E-30
Plasma cell-induced resident endoplasmic reticulum protein	PACAP	0.00081754
Plasma kallikrein;Plasma kallikrein heavy chain;Plasma kallikrein light chain	KLKB1	0.00057808
Plasma protease C1 inhibitor	SERPING1	9.5467E-07
Plasma serine protease inhibitor	SERPINA5	3.0194E-14
Plasminogen activator inhibitor 1 RNA-binding protein	SERBP1	0
Plasminogen;Plasmin heavy chain A;Activation peptide;Angiostatin;Plasmin heavy chain A, short form;Plasmin light chain B	PLG	1.579E-147
Plastin-2	LCP1	0.0017234
Platelet basic protein;Connective tissue-activating peptide III;TC-2;Connective tissue-activating peptide III(1-81);Beta-thromboglobulin;Neutrophil-activating peptide 2(74);Neutrophil-activating peptide 2(73);Neutrophil-activating peptide 2;TC-1;Neutrophil-activating peptide 2(1-66);Neutrophil-activating peptide 2(1-63)	PPBP	0.000028055
Platelet endothelial cell adhesion molecule	PECAM1	6.3772E-73
Platelet glycoprotein Ib alpha chain;Glycocalicin	GP1BA	3.3171E-91
Platelet glycoprotein VI	GP6	2.9529E-37
Platelet-activating factor acetylhydrolase IB subunit alpha	PAFAH1B1	1.5898E-86
Platelet-derived growth factor receptor alpha	PDGFRA	0.000001941
Platelet-derived growth factor receptor beta	PDGFRB	7.947E-07
Platelet-derived growth factor subunit B	PDGFB	2.5177E-17
Pleckstrin	PLEK	0.00005506

Pleckstrin homology domain-containing family G member 5	PLEKHG5	2.855E-07
Pleckstrin homology domain-containing family O member 2	PLEKHO2	6.9268E-07
Plectin	PLEC	8.4105E-14
PML-RARA-regulated adapter molecule 1	PRAM1	1.0377E-225
Podocalyxin	PODXL	7.1584E-14
Poliovirus receptor	PVR	0
Poliovirus receptor-related protein 1	PVRL1	0.000045827
Poliovirus receptor-related protein 2	PVRL2	2.2558E-21
Poliovirus receptor-related protein 4;Processed poliovirus receptor-related protein 4	PVRL4	0
Poly [ADP-ribose] polymerase 1	PARP1	9.4337E-67
Poly [ADP-ribose] polymerase 4	PARP4	2.4401E-86
Poly(rC)-binding protein 1;Poly(rC)-binding protein 3;Poly(rC)-binding protein 2	PCBP1;PCBP3;PCBP2	1.124E-11
Poly(rC)-binding protein 2	PCBP2	8.6827E-280
Polyadenylate-binding protein 1;Polyadenylate-binding protein 3	PABPC1;PABPC3	2.9738E-253
Polyadenylate-binding protein 2	PABPN1	6.9619E-22
Polymeric immunoglobulin receptor;Secretory component	PIGR	2.3624E-10
Polypyrimidine tract-binding protein 1	PTBP1	0.00054171
Polyubiquitin-C;Ubiquitin;Polyubiquitin-B;Ubiquitin;Ubiquitin-40S ribosomal protein S27a;Ubiquitin;40S ribosomal protein S27a;Ubiquitin-60S ribosomal protein L40;Ubiquitin;60S ribosomal protein L40	UBC;UBB;RPS27A;UBA52	1.4682E-07
Prefoldin subunit 2	PFDN2	0.0001007
Prefoldin subunit 3	VBP1	1.7412E-13
Prefoldin subunit 4	PFDN4	1.5891E-90
Prefoldin subunit 5	PFDN5	0
Prefoldin subunit 6	PFDN6	1.5871E-50
Pregnancy zone protein	PZP	0.000014667
Prelamin-A/C;Lamin-A/C	LMNA	1.42E-19
Pre-mRNA branch site protein p14	SF3B14	0.0018698
Pre-mRNA-processing-splicing factor 8	PRPF8	3.3983E-09
Prion-like protein doppel	PRND	2.1107E-47
Proactivator polypeptide;Saposin-A;Saposin-B-Val;Saposin-B;Saposin-C;Saposin-D	PSAP	0.00030156

Proactivator polypeptide-like 1;Saposin A-like;Saposin B-Val-like;Saposin B-like;Saposin C-like;Saposin D-like	PSAPL1	6.0607E-150
Probable ATP-dependent RNA helicase DDX46	DDX46	0.000025976
Probable ATP-dependent RNA helicase DDX6	DDX6	3.5699E-29
Probable G-protein coupled receptor 142	GPR142	1.2399E-209
Probable serine carboxypeptidase CPVL	CPVL	1.8739E-41
Pro-cathepsin H;Cathepsin H mini chain;Cathepsin H;Cathepsin H heavy chain;Cathepsin H light chain	CTSH	0
Procollagen C-endopeptidase enhancer 1	PCOLCE	7.4454E-226
Procollagen C-endopeptidase enhancer 2	PCOLCE2	3.2025E-47
Proenkephalin-A;Synenkephalin;Met-enkephalin;PENK(114-133);PENK(143-183);Met-enkephalin-Arg-Gly-Leu;Leu-enkephalin;PENK(237-258);Met-enkephalin-Arg-Phe	PENK	1.1552E-18
Pro-epidermal growth factor;Epidermal growth factor	EGF	3.3017E-33
Profilin-1	PFN1	1.6784E-18
Programmed cell death 1 ligand 2	PDCD1LG2	9.838E-07
Programmed cell death 6-interacting protein	PDCD6IP	4.5872E-63
Programmed cell death protein 10	PDCD10	0.00098775
Programmed cell death protein 6	PDCD6	0
Prolactin-inducible protein	PIP	4.8911E-07
Prolargin	PRELP	8.3793E-225
Proline-rich acidic protein 1	ZNF511;PRAP1	5.4105E-69
Proline-serine-threonine phosphatase-interacting protein 1	PSTPIP1	1.0535E-42
Prolow-density lipoprotein receptor-related protein 1;Low-density lipoprotein receptor-related protein 1 85 kDa subunit;Low-density lipoprotein receptor-related protein 1 515 kDa subunit;Low-density lipoprotein receptor-related protein 1 intracellular domain	LRP1	2.2504E-119
Promotilin;Motilin;Motilin-associated peptide	MLN	1.4026E-08
Properdin	CFP	0.00025582
ProSAAS;KEP;Big SAAS;Little SAAS;Big PEN-LEN;PEN;Little LEN;Big LEN	PCSK1N	1.6341E-20
Prostaglandin E synthase 3	PTGES3	0.00017051
Prostaglandin E2 receptor EP3 subtype	PTGER3;EP3-I	1.2709E-35
Prostaglandin reductase 1	PTGR1	0

Prostaglandin-H2 D-isomerase	PTGDS	6.6223E-85
Prostasin;Prostasin light chain;Prostasin heavy chain	PRSS8	3.9109E-07
Prostate stem cell antigen	PSCA	5.1348E-50
Prostate-specific antigen	KLK3	9.8675E-14
Prostatic acid phosphatase;PAPf39	ACPP	0.00098746
Proteasomal ubiquitin receptor ADRM1	ADRM1	8.7697E-180
Proteasome activator complex subunit 1	PSME1	2.6041E-42
Proteasome activator complex subunit 2	PSME2	2.7771E-27
Proteasome inhibitor PI31 subunit	PSMF1	1.7527E-92
Proteasome subunit alpha type-5	PSMA5	7.7842E-189
Proteasome subunit alpha type-7	PSMA7	3.8074E-15
Proteasome subunit beta type-4	PSMB4	0
Proteasome subunit beta type-8	PSMB8	8.6375E-91
Protein AHNAK2	AHNAK2	9.1572E-60
Protein AMBP;Alpha-1-microglobulin;Inter-alpha-trypsin inhibitor light chain;Trypstatin	AMBP	1.0683E-06
Protein BRICK1	BRK1	0.0055026
Protein canopy homolog 2	CNPY2	2.8403E-29
Protein canopy homolog 3	CNPY3	1.6561E-112
Protein CDV3 homolog	CDV3	5.8154E-17
Protein CREG1	CREG1	0.000003366
Protein CutA	CUTA	9.9577E-43
Protein CYR61	CYR61	0
Protein DDI1 homolog 2	DDI2	4.4783E-55
Protein DEK	DEK	0
Protein delta homolog 1;Fetal antigen 1	DLK1	1.2542E-27
Protein delta homolog 2	DLK2;EGFL9	5.2808E-08
Protein disulfide-isomerase	P4HB	0
Protein disulfide-isomerase A3	PDIA3	1.1652E-12
Protein disulfide-isomerase A4	PDIA4	0.00075573
Protein disulfide-isomerase A6	PDIA6	1.561E-17

Protein DJ-1	PARK7	0.0002899
Protein dpy-30 homolog	DPY30	0.00025685
Protein FADD	FADD	2.2308E-15
Protein FAM107B	FAM107B	1.9331E-33
Protein FAM3B	FAM3B	6.599E-34
Protein FAM3C	FAM3C	1.6813E-85
Protein FAM49B	FAM49B	6.2952E-16
Protein FAM63A	FAM63A	3.4913E-32
Protein flightless-1 homolog	FLII	4.7671E-16
Protein HEG homolog 1	HEG1	5.103E-19
Protein Hook homolog 3	HOOK3	1.4579E-15
Protein kinase C and casein kinase substrate in neurons protein 2	PACSIN2	2.1398E-08
Protein kinase C beta type	PRKCB	1.7941E-07
Protein NOV homolog	NOV	1.4271E-16
Protein NOXP20	FAM114A1;DKFZp686F20250	4.1668E-06
Protein phosphatase 1 regulatory subunit 12A	PPP1R12A	0.00026672
Protein phosphatase 1B	PPM1B	3.1114E-21
Protein S100-A11	S100A11	0.000037022
Protein S100-A12;Calcitermin	S100A12	1.1599E-09
Protein S100-A2	S100A2	0.000011308
Protein S100-A4	S100A4	8.3206E-16
Protein S100-A6	S100A6	5.4153E-39
Protein S100-A7;Protein S100-A7A	S100A7;S100A7A	5.6439E-07
Protein S100-A8	S100A8	0.00066527
Protein S100-A9	S100A9	1.5113E-36
Protein S100-P	S100P	2.8438E-06
Protein SET	SET	3.7341E-228
Protein shisa-2 homolog	SHISA2	8.2721E-10
Protein shisa-5	SHISA5	1.1944E-19

Protein shisa-6 homolog	SHISA6	1.037E-141
Protein shisa-7	SHISA7	1.99E-14
Protein sidekick-1	SDK1	1.3575E-48
Protein unc-13 homolog D	UNC13D	5.5821E-51
Protein unc-79 homolog	UNC79	0
Protein Wnt-7a	WNT7A	6.2955E-07
Protein YIPF3	YIPF3	3.2234E-11
Protein-arginine deiminase type-4	PADI4	4.1421E-36
Protein-glutamine gamma-glutamyltransferase 4	TGM4	3.4372E-07
Protein-L-isoaspartate O-methyltransferase;Protein-L-isoaspartate(D-aspartate) O-methyltransferase	PCMT1	4.1822E-57
Proteoglycan 4;Proteoglycan 4 C-terminal part	PRG4	2.5048E-26
Prothrombin;Activation peptide fragment 1;Activation peptide fragment 2;Thrombin light chain;Thrombin heavy chain	F2	4.7376E-12
Prothymosin alpha;Thymosin alpha-1	PTMA	9.7613E-16
Protocadherin gamma-C3	PCDHGC3	9.2144E-88
Protocadherin-1	PCDH1	3.7166E-07
Protocadherin-16	DCHS1	1.1146E-11
PTB domain-containing engulfment adapter protein 1	GULP1	6.4306E-07
Purine nucleoside phosphorylase	PNP	0.0016591
Putative coiled-coil-helix-coiled-coil-helix domain-containing protein CHCHD2P9, mitochondrial;Coiled-coil-helix-coiled-coil-helix domain-containing protein 2, mitochondrial	CHCHD2P9;CHCHD2	2.015E-70
Putative macrophage-stimulating protein MSTP9;Hepatocyte growth factor-like protein;Hepatocyte growth factor-like protein alpha chain;Hepatocyte growth factor-like protein beta chain	MST1;MST1P9	2.2778E-101
Putative neutrophil cytosol factor 1B;Neutrophil cytosol factor 1;Putative neutrophil cytosol factor 1C	NCF1B;NCF1;NCF1C	3.0812E-09
Putative phospholipase B-like 2;Putative phospholipase B-like 2 32 kDa form;Putative phospholipase B-like 2 45 kDa form	PLBD2	1.9824E-12
Putative RNA-binding protein Luc7-like 2	LUC7L2	2.026E-69
Pyridoxal kinase	PDXK	0.00034511

Pyruvate kinase isozymes M1/M2;Pyruvate kinase	PKM2	0.000007596
Quinone oxidoreductase-like protein 1	CRYZL1	5.7297E-11
Rab GDP dissociation inhibitor beta	GDI2	4.6328E-60
Rab11 family-interacting protein 1	RAB11FIP1	0.000072498
Radixin	RDX	0.0021553
Ras association domain-containing protein 5	RASSF5	9.8774E-09
Ras GTPase-activating-like protein IQGAP1	IQGAP1	9.5937E-12
Ras GTPase-activating-like protein IQGAP2	IQGAP2	7.2012E-71
RAS protein activator like-3	RASAL3	1.3061E-08
Ras-related C3 botulinum toxin substrate 1;Ras-related C3 botulinum toxin substrate 2;Ras-related C3 botulinum toxin substrate 3	RAC1;RAC2;RAC3	3.6278E-50
Ras-related protein Rab-10	RAB10	5.9116E-07
Ras-related protein Rab-11B;Ras-related protein Rab-11A	RAB11B;RAB11A	9.979E-24
Ras-related protein Rab-14	RAB14	8.3438E-50
Ras-related protein Rab-18	RAB18	0.00096262
Ras-related protein Rab-1A;Putative Ras-related protein Rab-1C;Ras-related protein Rab-1B	RAB1A;RAB1C;RAB1B	7.9102E-113
Ras-related protein Rab-1B;Putative Ras-related protein Rab-1C;Ras-related protein Rab-15;Ras-related protein Rab-8A;Ras-related protein Rab-8B;Ras-related protein Rab-1A;Ras-related protein Rab-13;Ras-related protein Rab-10	RAB1B;RAB1C;RAB15; RAB8A;RAB8B;RAB1A; RAB13;RAB10	0
Ras-related protein Rab-21	RAB21	2.0185E-12
Ras-related protein Rab-24	RAB24	3.5101E-12
Ras-related protein Rab-2A	RAB2A	0.0016782
Ras-related protein Rab-3D	RAB3D	0.00020272
Ras-related protein Rab-44	RAB44	7.1605E-179
Ras-related protein Rab-4B	RAB4B	0.000016823
Ras-related protein Rab-5A	RAB5A	3.1766E-11
Ras-related protein Rab-5B	RAB5B	9.316E-07
Ras-related protein Rab-5C	RAB5C	8.4937E-07
Ras-related protein Rab-6A;Ras-related protein Rab-6B	RAB6A;RAB6B	0
Ras-related protein Rab-6A;Ras-related protein Rab-6B;Ras-related protein Rab-39A	RAB6A;RAB6B;RAB39A	3.0053E-06

Ras-related protein Rab-7a	RAB7A	1.272E-48
Ras-related protein Rab-8B;Ras-related protein Rab-8A	RAB8B;RAB8A	3.2455E-15
Ras-related protein Rap-1b;Ras-related protein Rap-1b-like protein;Ras-related protein Rap-1A	RAP1B;RAP1A	0.000067132
Receptor-type tyrosine-protein phosphatase C	PTPRC	1.8677E-15
Receptor-type tyrosine-protein phosphatase delta	PTPRD	1.8014E-12
Receptor-type tyrosine-protein phosphatase eta	PTPRJ	1.4667E-112
Receptor-type tyrosine-protein phosphatase F	PTPRF	0.000050681
Receptor-type tyrosine-protein phosphatase gamma	PTPRG	1.9184E-06
Receptor-type tyrosine-protein phosphatase S	PTPRS	1.3243E-06
Receptor-type tyrosine-protein phosphatase-like N	PTPRN	0.000092717
Replication protein A 70 kDa DNA-binding subunit	RPA1	0.000012743
Resistin	RETN	0
Reticulocalbin-1	RCN1	0
Reticulon-3	RTN3	5.9758E-18
Reticulon-4	RTN4	2.1174E-12
Reticulon-4 receptor-like 2	RTN4RL2	2.1263E-112
Retinoic acid-induced protein 3	GPRC5A	0.0010332
Retinoid-binding protein 7	RBP7	0.000095674
Retinoid-inducible serine carboxypeptidase	SCPEP1	0.00096754
Retinol-binding protein 4;Plasma retinol-binding protein(1-182);Plasma retinol-binding protein(1-181);Plasma retinol-binding protein(1-179);Plasma retinol-binding protein(1-176)	RBP4	0.00041337
Rho GDP-dissociation inhibitor 1	ARHGDIA	1.9685E-61
Rho GDP-dissociation inhibitor 2	ARHGDIB	1.7975E-26
Rho GTPase-activating protein 1	ARHGAP1	0.00000766
Rho GTPase-activating protein 25	ARHGAP25	0.000025493
Rho GTPase-activating protein 4	ARHGAP4	2.4324E-137
Rho guanine nucleotide exchange factor 1	ARHGEF1	3.7524E-07
Rho guanine nucleotide exchange factor 6	ARHGEF6	0.00098066
Rho-associated protein kinase 1	ROCK1	2.2057E-12

Rho-related GTP-binding protein RhoG;Rho-related GTP-binding protein RhoJ;Rho-related GTP-binding protein RhoQ	RHOG;RHOJ;RHOQ;RAC2	8.0379E-09
Ribonuclease 4	RNASE4	0.0013425
Ribonuclease K6	RNASE6	1.0151E-13
Ribonuclease pancreatic	RNASE1	5.7408E-09
Ribonuclease T2	RNASET2	8.0344E-09
Ribonuclease UK114	HRSP12	0.0026228
Ribosyldihydronicotinamide dehydrogenase [quinone]	NQO2	3.3057E-106
RNA-binding motif protein, X chromosome;RNA-binding motif protein, X chromosome, N-terminally processed;RNA binding motif protein, X-linked-like-1;RNA binding motif protein, X-linked-like-1, N-terminally processed;RNA-binding motif protein, X-linked-like-3;RNA-binding motif protein, X-linked-like-2	RBMX;RBMXL1;RBMXL3;RBMXL2	7.413E-12
RNA-binding protein 39	RBM39	0.00044215
RNA-binding protein EWS	EWSR1	0.0014263
RNA-binding protein Raly	RALY	1.5902E-54
Roundabout homolog 4	ROBO4	8.9708E-54
RUN and FYVE domain-containing protein 1	RUFY1	4.9893E-16
Salivary acidic proline-rich phosphoprotein 1/2;Salivary acidic proline-rich phosphoprotein 1/2;Salivary acidic proline-rich phosphoprotein 3/4;Peptide P-C	PRH1	1.3073E-22
SAM and SH3 domain-containing protein 3	SASH3	2.7761E-45
Scaffold attachment factor B1	SAFB	1.6018E-151
Sec1 family domain-containing protein 1	SCFD1	0.00084326
Secreted and transmembrane protein 1	SECTM1	1.7171E-11
Secreted frizzled-related protein 3	FRZB	2.8259E-15
Secreted frizzled-related protein 4	SFRP4	2.4068E-08
Secreted Ly-6/uPAR-related protein 1	SLURP1	8.2543E-31
Secretoglobin family 1D member 2	SCGB1D2	0.0022081
Secretogranin-1;GAWK peptide;CCB peptide	CHGB	0
Secretogranin-2;Secretoneurin	SCG2	2.0418E-06
Selenium-binding protein 1	SELENBP1	3.5301E-08
Selenoprotein P	SEPP1	0.00040298

Semenogelin-1;Alpha-inhibin-92;Alpha-inhibin-31;Seminal basic protein	SEMG1	5.0184E-10
Semenogelin-2	SEMG2	5.7838E-12
Septin-6	06-Sep	3.5832E-07
Septin-9		4.2313E-19
Serglycin	SRGN	7.9032E-124
Serine palmitoyltransferase 2;Serine palmitoyltransferase 3	SPTLC2;SPTLC3	0.00018738
Serine protease 23	PRSS23	0.000042995
Serine protease hepsin;Serine protease hepsin non-catalytic chain;Serine protease hepsin catalytic chain	HPN	2.9022E-25
Serine protease inhibitor Kazal-type 5;Hemofiltrate peptide HF6478;Hemofiltrate peptide HF7665	SPINK5	8.4887E-57
Serine/arginine-rich splicing factor 3	SRSF3	0.00072408
Serine/arginine-rich splicing factor 7	SRSF7	0.0010681
Serine/threonine-protein kinase 10	STK10	2.5472E-26
Serine/threonine-protein kinase 4;Serine/threonine-protein kinase 4 37kDa subunit;Serine/threonine-protein kinase 4 18kDa subunit	STK4	5.4856E-06
Serine/threonine-protein kinase OSR1	OXSR1	0.0011193
Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	PPP2R1A	4.9196E-41
Serine/threonine-protein phosphatase 2A activator	PPP2R4	5.9748E-11
Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	PPP2CA;SKP1	1.0768E-280
Serine/threonine-protein phosphatase 2A catalytic subunit beta isoform;Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform	PPP2CB;PPP2CA	0.0017343
Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform;Serine/threonine-protein phosphatase;Serine/threonine-protein phosphatase 2B catalytic subunit beta isoform	PPP3CA;PPP3CB	0.000012134
Serine/threonine-protein phosphatase 4 regulatory subunit 3B	SMEK2	3.796E-128
Serine/threonine-protein phosphatase 5;Serine/threonine-protein phosphatase	PPP5C	6.9518E-06
Serine/threonine-protein phosphatase;Serine/threonine-protein phosphatase PP1-alpha catalytic subunit;Serine/threonine-protein phosphatase PP1-beta catalytic subunit;Serine/threonine-protein phosphatase PP1-gamma catalytic subunit	PPP1CC;PPP1CA;PPP1CB	0.0011331
Serotransferrin	TF	9.2335E-20
Serpin B10	SERPINB10	2.6072E-12

Serpin B13	SERPINB13	1.5895E-14
Serpin B3;Serpin B4	SERPINB3;SERPINB4	4.3017E-08
Serpin B4	SERPINB4	0.00057836
Serpin B6	SERPINB6	0.001738
Serrate RNA effector molecule homolog	SRRT	4.3149E-39
Serum amyloid A protein	SAA1	0.0021199
Serum amyloid A-4 protein	SAA4	0.000094879
Serum amyloid P-component;Serum amyloid P-component(1-203)	APCS	9.26E-136
Serum paraoxonase/arylesterase 1	PON1	1.9185E-25
Serum paraoxonase/arylesterase 1	PON1	2.7022E-15
Sex hormone-binding globulin	SHBG	0.000046426
S-formylglutathione hydrolase	ESD	2.7478E-24
SH3 domain-binding glutamic acid-rich-like protein	SH3BGRL	0
SH3 domain-binding glutamic acid-rich-like protein 3	SH3BGRL3	1.0566E-53
SH3 domain-containing kinase-binding protein 1	SH3KBP1	0.000033055
Sialate O-acetylesterase	SIAE	4.1032E-07
Sialic acid-binding Ig-like lectin 16	SIGLEC16	0.000042457
Signal peptide, CUB and EGF-like domain-containing protein 1	SCUBE1	1.2607E-141
Signal peptide, CUB and EGF-like domain-containing protein 2	SCUBE2	2.5477E-24
Signal transducer and activator of transcription 3	STAT3	4.0259E-12
Signal transducer and activator of transcription 6	STAT6	0.00079751
Signal transducing adapter molecule 1	STAM	3.5512E-148
Signal-regulatory protein beta-1	SIRPB1	2.9306E-11
SLAIN motif-containing protein 1	SLAIN1	1.1016E-12
SLAM family member 5	CD84	8.736E-24
SLIT and NTRK-like protein 1	SLITRK1	4.6265E-12
Slit homolog 3 protein	SLIT3	2.7564E-07
Small nuclear ribonucleoprotein F	SNRPF	1.5259E-33
Small nuclear ribonucleoprotein Sm D1	SNRPD1	0.00018392
Small nuclear ribonucleoprotein Sm D2	SNRPD2	0.00011433

Small nuclear ribonucleoprotein Sm D3	SNRPD3	0.00028981
Small nuclear ribonucleoprotein-associated protein;Small nuclear ribonucleoprotein-associated proteins B and B;Small nuclear ribonucleoprotein-associated protein N	SNRPN;SNRPB	8.9763E-60
Small proline-rich protein 2G;Small proline-rich protein 2E;Small proline-rich protein 2D;Small proline-rich protein 2B;Small proline-rich protein 2A;Small proline-rich protein 2F	SPRR2G;SPRR2E;SPRR2D;SPRR2B;SPRR2A;SPRR2F	2.2313E-16
Small proline-rich protein 3	SPRR3	9.9192E-10
Small ubiquitin-related modifier 2	SUMO2	3.2962E-28
Small ubiquitin-related modifier 3;Small ubiquitin-related modifier 2;Small ubiquitin-related modifier 4	SUMO3;SUMO2;SUMO4	0.00002921
S-methyl-5-thioadenosine phosphorylase	MTAP	1.3227E-06
SNARE-associated protein Snapin	SNAPIN	7.7213E-06
Sodium/nucleoside cotransporter 1	SLC28A1	1.4882E-14
Sodium/potassium-transporting ATPase subunit gamma	FXYD2	2.6995E-41
Solute carrier family 12 member 2	SLC12A2	0.00005537
Solute carrier family 12 member 3	SLC12A3	1.3756E-09
Somatostatin;Somatostatin-28;Somatostatin-14	SST	0.00079751
Sorcin	SRI	5.1601E-15
Sortilin	SORT1	3.4444E-16
Sortilin-related receptor	SORL1	4.1025E-06
Sorting nexin-12	SNX12	6.7618E-59
Sorting nexin-27	SNX27	3.0303E-41
Sorting nexin-5	SNX5	0.000087774
SPARC-like protein 1	SPARCL1	6.7703E-53
Spastin	SPAST	1.4846E-20
Spectrin alpha chain, brain	SPTAN1	6.7188E-87
Spectrin beta chain, brain 1	SPTBN1	7.8242E-196
S-phase kinase-associated protein 1	SKP1	3.1333E-17
Splicing factor 1	SF1	1.286E-34
Splicing factor 3A subunit 1	SF3A1	0.000016194
Splicing factor 3B subunit 1	SF3B1	6.78E-13

Splicing factor 3B subunit 2	SF3B2	2.1051E-11
Splicing factor U2AF 65 kDa subunit	U2AF2	1.9511E-07
Splicing factor, proline- and glutamine-rich	SFPQ	3.6778E-12
Src kinase-associated phosphoprotein 2	SKAP2	8.9439E-32
Src substrate cortactin	CTTN	9.1402E-09
Stabilin-1	STAB1	0.00037672
Stathmin	STMN1	6.3344E-12
Stathmin	STMN1	0.000016786
Stathmin-2;Stathmin	STMN2	8.7896E-06
Steroid receptor RNA activator 1	SRA1	0.00054309
Stress-induced-phosphoprotein 1	STIP1	9.9047E-37
Stromal cell-derived factor 1;SDF-1-beta(3-72);SDF-1-alpha(3-67)	CXCL12	1.4656E-68
Structural maintenance of chromosomes protein 1A	SMC1A	0.000028059
Structural maintenance of chromosomes protein 3	SMC3	0.0023337
Submaxillary gland androgen-regulated protein 3B;Peptide P-A;Peptide D1A	SMR3B	1.7577E-23
Sulfhydryl oxidase 1	QSOX1	0.00022488
SUMO-conjugating enzyme UBC9	UBE2I	1.6548E-136
SUN domain-containing protein 2	SUN2	0.000011853
Superoxide dismutase [Cu-Zn]	SOD1	7.8395E-50
Superoxide dismutase [Mn], mitochondrial;Superoxide dismutase	SOD2	0.00048122
Sushi domain-containing protein 2	SUSD2	0.000010205
Sushi domain-containing protein 5	SUSD5	0.00003183
SWI/SNF complex subunit SMARCC2	SMARCC2	2.7261E-122
Switch-associated protein 70	SWAP70	8.1493E-43
Synaptic vesicle membrane protein VAT-1 homolog	VAT1	7.0519E-09
Syndecan-1	SDC1	0.00094046
Syndecan-4;Syndecan	SDC4	8.3089E-35
Syntaxin-12	STX12	0.00015106
Syntaxin-3	STX3	6.2432E-16
Syntaxin-7	STX7	1.1868E-06

Syntaxin-binding protein 2	STXBP2	1.8344E-30
Syntenin-1	SDCBP	2.189E-32
Talin-1	TLN1	2.0655E-12
Target of Myb protein 1	TOM1	1.2251E-16
Target of Nesh-SH3	ABI3BP	6.4538E-06
Tax1-binding protein 1	TAX1BP1	8.4953E-102
TBC1 domain family member 13	TBC1D13	9.8892E-72
TCDD-inducible poly [ADP-ribose] polymerase	TIPARP	5.5695E-109
T-cell antigen CD7	CD7	8.5851E-06
T-complex protein 1 subunit beta	CCT2	0
T-complex protein 1 subunit delta	CCT4	0.000081077
T-complex protein 1 subunit gamma	CCT3	1.317E-22
T-complex protein 1 subunit theta	CCT8	1.7352E-07
Tenascin	TNC	1.0622E-29
Tenascin-R	TNR	0.0013675
Tenascin-X	TNXB	1.7813E-08
Teratocarcinoma-derived growth factor 1;Putative teratocarcinoma-derived growth factor 3	TDGF1;TDGF1P3	5.0796E-41
Testican-1	SPOCK1	1.4749E-38
Testican-2	SPOCK2	1.6774E-45
Testis- and ovary-specific PAZ domain-containing protein 1	TOPAZ1	4.6426E-07
Testis-expressed sequence 2 protein	TEX2	1.3876E-08
Tetranectin	CLEC3B	2.2205E-07
Tetraspanin-1	TSPAN1	1.1651E-06
Tetratricopeptide repeat protein 26	TTC26	1.1754E-85
TGF-beta receptor type-1	TGFBR1	0.00054177
TGF-beta receptor type-2	TGFBR2	2.6322E-24
Thioredoxin	TXN	0.00093241
Thioredoxin domain-containing protein 17	TXNDC17	0
Thioredoxin domain-containing protein 5	TXNDC5	0.0013631

Thioredoxin reductase 1, cytoplasmic	TXNRD1	2.4647E-11
Thioredoxin-dependent peroxide reductase, mitochondrial	PRDX3	0.00051791
Thioredoxin-like protein 1	TXNL1	0.0043149
Thioredoxin-related transmembrane protein 1	TMX1	0.00012113
THO complex subunit 4	ALYREF	0.0017334
Thrombomodulin	THBD	0.00091547
Thrombospondin-1	THBS1	0.0014005
Thrombospondin-4	THBS4	0.00066992
Thy-1 membrane glycoprotein	THY1	0.000050556
Thymidine phosphorylase	TYMP	0.0065561
Thymidine phosphorylase	TYMP	0.00087215
Thymocyte nuclear protein 1	THYN1	0.00034597
Thymosin beta-10	TMSB10	0.00067785
Thymosin beta-4;Hematopoietic system regulatory peptide	TMSB4X;TMSL4	0.00093329
Thyroxine-binding globulin	SERPINA7	0.0014908
TIP41-like protein	TIPRL	0
Tissue alpha-L-fucosidase	FUCA1	0.00050889
Tissue factor pathway inhibitor	TFPI	4.38E-09
Titin	TTN	0.0010078
Titin	TTN	0
Toll-interacting protein	TOLLIP	0.00058096
Trafficking protein particle complex subunit 3	TRAPPC3	5.3682E-07
Transaldolase	TALDO1	0.00038243
Transcobalamin-1	TCN1	3.2223E-168
Transcobalamin-2	TCN2	0.000025731
Transcription elongation factor A protein-like 5;Transcription elongation factor A protein-like 3	TCEAL5;TCEAL3	7.6616E-16
Transcription initiation factor TFIID subunit 10	TAF10	0.000022255
Transcription intermediary factor 1-beta	TRIM28	7.123E-201
Transformer-2 protein homolog beta	TRA2B	0

Transforming growth factor beta receptor type 3	TGFBR3	1.1147E-79
Transforming growth factor beta-1;Latency-associated peptide	TGFB1	3.1312E-45
Transforming protein RhoA;Rho-related GTP-binding protein RhoC	RHOA;RHOC	0.00026993
Transgelin-2	TAGLN2	1.9228E-26
Trans-Golgi network integral membrane protein 2	TGOLN2	2.2742E-67
Transitional endoplasmic reticulum ATPase	VCP	4.6828E-139
Transketolase	TKT	4.2324E-39
Transketolase	TKT	5.8892E-35
Translationally-controlled tumor protein	TPT1	8.2526E-33
Translin	TSN	8.7261E-14
Transmembrane emp24 domain-containing protein 7	TMED7	6.7463E-56
Transmembrane emp24 domain-containing protein 7	TIRAP3;TMED7;TMED7-TICAM2	4.719E-11
Transmembrane protease serine 2;Transmembrane protease serine 2 non-catalytic chain;Transmembrane protease serine 2 catalytic chain	TMPRSS2	3.0325E-92
Transmembrane protease serine 2;Transmembrane protease serine 2 non-catalytic chain;Transmembrane protease serine 2 catalytic chain	TMPRSS2	0
Transmembrane protein 106B	TMEM106B	0
Transmembrane protein PVRIG	PVRIG	0.00014285
Transthyretin	TTR	2.1834E-28
Trefoil factor 1	TFF1	1.6926E-31
Trefoil factor 2	TFF2	1.8627E-122
Trefoil factor 3	TFF3	1.361E-39
Trefoil factor 3	TFF3	2.9948E-25
Triosephosphate isomerase	TPI1	4.2702E-16
Tripeptidyl-peptidase 1	TPP1	0
Tropomodulin-3	TMOD3	1.7557E-197
Tropomyosin alpha-1 chain	TPM1	2.5756E-08
Tropomyosin alpha-3 chain	TPM3	0
Tropomyosin alpha-3 chain	TPM3	0.0015185
Tropomyosin alpha-4 chain	TPM4	0.00028957

Trypsin-2;Putative trypsin-6	PRSS2;TRY6	0.000034624
Tryptophan--tRNA ligase, cytoplasmic;T1-TrpRS;T2-TrpRS	WARS	5.4196E-06
Tubulin alpha-1B chain;Tubulin alpha-1C chain;Tubulin alpha-1A chain;Tubulin alpha-4A chain	TUBA1C;TUBA1B;TUBA1A;TUBA4A	0.000025223
Tubulin beta chain;Tubulin beta-3 chain;Tubulin beta-4B chain;Tubulin beta-2A chain;Tubulin beta-2B chain	TUBB;TUBB3;TUBB4B;TUBB2A;TUBB2B	2.3574E-06
Tubulin beta-4B chain	TUBB4B	2.1157E-113
Tubulin-specific chaperone A	TBCA	3.1005E-10
Tubulin-specific chaperone A	TBCA	5.7819E-17
Tubulointerstitial nephritis antigen	TINAG	0
Tumor necrosis factor receptor superfamily member 10C	TNFRSF10C	0.00050908
Tumor necrosis factor receptor superfamily member 11A	TNFRSF11A	0.000027912
Tumor necrosis factor receptor superfamily member 12A	TNFRSF12A	3.6316E-13
Tumor necrosis factor receptor superfamily member 14	TNFRSF14	1.7019E-11
Tumor necrosis factor receptor superfamily member 16	NGFR	1.5651E-11
Tumor necrosis factor receptor superfamily member 17	TNFRSF17	1.895E-65
Tumor necrosis factor receptor superfamily member 19	TNFRSF19	4.9731E-08
Tumor necrosis factor receptor superfamily member 19L	RELT	2.5943E-78
Tumor necrosis factor receptor superfamily member 1A;Tumor necrosis factor receptor superfamily member 1A, membrane form;Tumor necrosis factor-binding protein 1	TNFRSF1A	5.2036E-13
Tumor necrosis factor receptor superfamily member 1B;Tumor necrosis factor receptor superfamily member 1b, membrane form;Tumor necrosis factor-binding protein 2	TNFRSF1B;TNFR1B	2.2422E-11
Tumor necrosis factor receptor superfamily member 5	CD40	0.00007015
Tumor protein D52	TPD52	0.00055623
Tumor protein D54	TPD52L2	4.7253E-16
Tumor susceptibility gene 101 protein	TSG101	2.6617E-06
Tumor-associated calcium signal transducer 2	TACSTD2	0
Twinfilin-2	TWF2	0.000017925
Twisted gastrulation protein homolog 1	TWSG1	0.00048633
Tyrosine-protein kinase CSK	CSK	3.0135E-42
Tyrosine-protein kinase HCK	HCK	4.8411E-13

Tyrosine-protein kinase Lyn	LYN	3.0243E-06
Tyrosine-protein kinase receptor UFO	AXL	4.1849E-25
Tyrosine-protein kinase transmembrane receptor ROR1	ROR1	0.00031699
Tyrosine-protein kinase transmembrane receptor ROR2	ROR2	0.00032388
Tyrosine-protein phosphatase non-receptor type 18	PTPN18	0.00013026
Tyrosine-protein phosphatase non-receptor type 6	PTPN6	9.0045E-11
Tyrosine-protein phosphatase non-receptor type substrate 1;Signal-regulatory protein beta-1 isoform 3	SIRPA;SIRPB1	0.000012445
U1 small nuclear ribonucleoprotein A;U2 small nuclear ribonucleoprotein B	SNRPA;SNRPB2	2.2664E-08
U1 small nuclear ribonucleoprotein C	SNRPC	0.00012821
U2 small nuclear ribonucleoprotein A	SNRPA1	3.6315E-16
U2 snRNP-associated SURP motif-containing protein	U2SURP	6.4492E-93
U5 small nuclear ribonucleoprotein 200 kDa helicase	SNRNP200	0.00059929
U6 snRNA-associated Sm-like protein LSm2	LSM2	5.8981E-09
U6 snRNA-associated Sm-like protein LSm3	LSM3	1.7716E-18
U6 snRNA-associated Sm-like protein LSm5	LSM5	0.000080043
U8 snoRNA-decapping enzyme	NUDT16	0.00025739
Ubiquilin-1;Ubiquilin-2;Ubiquilin-4	UBQLN1;UBQLN2;UBQLN4	0.00010789
Ubiquilin-2;Ubiquilin-4;Ubiquilin-1	UBQLN2;UBQLN4;UBQLN1	5.8053E-06
Ubiquitin carboxyl-terminal hydrolase 14;Ubiquitin carboxyl-terminal hydrolase	USP14	0.00080028
Ubiquitin carboxyl-terminal hydrolase 15;Ubiquitin carboxyl-terminal hydrolase	USP15	0.00086398
Ubiquitin carboxyl-terminal hydrolase 5	USP5	2.885E-08
Ubiquitin carboxyl-terminal hydrolase 7;Ubiquitin carboxyl-terminal hydrolase	USP7	2.5882E-152
Ubiquitin carboxyl-terminal hydrolase;Ubiquitin carboxyl-terminal hydrolase 19	USP19	2.1861E-11
Ubiquitin thioesterase OTUB1	OTUB1	0.00008151
Ubiquitin-associated and SH3 domain-containing protein B	UBASH3B	0.00028411
Ubiquitin-conjugating enzyme E2 N;Putative ubiquitin-conjugating enzyme E2 N-like	UBE2N;UBE2NL	0.000018375
Ubiquitin-conjugating enzyme E2 variant 1;Ubiquitin-conjugating enzyme E2 variant 2	UBE2V1;TMEM189;UBE2V2	0.000049275

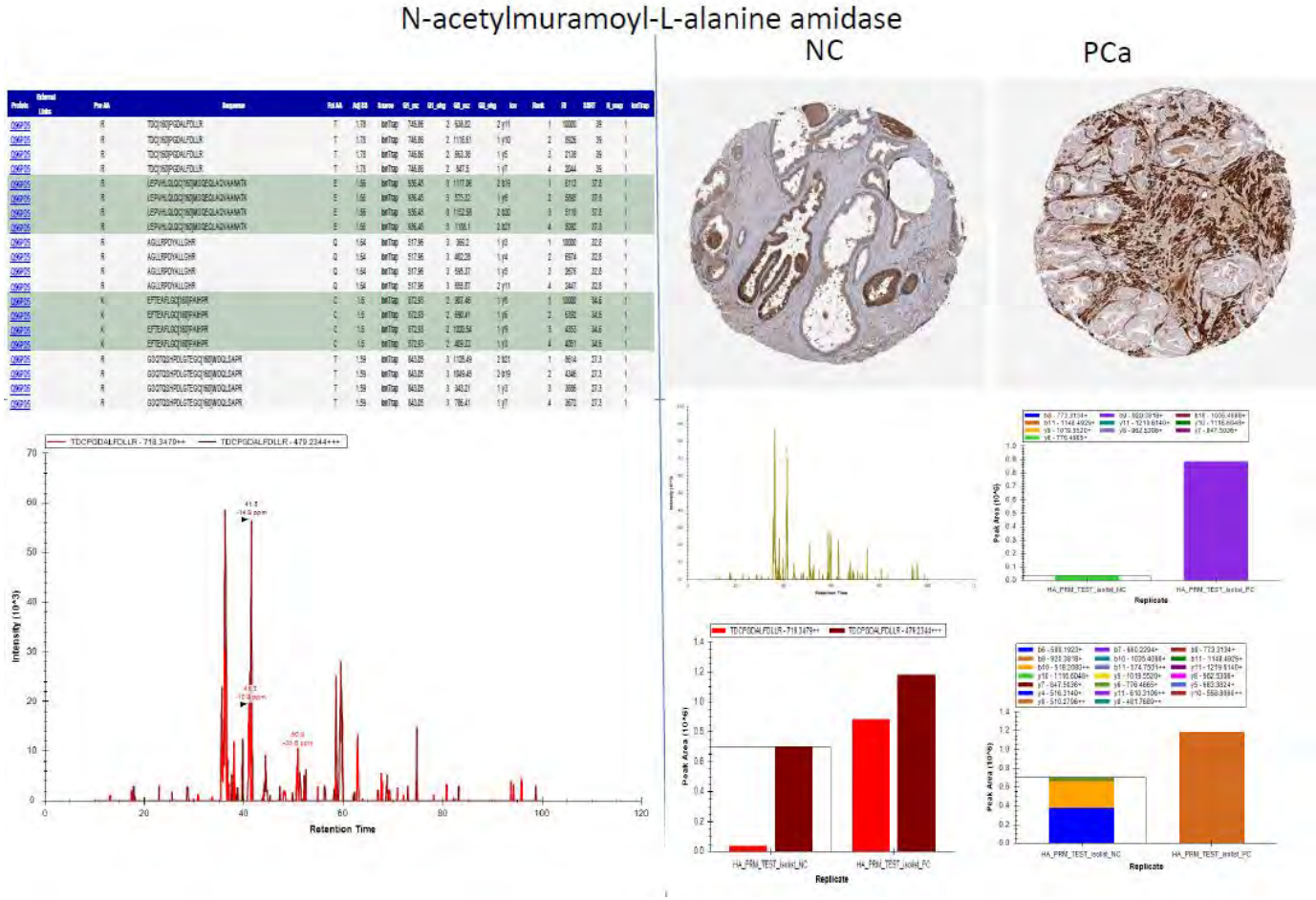
Ubiquitin-conjugating enzyme E2 variant 1;Ubiquitin-conjugating enzyme E2 variant 2	TMEM189;UBE2V2;UBE2V1	6.1679E-31
Ubiquitin-like modifier-activating enzyme 1	UBA1	2.0955E-07
Ubiquitin-like modifier-activating enzyme ATG7	ATG7	5.1173E-14
UDP-glucose:glycoprotein glucosyltransferase 1	UGGT1	8.9865E-09
Uncharacterized protein C16orf46	C16orf46	0.00075022
Uncharacterized protein C4orf21;Uncharacterized protein FLJ44066	C4orf21	0.000037782
Uncharacterized protein C6orf72	C6orf72	4.0187E-20
Uncharacterized protein C9orf142	C9orf142	0.0014118
Unconventional myosin-If	MYO1F	6.7456E-10
UPF0510 protein INM02	C19orf63	0.000057992
UPF0533 protein C5orf44	C5orf44	0.0011151
UPF0539 protein C7orf59	C7orf59	5.2269E-06
UPF0556 protein C19orf10	C19orf10	0.00025337
UPF0587 protein C1orf123	C1orf123	0
UPF0669 protein C6orf120	C6orf120	0.000015806
UPF0687 protein C20orf27	C20orf27	2.267E-21
UPF0696 protein C11orf68	C11orf68	2.9915E-53
Urokinase plasminogen activator surface receptor	PLAUR	0.0014893
Urokinase-type plasminogen activator;Urokinase-type plasminogen activator long chain A;Urokinase-type plasminogen activator short chain A;Urokinase-type plasminogen activator chain B	PLAU	0.001607
Uromodulin;Uromodulin, secreted form	UMOD	0.000021042
Uteroglobin	SCGB1A1	0.00018236
Utrophin	UTRN	0.001713
UV excision repair protein RAD23 homolog A;UV excision repair protein RAD23 homolog B	RAD23A;RAD23B	2.256E-07
UV excision repair protein RAD23 homolog B	RAD23B	1.7825E-201
Vacuolar fusion protein CCZ1 homolog B;Vacuolar fusion protein CCZ1 homolog	CCZ1B;CCZ1	0.000021014
Vacuolar protein sorting-associated protein 35	VPS35	1.5809E-27
Vacuolar protein sorting-associated protein 37B	VPS37B	0.00055276
Vacuolar protein sorting-associated protein 4B;Fidgetin-like protein 1;Vacuolar protein	VPS4B;FIGNL1;VPS4A	0

sorting-associated protein 4A		
Vacuolar protein sorting-associated protein VTA1 homolog	VTA1	3.5432E-06
Vascular cell adhesion protein 1	VCAM1	2.4244E-28
Vascular non-inflammatory molecule 2	VNN2	1.1828E-20
Vasodilator-stimulated phosphoprotein	VASP	2.8899E-17
Vasorin	VASN	1.2166E-07
Versican core protein	VCAN	0.000014931
Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	ACADVL	9.5752E-12
Very long-chain specific acyl-CoA dehydrogenase, mitochondrial	ACADVL	3.7423E-220
Vesicle-associated membrane protein 2	VAMP2	0.000052205
Vesicle-associated membrane protein 2	VAMP2	0
Vesicle-associated membrane protein-associated protein A	VAPA	5.8694E-28
Vesicle-fusing ATPase	NSF	1.1173E-52
Vesicle-trafficking protein SEC22b	SEC22B	2.5166E-20
Vesicular integral-membrane protein VIP36	LMAN2	1.805E-120
Villin-like protein	VILL	4.3291E-11
Vimentin	VIM	4.4962E-07
Vinculin	VCL	0.00144
Vitamin D-binding protein	GC	2.9781E-06
Vitamin K-dependent protein S	PROS1	0.0014254
Vitamin K-dependent protein Z	PROZ	9.5592E-07
Vitelline membrane outer layer protein 1 homolog	VMO1	0.00049303
Vitronectin;Vitronectin V65 subunit;Vitronectin V10 subunit;Somatomedin-B	VTN	3.2131E-12
Voltage-dependent L-type calcium channel subunit alpha-1D	CACNA1D	2.9486E-18
Voltage-gated potassium channel subunit beta-2	KCNAB2	0.00070215
von Willebrand factor;von Willebrand antigen 2	VWF	0.0012628
V-set and immunoglobulin domain-containing protein 4	VSIG4	0.0003412
V-type proton ATPase subunit C 1	ATP6V1C1	0.000030071
V-type proton ATPase subunit G 1;V-type proton ATPase subunit G 2	ATP6V1G1;ATP6V1G2; ATP6V1G2-DDX39B	0

V-type proton ATPase subunit S1	ATP6AP1	2.9932E-06
WAP four-disulfide core domain protein 2	WFDC2	1.8753E-06
WAS/WASL-interacting protein family member 1	WIPF1	3.9691E-10
WASH complex subunit FAM21A;WASH complex subunit FAM21C;WASH complex subunit FAM21B	FAM21A;FAM21C;FAM21B	0.00010427
WD repeat-containing protein 1	WDR1	8.447E-147
WD repeat-containing protein 87	WDR87	1.4302E-13
Wiskott-Aldrich syndrome protein	WAS	0.0052673
Wiskott-Aldrich syndrome protein family member 2	WASF2	9.3462E-117
WNT1-inducible-signaling pathway protein 2	WISP2	3.9617E-110
WW domain-binding protein 2	WBP2	4.59E-119
Xaa-Pro dipeptidase	PEPD	1.1864E-10
X-linked retinitis pigmentosa GTPase regulator-interacting protein 1	RPGRIP1	1.1242E-06
X-ray repair cross-complementing protein 5	XRCC5	4.3449E-06
Xylosyltransferase 1	XYLT1	0.0039844
Zinc finger matrin-type protein 5	ZMAT5	4.0316E-15
Zinc finger protein 33B;Zinc finger protein 33A	ZNF33B;ZNF33A	0.00025854
Zinc-alpha-2-glycoprotein	AZGP1	1.6271E-06
Zymogen granule protein 16 homolog B	ZG16B	1.622E-08
Zyxin	ZYX	0.000036578

8.2. Annexure II- Snapshot of 12 top ranking urinary biomarkers of PCa by In-silico and PRM analyses

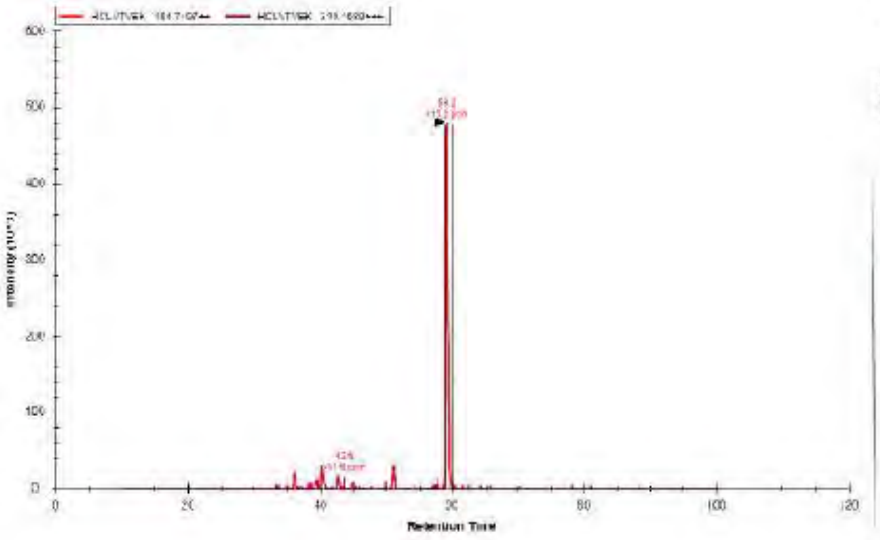
The upper left panels show the transitions previously used in SRMAtlas to identify these peptides; while the upper right panel shows a comparative immunohistochemical staining of these proteins between cancerous and normal prostatic tissues. The lower right left panels shows superimposed PRM traces of +2 and +3 charges of the top ranking potential biomarkers; while the lower right panel shows stacked bar graph replicate comparison of various transitions identified by PRM for both the +2 and +3 charge states of the potential biomarkers.

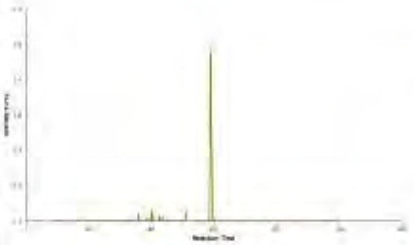


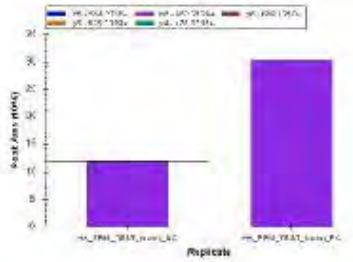
Vitamin K-dependent protein S

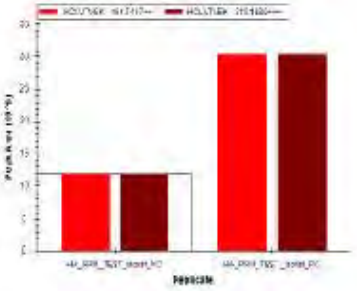
NC

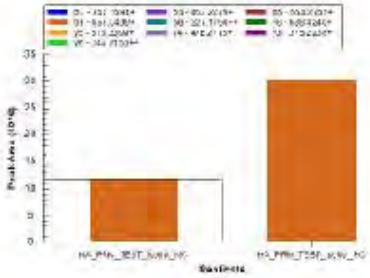
PCa



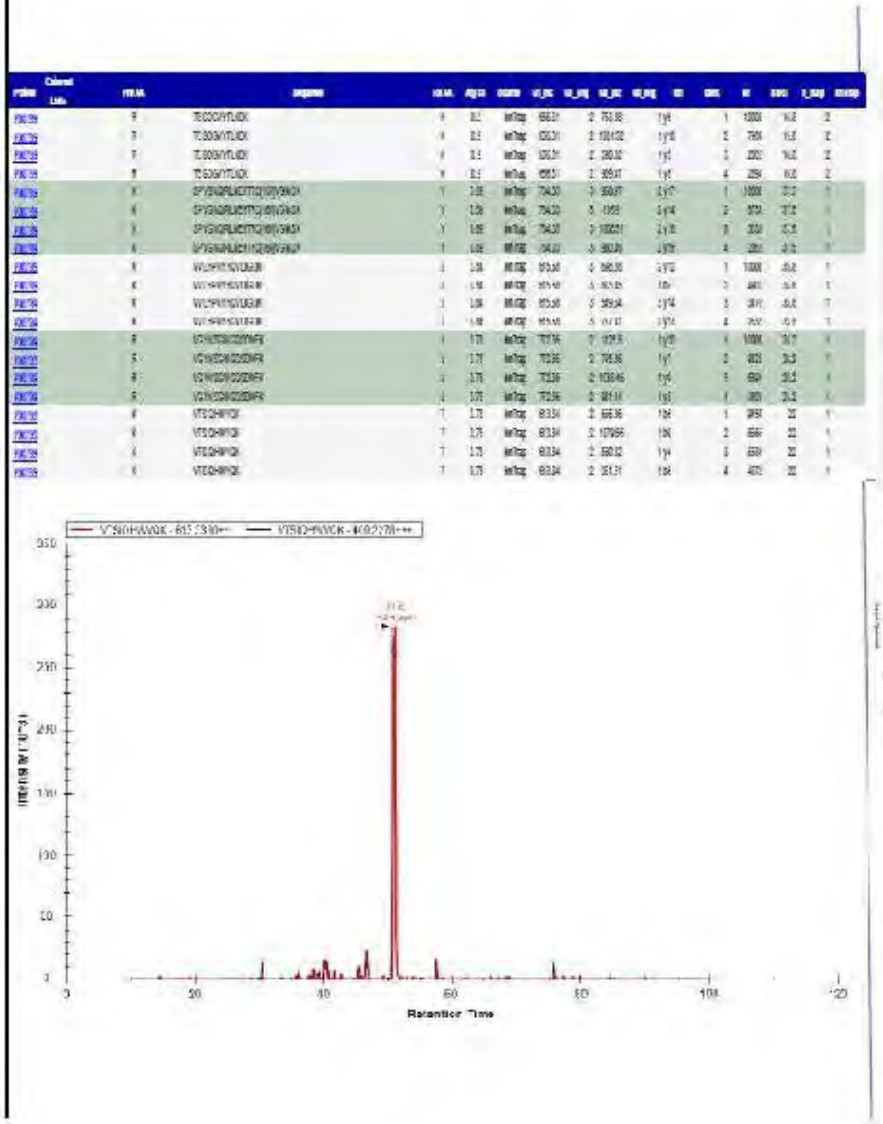


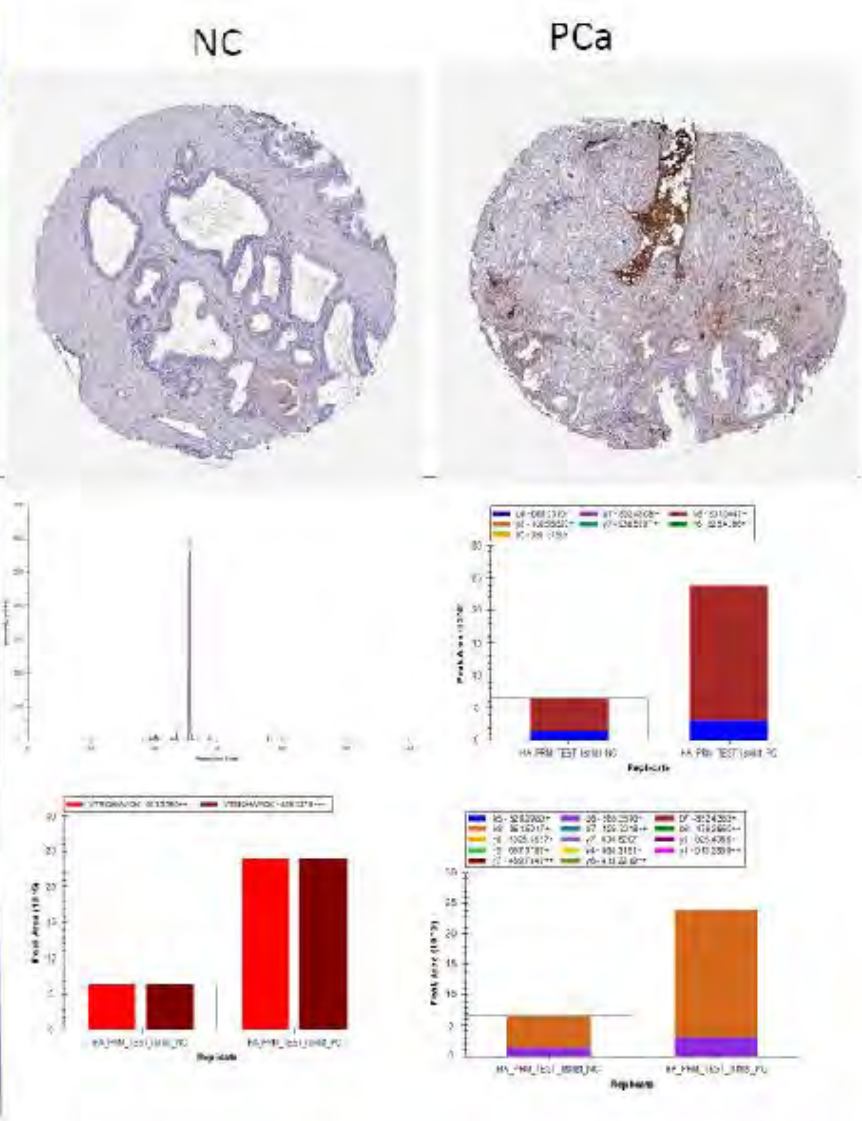






Haptoglobin-related protein

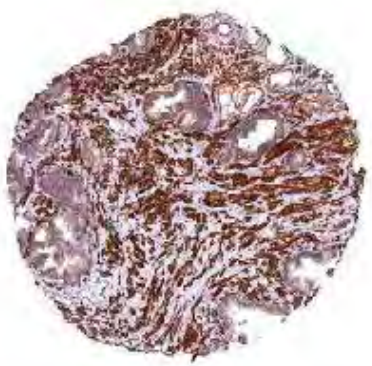


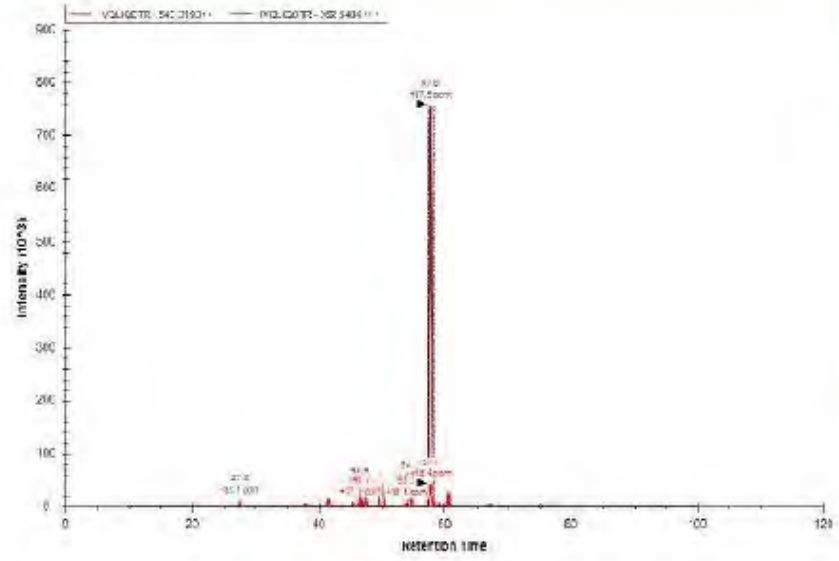


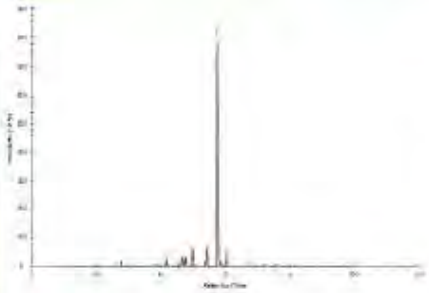
Carboxypeptidase N catalytic chain

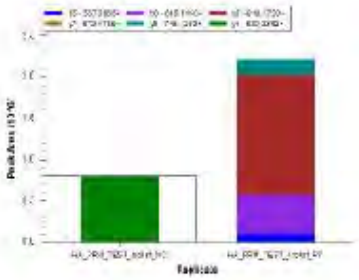
NC

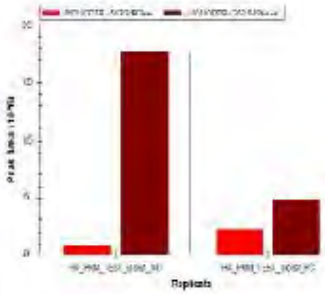
PCa

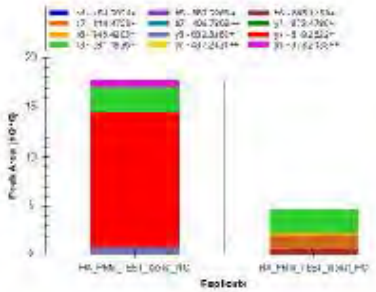




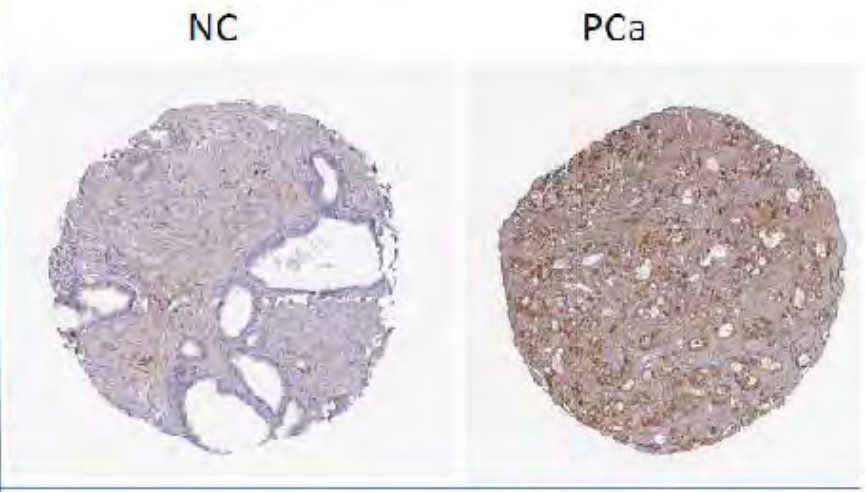


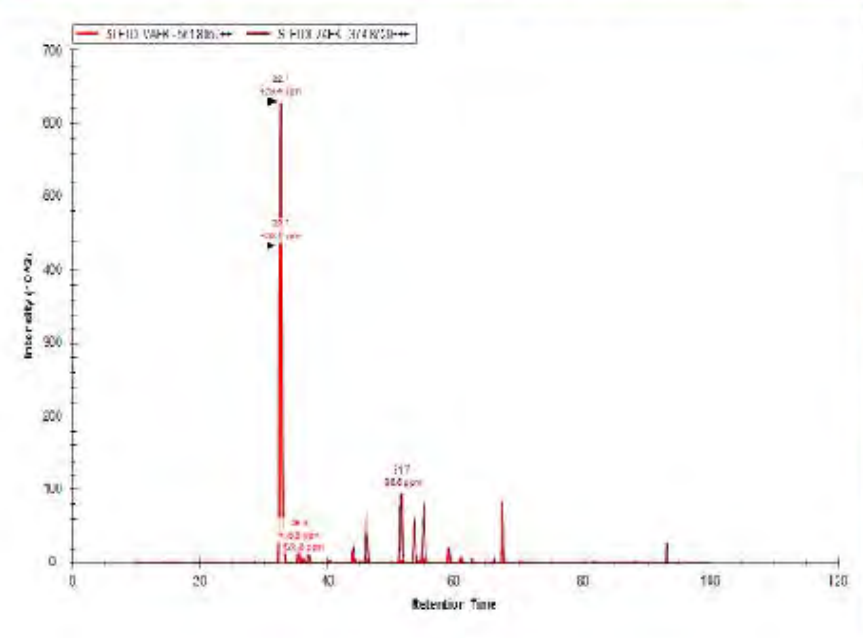


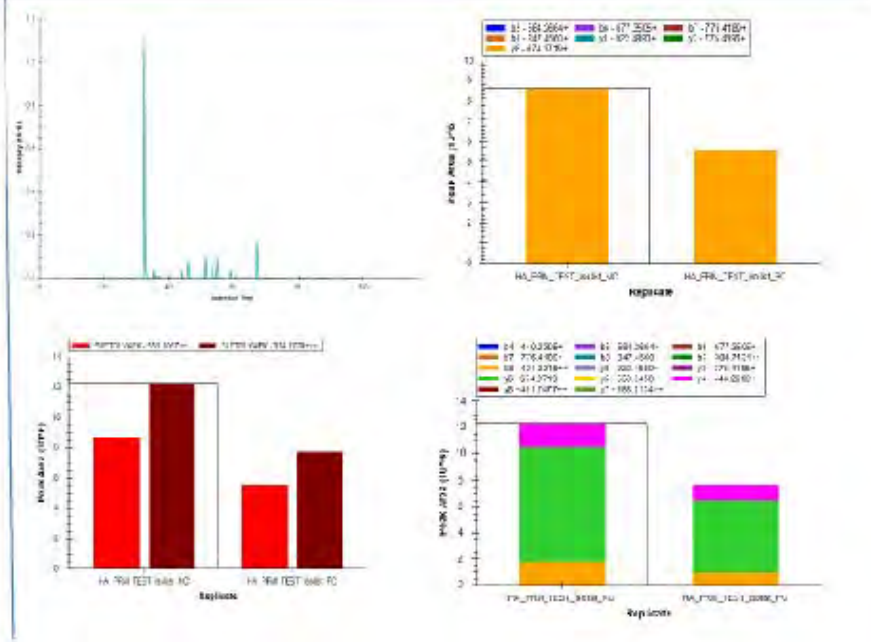




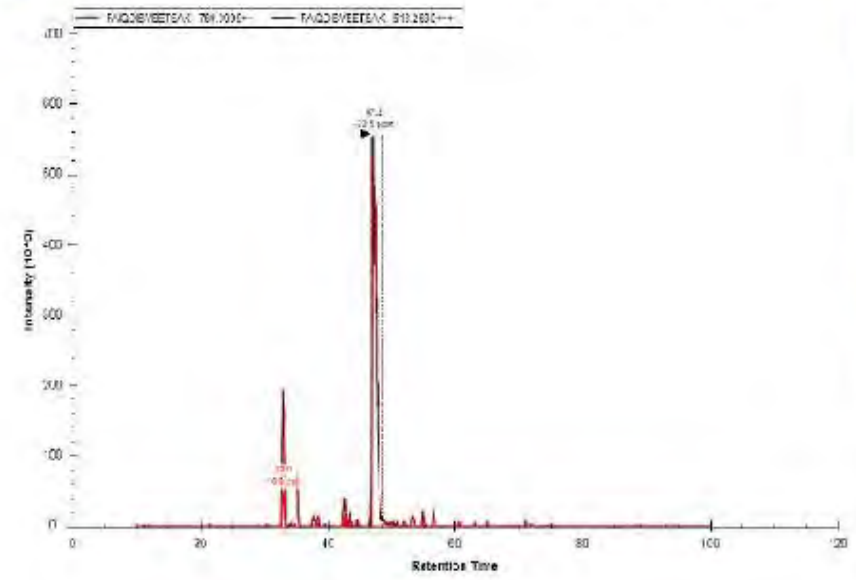
Pregnancy Zone Protein

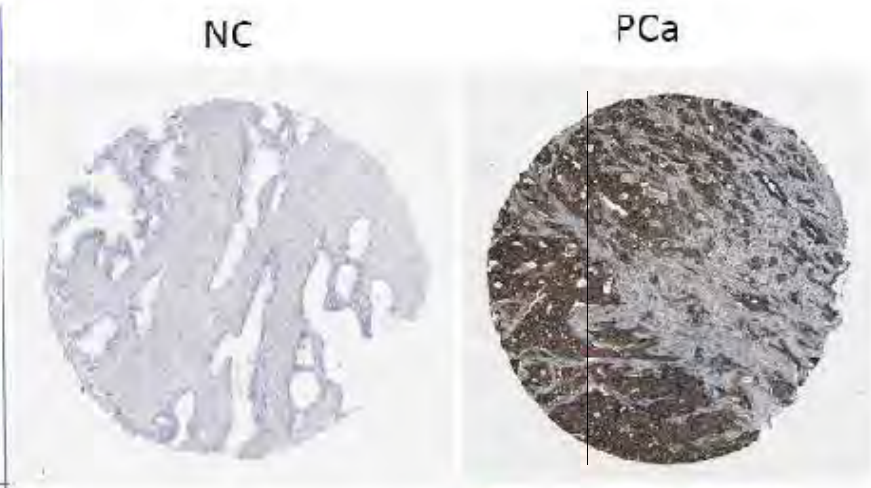


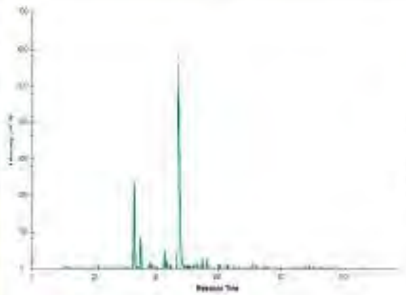


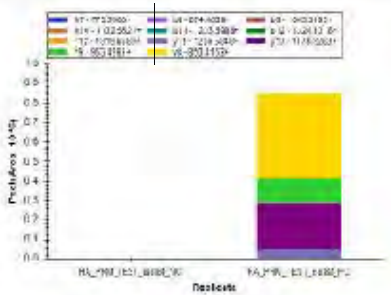


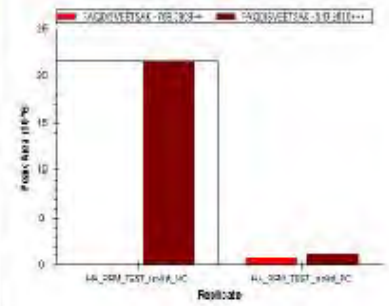
Alpha-actinin-1

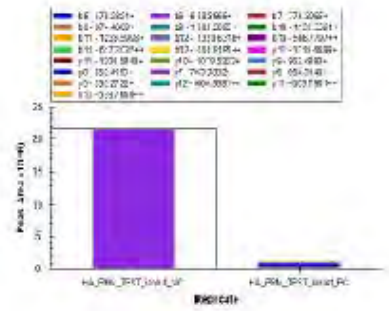






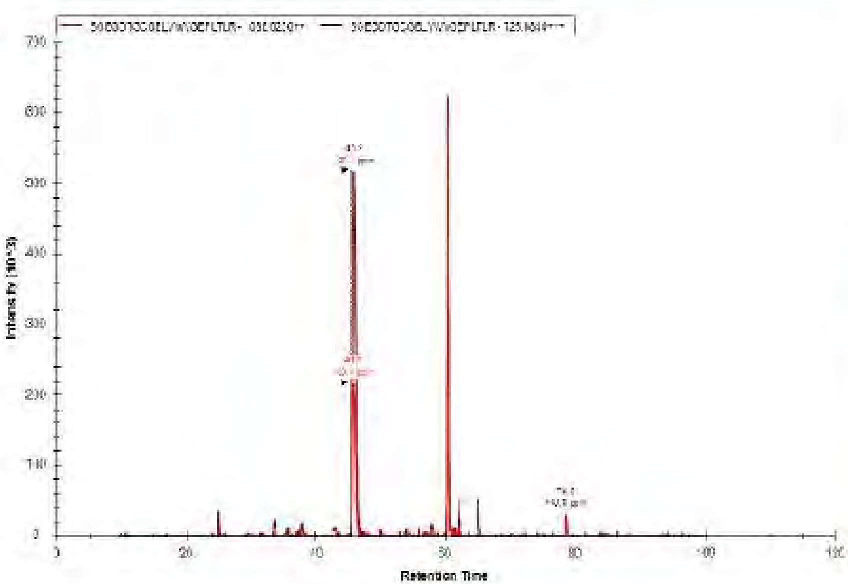




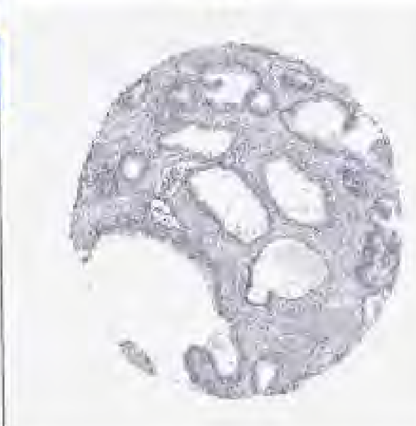


Myocilin

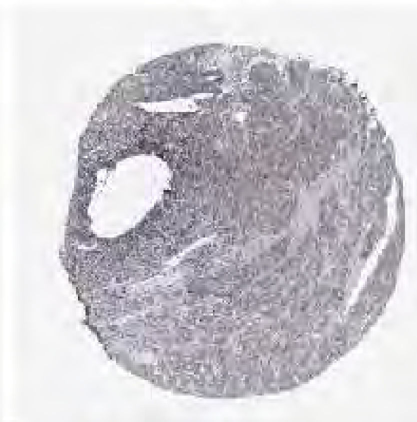
Protein	External Links	PreAA	Sequence	[illegible]	[illegible]	Score	Q1_mz	Q1_chg	Q3_mz	Q3_chg	Ion	[illegible]	R	SSRT	[illegible]	[illegible]
[illegible]		R	[illegible]	T	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	1	[illegible]	[illegible]	1	
[illegible]		R	[illegible]	T	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	1	
[illegible]		R	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	1	
[illegible]		R	[illegible]	T	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	1	

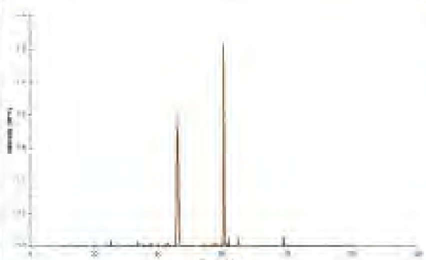


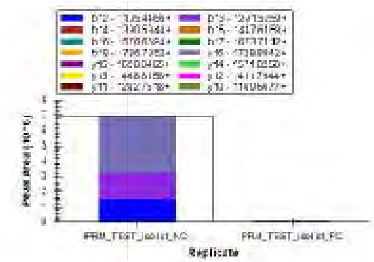
NC

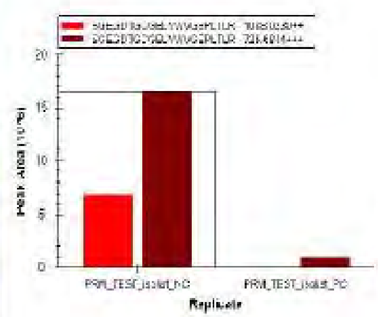


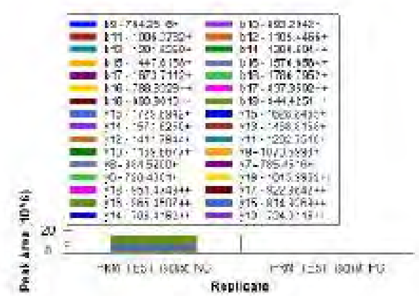
PCa



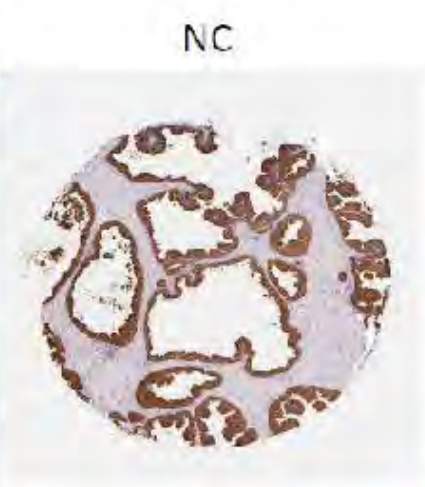


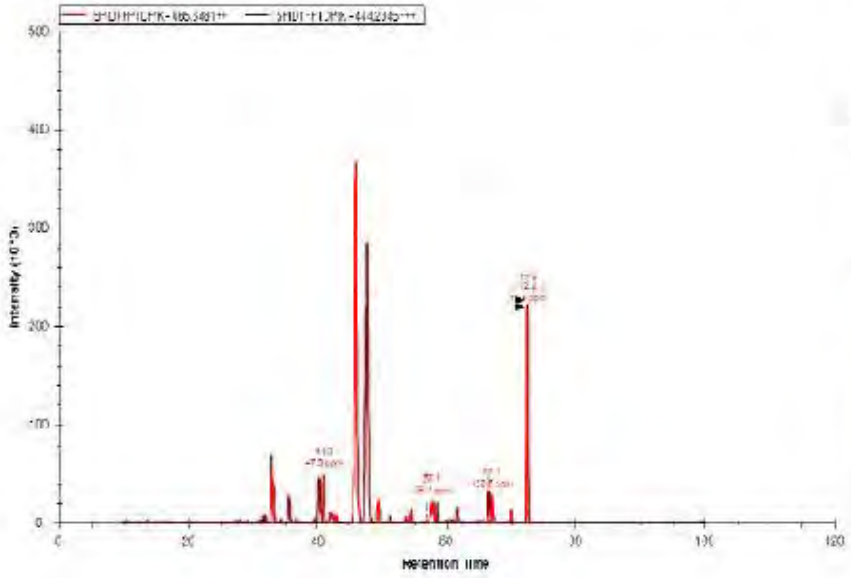


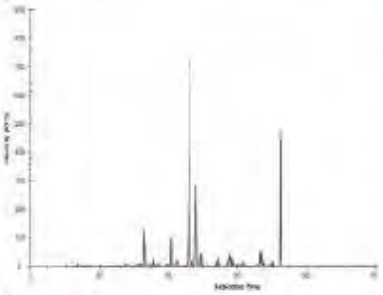


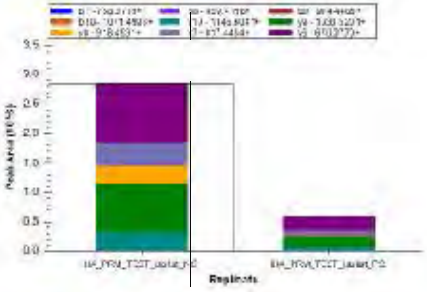


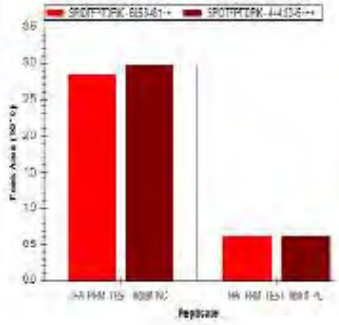
Prostatic acid phosphatase

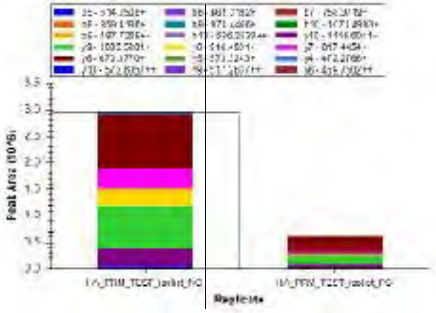




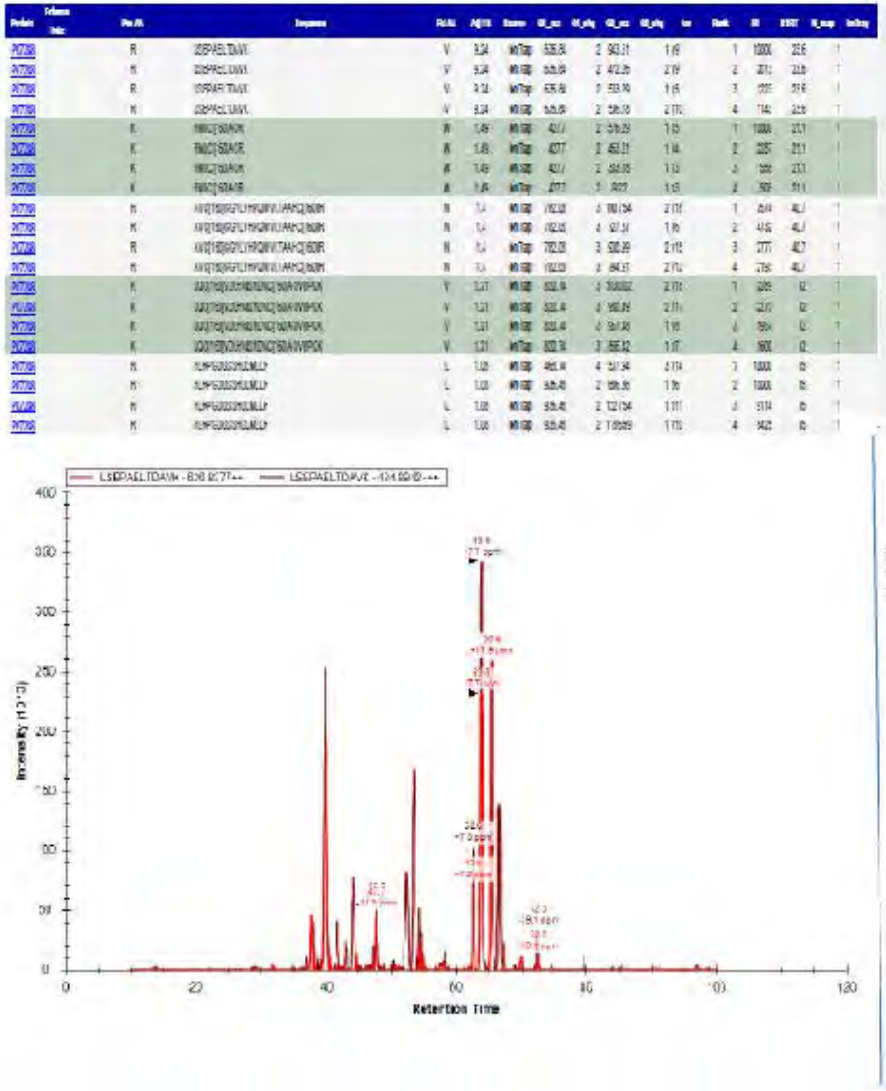


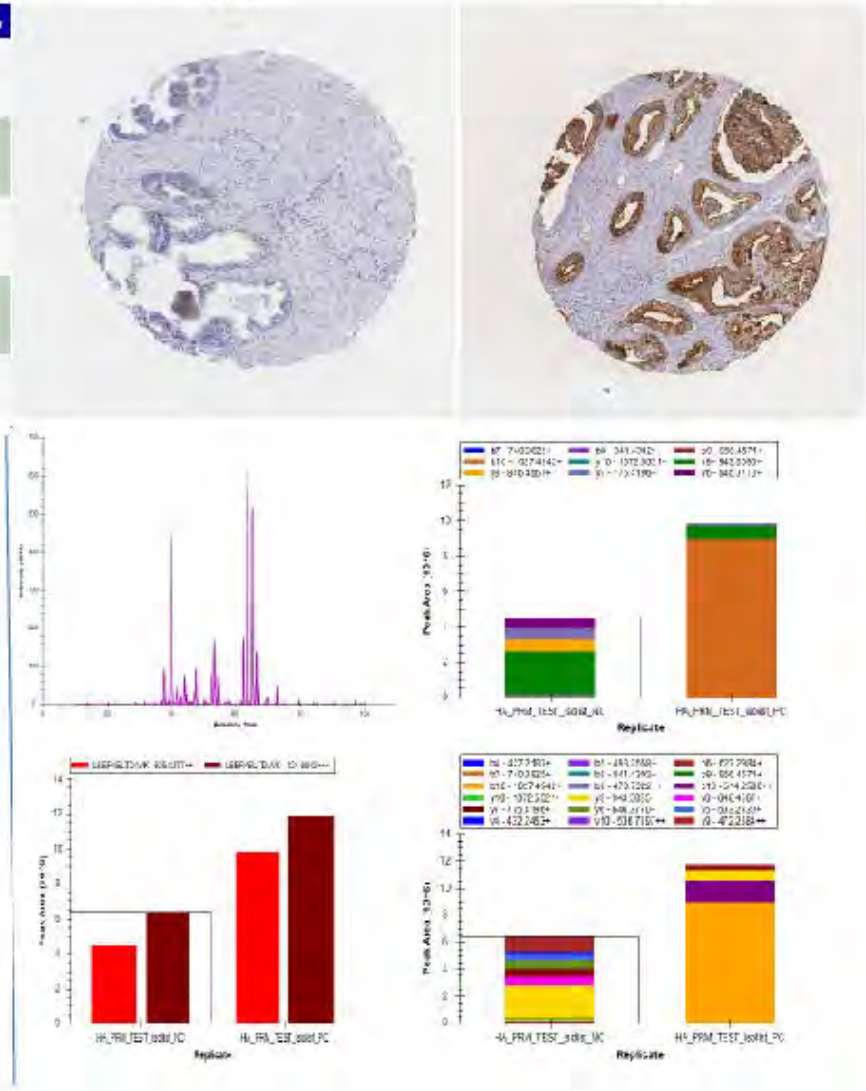




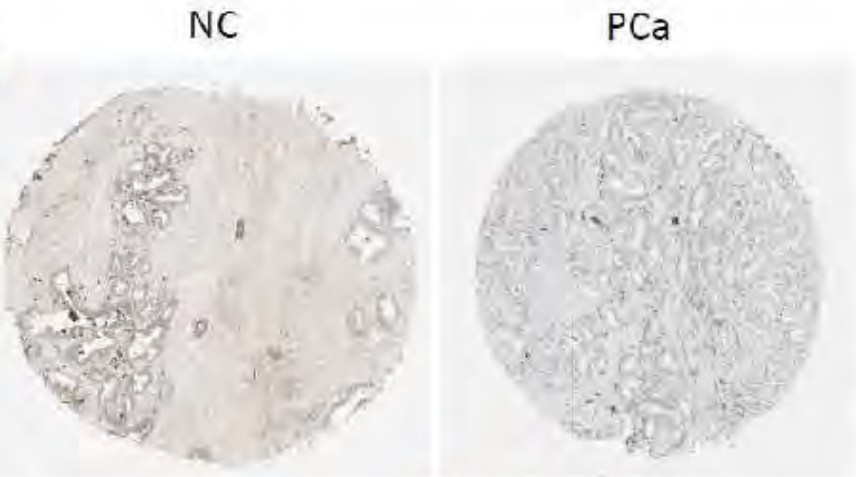


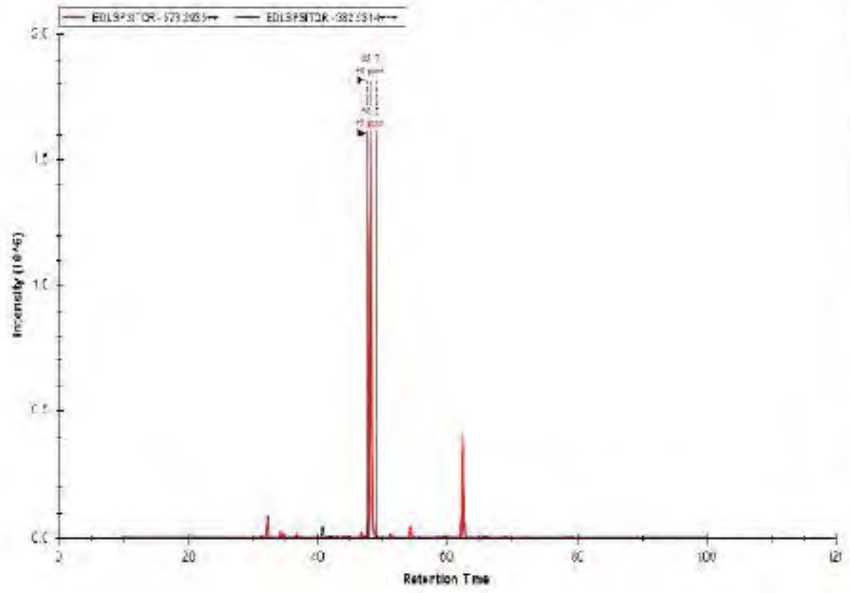
Prostate-specific antigen

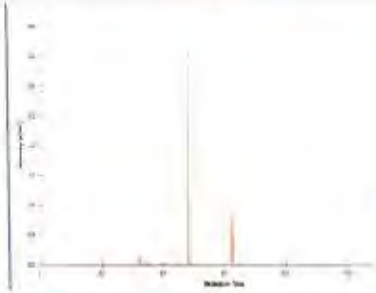


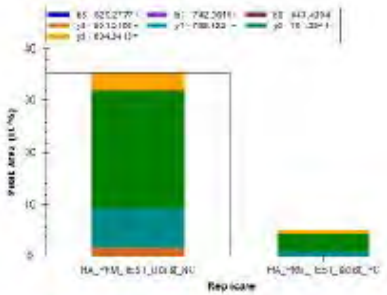


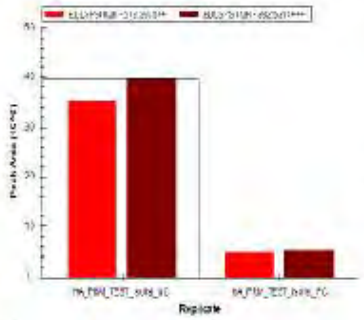
Nidogen-1

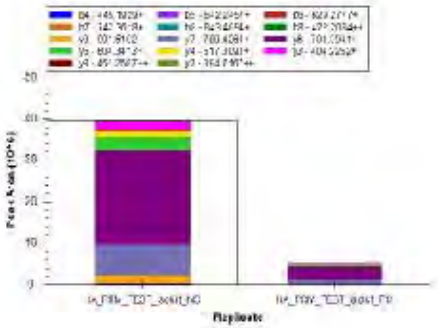

NC
PCa


Intensity (10^6)
Retention Time




Peak Area (10^6)
Replicate

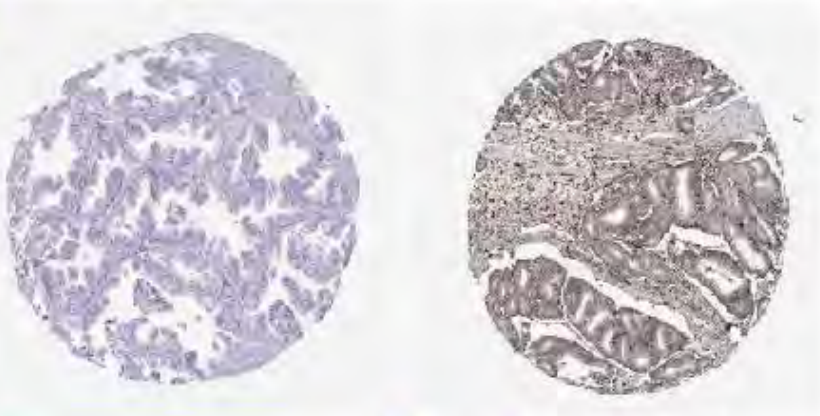

Peak Area (10^6)
Replicate

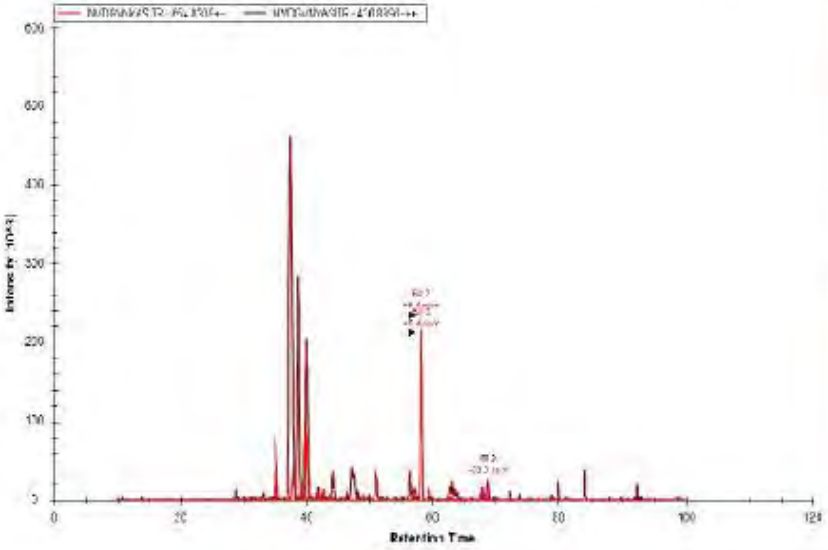

Peak Area (10^6)
Replicate

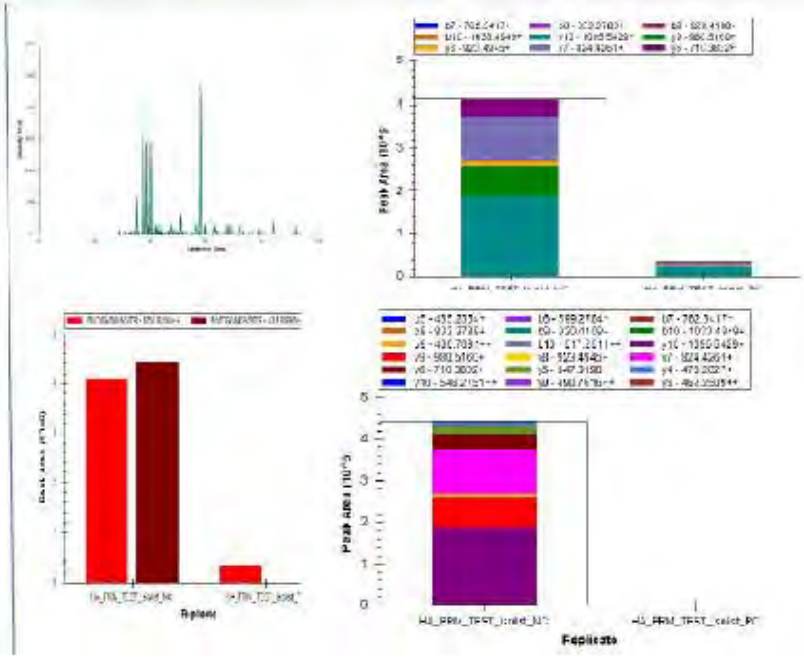
Cathepsin Z

NC

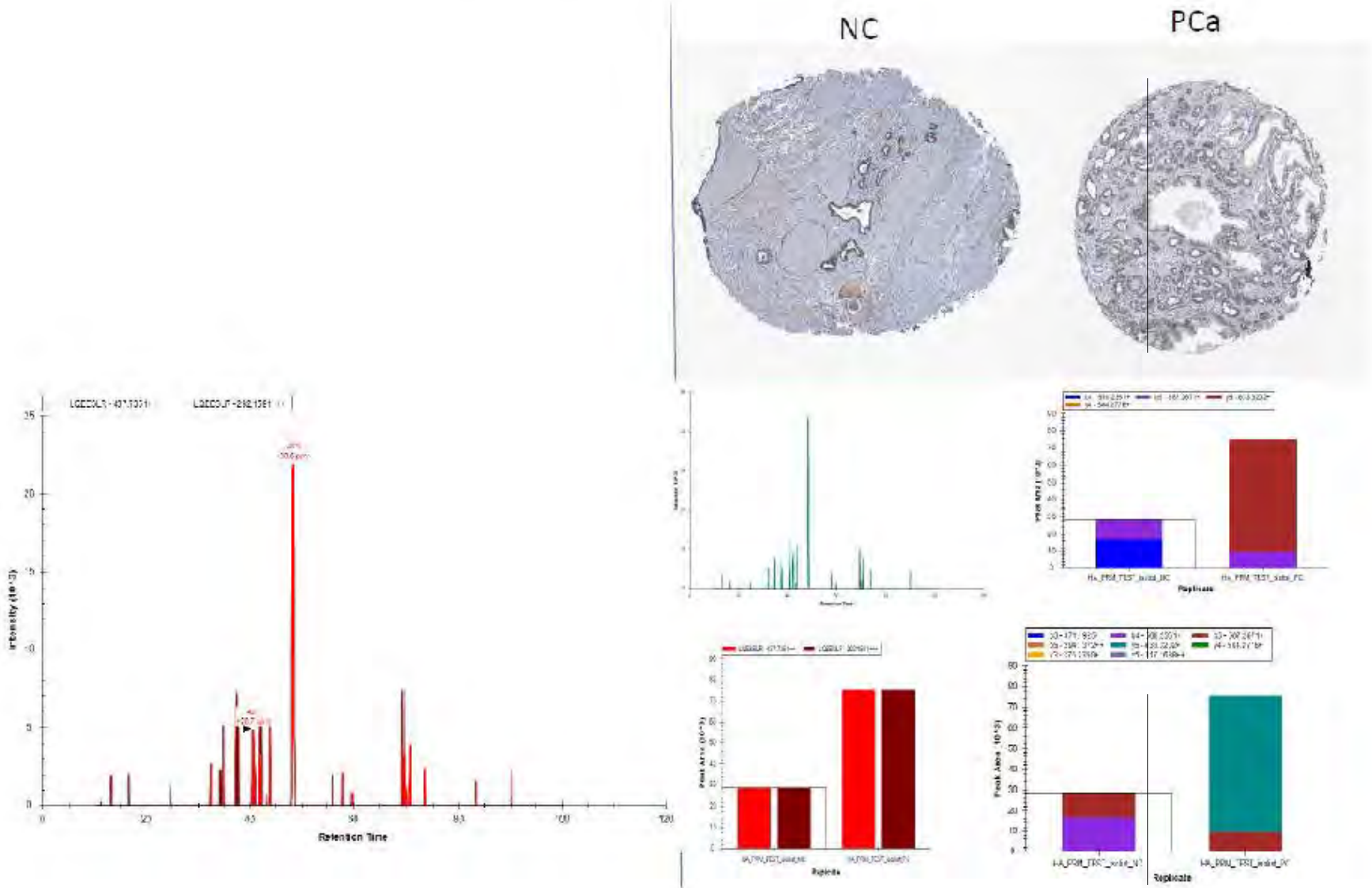
PCa







Slain motif-containing protein 1



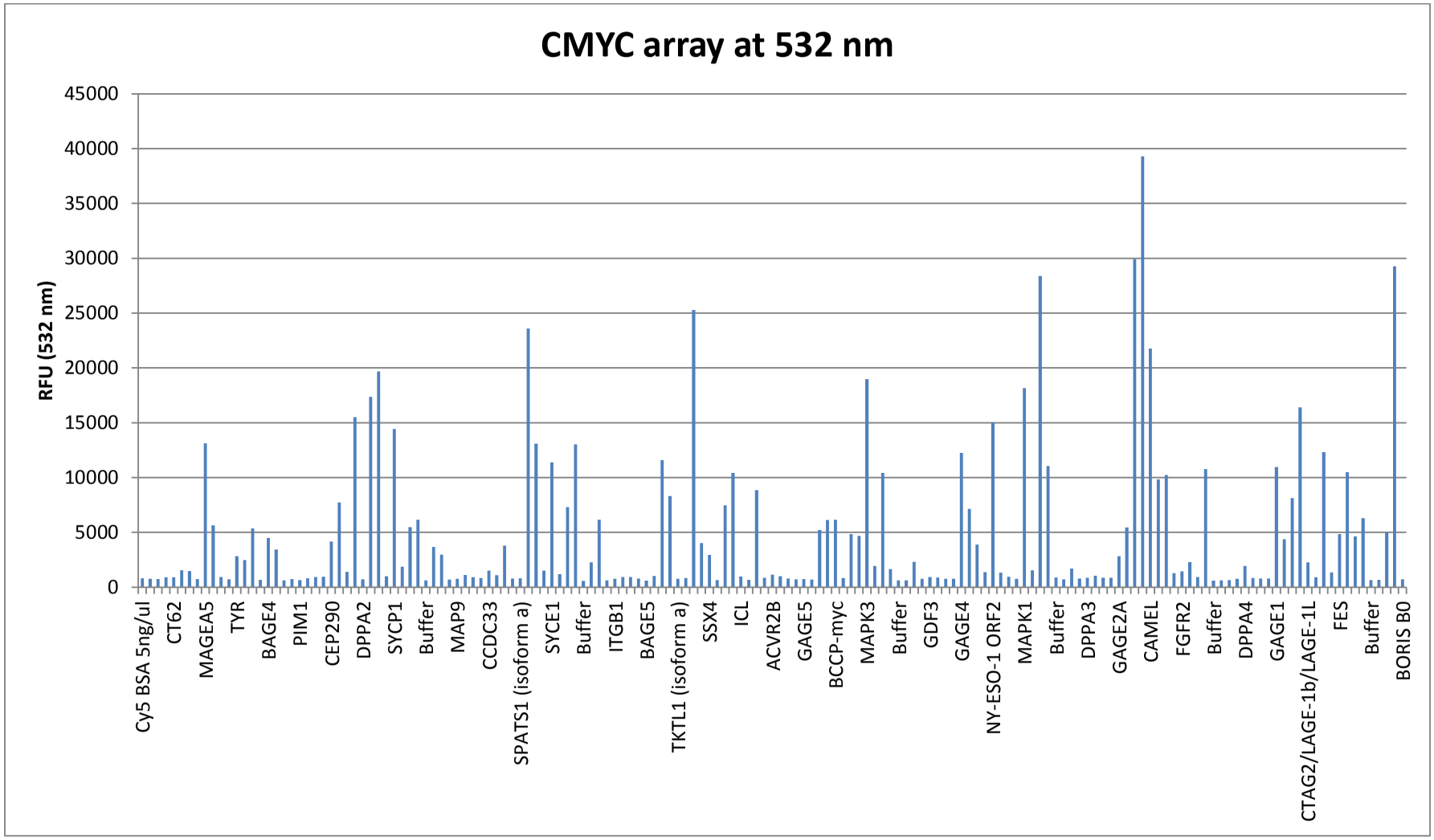
8.3. Annexure III- Default and Minor modification Settings used for MaxQuant analysis

Default Settings	Minor Modifications
Variable Modification= Oxidation (M) and Acetyl (Protein N-term)	Increase the threads to 8
Multiplicity= 2	Tick IBAQ
Enzyme= Trypsin (P)	Tick LFQ
First search PPM= 20	Increase minimum peptide length to 7
Main search PPM=6	Upload a FASTA file
Separate variable modifications for first search= unticked	Upload experimental design template
Maximum number of modifications per peptide=5	
Maximum labeled AAs=3	
Maximum missed cleavage=2	
Maximum charge=7	
Individual peptide mass tolerance= ticked	
Type = standard	
MS/MS tolerance for FTMS= 20	
MS/MS tolerance for ITMS= 0.5	
MS/MS tolerance for TOF= 0.1	
Unknown= 0.5	
Fixed modification= Carbamidomethyl (C)	

Peptide FDR= 0.01	
Site FDR= 0.01	
Max peptide PEP= 1	
Min. peptides= 1	
Min Razor+ Unique peptide= 1	
Min. unique peptide=0	
Protein FDR= 0.01	
Apply site FRD separately= Ticked	
Min. peptide length= 6	
Min. score= 0	
Filter labelled amino acid= Ticked	
Second peptide= Ticked	
Discard unmodified counterpart peptides= Ticked	
Min. ratio count= 2	
Site quantification mode= use least modified peptide”	
Use of occupancies= normalized ratio	
Protein quantification use only unmodified peptides= Ticked	
Requantify= Ticked	
Keep low scoring version of identified peptides=no	
Match between runs= Unticked	

IBAQ= Unticked	
Label free quantification= Unticked	
Thread=1	

8.4. Annexure IV- Bar graph showing anti-c-myc-Cy3 assay which was used to ensure that individual antigens were successfully immobilised to the array during printing.



8.5. Annexure V- Table showing levels of androgen measured in human males across different ethnic backgrounds

This table (adapted from [518]) shows variations in the level of androgen in men from various ethnic backgrouds. These androgen levels were measure in blood samples collected from men (n=1899) enrolled in the BACH study from 2002-2005.

	Black	Hispanic	White	Total	*P* value[d]
Testosterone (ng/dl)[e]					
n	531	648	702	1881	
Mean	454	441	434	440	0.956
Median (IQR)	425 (274)	404 (219)	418 (249)	418 (253)	
Bioavailable testosterone (ng/dl)[e]					
n	531	648	701	1880	
Mean	217	223	210	213	0.313
Median (IQR)	206 (116)	214 (117)	196 (104)	201 (109)	
DHT (ng/dl)[e]					
n	529	648	701	1878	
Mean	47.8	43.6	44.4	45.2	0.121
Median (IQR)	36.9 (25.4)	35.3 (20.0)	35.1 (19.7)	35.4 (20.9)	
DHT to testosterone (DHT:T) ratio[e]					
n	528	648	701	1877	
Mean	0.1083	0.0996	0.1045	0.1048	0.051
Median (IQR)	0.0969 (0.0580)	0.0881 (0.0494)	0.0877 (0.0520)	0.0903 (0.0529)	
SHBG (nmol/liter)					
n	531	648	701	1880	
Mean	36.0	32.0	34.2	34.3	0.128
Median (IQR)	30.6 (20.2)	27.5 (17.9)	30.3 (19.5)	30.3 (19.2)	
DHEAS (μg/ml)[f]					
n	531	648	701	1880	
Mean	1.93	2.10	2.04	2.02	0.027
Median (IQR)	1.80 (1.67)	1.97 (1.38)	1.90 (1.44)	1.88 (1.50)	

SHBG=Sex hormone-binding globulin, DHEAS=Dehydroepiandrosterone sulfate, n=Number, IQR=Interquartile range, DHT=Dihydrotestosterone.

8.6. Annexure VI- Comparison of MS1 and MS2 Quantitation

Ions formed in tandem spectrometry are separated by their mass-to-charge ratio (*m/z*) in the initial stage of mass spectrometry as shown below (**MS1**). Ions of a particular *m/z* which are known as precursor ion are further selected for fragmentation and daughter/fragment/product ions are created by higher collision energy dissociation (HCD), collision-induced dissociation (CID), photodissociation, ion-molecule reaction, or other fragmentation process. The ions generated in this manner are then separated and detected in the next stage of mass spectrometry (**MS2**). (Picture adapted from http://proteomicsnews.blogspot.co.za/)