Proteomic Identification and Systematic Verification of Biomarkers for Aggressive Prostate Cancer

by

Yunee Kim

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Medical Biophysics
University of Toronto

© Copyright by Yunee Kim 2015
Proteomic Identification and Systematic Verification of Biomarkers for Aggressive Prostate Cancer

Yunee Kim

Doctor of Philosophy
Medical Biophysics
University of Toronto

2015

Abstract

Despite prostate cancer’s status as the most common cancer to affect men in the Western world, only an estimated 20% of men will die from it. Due to the lack of effective means of prognostication, many men are unnecessarily biopsied and treated. The early and non-invasive identification of lethal disease would have a profound impact on the current practice of prostate cancer management. To this end, the work presented in this thesis describes efforts to characterize proteins in a prostatic fluid known as expressed prostatic secretions (EPS), and to delineate factors in this fluid that demonstrate a high potential as prognostic biomarkers. Using different mass spectrometry-based platforms, I present both a global view of the EPS proteome, as well as targeted, quantitative measures of promising biomarker candidates in this fluid.

In order to generate an archive of soluble and secreted factors present in the proteome of EPS, I used shotgun proteomics to analyze direct-EPS and EPS-urine. This led to the identification of over 2000 proteins that comprise the EPS proteome. The data was integrated with other publicly accessible information, such as gene expression profiles and immunohistochemistry images, to characterize this proteome. Furthermore, a collection of prostate-enriched proteins was identified...
by comparing the proteomes of EPS-urine and urine, demonstrating the unique protein profile of EPS.

Next, comparative proteomic profiling of direct-EPS from organ confined and extracapsular tumors produced a long list of potential biomarker candidates. A targeted assay was systematically developed in order to quantitatively assess the performance of the most promising candidates in cohorts of EPS-urines from a heterogeneous population of prostate cancer patients and controls. Utilizing the assay, a number of candidates were verified to be putative biomarkers for prostate cancer diagnosis and prognosis. This information needs to be further validated and holds promise for improved prostate cancer management.
Acknowledgments

Let me begin with my deepest gratitude for the support and encouragement I received from my supervisor, Dr. Thomas Kislinger. His integrity, courage, and commitment to contribute high caliber research to advancing medical science have been great driving forces in my own pursuit towards a productive Ph.D.

The members of my advisory committee, Drs. Theodorus Van der Kwast and Rama Khokha, have been instrumental in my journey. I thank them for taking the time to engage in stimulating discussions and offering their valuable insights.

This thesis is a culmination of the dedication of numerous individuals. I thank Drs. Richard Drake, John Semmes, and Raymond Lance for organizing and providing the samples. Thanks to Dr. Anthony Gramolini for permitting the unlimited use of the mass spectrometer. Thank you to Drs. Paul Boutros, Clare Jeon, and Irina Kalatskaya, and Miss Cindy Yao for performing the bioinformatics analyses and patiently answering my often naïve questions. Lastly, I thank the patients and donors. None of this would have been possible without their generosity.

My time in grad school was made enjoyable in large part due to the many friends and groups that became a part of my life. I have been fortunate to work alongside the past and present members of the Kislinger lab: Ankit Sinha, Dr. Salvador Mejia, Alexandr Ignatchenko, Vladimir Ignatchenko, Javier Alfaro, Dr. Erwin Schoof, Dr. Lusia Sepiashvili, Dr. Sarah Elschenbroich, Lyudmila Nedyalkova, Dr. Simona Fontana and Dr. Simona Principe. I admire each and every one of them for their bright minds and am grateful for the lasting friendships. A special shout out to Sarah for patiently teaching me the way around a mass spectrometer and guiding me through my first attempts at analyzing the data. The memories of the happy hours we spent together at Café Volo after a hard day’s work will be forever cherished. I’m grateful to my lab mate and friend Simona, with whom I shared countless lunches, washroom breaks, and random adventures. I admire her tremendous abilities as a scientist and enormous warm heart. Thank you to Lusia, for always having my back and taking the time to share your brilliant ideas with me. In addition to my lab mates, I thank the Rottapel, Raught, and Gramolini labs for making the work environment so pleasurable. The countless hours spent in the windowless mass spec room were made enjoyable by the many friends that kept me company.
I am grateful to my best friends, Serge and Laura, for being my supporters and therapists, even from halfway around the world. I’m thankful to my #yologirls for being those with whom I can be downright silly. Thanks to my MPC friends for their openness, kindness, generosity, and deep, thought-provoking insights. They have inspired an exploration of myself to a depth and truth I have never experienced before.

I thank my family for giving me the best foundation to grow from and showering me with their warm love every day. With their strong support, I am fearless about whatever life throws my way.

Last but not least, I thank my partner, Albert, for his quiet patience and unwavering love.
Summary of Accomplishments

Publications (10)

Original research articles


Reviews, book chapter, commentary


**Presentations to Scientific Audiences (12)**

*Oral presentations*

4. Lake Louise Tandem Mass Spectrometry Workshop, 2011. Lake Louise, AB, Canada
5. James Lepock Memorial Symposium Ph.D. Talk, 2011. Toronto, ON, Canada
6. Ontario Cancer Institute Cancer Proteomics Group Meeting, 2010. Toronto, ON, Canada

*Poster presentations*

10. Institute of Medical Science Day, 2010. Toronto, ON, Canada
11. 22nd Annual Canadian Society for Immunology Conference, 2009. Whistler, BC, Canada

12. Centre for Genetic Improvement of Livestock Workshop, 2009. Guelph, ON, Canada

Abstracts (7)


Academic Awards, Scholarships, Fellowships, and Honours (20)

1. 2014 Paul Starita Graduate Student Fellowships

2. 2014 Peterborough K.M. Hunter Graduate Studentships
3. 2014 Frank Fletcher Memorial Fund
4. 2014 Canadian National Proteomics Network Travel Award
5. 2014 Canadian National Proteomics Network Best Oral Presentation by a Trainee
6. 2013 Ontario Graduate Scholarship
7. 2013 Scace Graduate Fellowship in Prostate Cancer Research
8. 2013 Paul Starita Graduate Student Fellowship
9. 2013 Medical Biophysics Travel Grant
10. 2012 Ontario Graduate Scholarship
11. 2012 Medical Biophysics Excellence Fellowship
12. 2011 James Lepock Memorial Symposium Best PhD Talk Award
13. 2011 Ontario Graduate Scholarship
14. 2011 Canadian Society for Mass Spectrometry Travel Award
15. 2010 Ontario Graduate Scholarship
16. 2010 Paul Starita Graduate Student Fellowship
17. 2010 Australian-Canadian Prostate Cancer Research Alliance Travel Award
18. 2010 Prostate Cancer Foundation of Australia Poster Contest
19. 2010 Paul Starita Graduate Student Fellowship
20. 2009 Institute of Medical Science Entry Award
# Table of Contents

Acknowledgments........................................................................................................ iv

Summary of Accomplishments...................................................................................... vi

Table of Contents........................................................................................................... x

List of Tables .................................................................................................................. xv

List of Figures ................................................................................................................ xvi

List of Appendices ......................................................................................................... xix

List of Abbreviations .................................................................................................... xxi

Chapter 1 Introduction and thesis overview ................................................................. 1

1 Introduction and thesis overview .............................................................................. 1

1.1 The prostate gland .................................................................................................. 1

1.1.1 Anatomy and histology of the prostate gland ..................................................... 1

1.2 Prostate cancer ........................................................................................................ 3

1.2.1 Prostate cancer epidemiology and statistics ..................................................... 3

1.2.2 Histopathology and clinical progression of prostate cancer ......................... 5

1.2.3 Current concepts of insignificant versus significant prostate cancer .......... 6

1.2.4 Genetic alterations and molecular subtypes of prostate cancer .................... 7

1.2.5 Current methods of prostate cancer and diagnosis ...................................... 8

1.3 Prostate cancer biomarkers ................................................................................... 11

1.3.1 Introduction to cancer biomarkers .................................................................. 11

1.3.2 Clinical and analytical considerations for cancer biomarker assay development. 12

1.3.3 Biological sources of biomarkers ................................................................... 15

1.3.3.1 Tissue proximal fluids .............................................................................. 16

1.3.3.2 Biomarkers in biological fluids ................................................................. 17

1.3.4 Adjunct prostate cancer biomarkers ................................................................. 18
1.3.4.1 Cellular markers .................................................................................. 19
1.3.4.2 Genetic aberrations as prognostic biomarkers................................. 20
1.3.4.3 MicroRNAs ....................................................................................... 21
1.3.4.4 Circulating biomarkers ..................................................................... 21
1.3.4.5 Urinary biomarkers .......................................................................... 22
1.4 Clinical proteomics for biomarker discovery ........................................... 23
  1.4.1 Discovery proteomics using shotgun, bottom-up proteomics ............... 23
  1.4.2 Protein quantification using mass spectrometry ................................. 27
  1.4.3 Targeted proteomics .......................................................................... 29
    1.4.3.1 Selected reaction monitoring mass spectrometry ......................... 30
    1.4.3.2 Application of selected reaction monitoring mass spectrometry to biomarker research .......................................................... 31
1.5 Rationale and objectives ......................................................................... 33
  1.5.1 Rationale ........................................................................................... 33
  1.5.2 Hypotheses and aims ......................................................................... 33
Chapter 2 In-depth proteomic analyses of prostate proximal tissue fluids .......... 36
  2 In-depth proteomic analyses of prostate proximal fluids .......................... 36
    2.1 Abstract .............................................................................................. 36
    2.2 Introduction ......................................................................................... 37
    2.3 Results ............................................................................................... 39
      2.3.1 In-depth proteomic analyses of direct expressed prostatic secretions .... 39
        2.3.1.1 MudPIT profiling of direct expressed prostatic secretions from prostate cancer patients ...................................................... 39
      2.3.2 Integrative data mining and comparison to other public databases .... 43
        2.3.2.1 Comparison of the direct-EPS proteome to other published proteomic datasets .................................................................... 45
        2.3.2.2 Multi-dataset integration .............................................................. 46
2.3.3 Identification of prostate-enriched proteins by in-depth proteomic analyses of expressed prostatic secretions in urine ................................................................. 50
  2.3.3.1 MudPIT profiling of expressed prostatic secretions in urine .......... 50
  2.3.3.2 Identification of prostate-enriched proteins .................................... 57
  2.3.3.3 Characterization of prostate-enriched proteins .................................. 62
  2.3.3.4 Generation of candidate shortlist ...................................................... 64
  2.3.3.5 Verification of proteomic data .......................................................... 64

2.4 Discussion ................................................................................................. 66

2.5 Materials and methods ............................................................................. 67
  2.5.1 Materials ............................................................................................... 67
  2.5.2 Sample collection and properties .......................................................... 67
  2.5.3 Protein digestion and sample preparation .............................................. 69
  2.5.4 MudPIT analyses ................................................................................ 69
  2.5.5 Protein identification and data analyses ................................................. 71
    2.5.5.1 Gene Ontology annotation and data comparisons ......................... 72
  2.5.6 Prostate-enriched proteins characterization ......................................... 73
    2.5.6.1 SDS-PAGE and Western blot analysis on urine and EPS-urine pools... 74

Chapter 3 Identification of candidate biomarkers of aggressive prostate cancer .......... 75

3 Identification of candidate biomarkers of aggressive prostate cancer ...................... 75
  3.1 Abstract .................................................................................................. 75
  3.2 Introduction ............................................................................................ 76
  3.3 Results .................................................................................................... 78
    3.3.1 Proteomic characterization of expressed prostatic secretions from organ-confined and extracapsular prostate cancers ........................................... 78
    3.3.2 Semi-quantitative comparison of expressed prostatic secretions from extracapsular and organ-confined prostate cancers ........................................ 84
    3.3.3 Generation of candidate shortlist ......................................................... 87
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.4</td>
<td>Verification of differentially expressed candidates in direct-EPS</td>
<td>90</td>
</tr>
<tr>
<td>3.3.5</td>
<td>Verification of differentially expressed candidates in EPS-urines</td>
<td>93</td>
</tr>
<tr>
<td>3.4</td>
<td>Discussion</td>
<td>99</td>
</tr>
<tr>
<td>3.5</td>
<td>Materials and methods</td>
<td>103</td>
</tr>
<tr>
<td>3.5.1</td>
<td>EPS Sample Properties</td>
<td>103</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Protein digestion and sample preparation for mass spectrometry</td>
<td>104</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Protein identification and data analysis</td>
<td>104</td>
</tr>
<tr>
<td>3.5.4</td>
<td>Gene Ontology annotation, pathway enrichment and network-based analysis of EPS proteins</td>
<td>105</td>
</tr>
<tr>
<td>3.5.5</td>
<td>Integrative data mining</td>
<td>106</td>
</tr>
<tr>
<td>3.5.6</td>
<td>Immunoassays in prostatic fluids</td>
<td>106</td>
</tr>
<tr>
<td>3.5.7</td>
<td>Statistical analysis</td>
<td>107</td>
</tr>
<tr>
<td>3.5.8</td>
<td>SRM-MS development</td>
<td>107</td>
</tr>
<tr>
<td>3.5.9</td>
<td>Relative quantification by SRM-MS in EPS-urines</td>
<td>109</td>
</tr>
<tr>
<td>4</td>
<td>Systematic large-scale development of targeted proteomics assays for biomarker verification in prostate-proximal fluids</td>
<td>110</td>
</tr>
<tr>
<td>4.1</td>
<td>Abstract</td>
<td>110</td>
</tr>
<tr>
<td>4.2</td>
<td>Introduction</td>
<td>111</td>
</tr>
<tr>
<td>4.3</td>
<td>Results</td>
<td>112</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Candidate selection for targeted quantification by selected reaction monitoring mass spectrometry</td>
<td>112</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Targeted relative quantification of candidates in a cohort of EPS-urines</td>
<td>114</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Selection of peptide candidates for verification</td>
<td>116</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Absolute quantification for the verification of candidates in EPS-urines</td>
<td>119</td>
</tr>
<tr>
<td>4.3.5</td>
<td>Comparative analysis between subgroups of patients</td>
<td>122</td>
</tr>
<tr>
<td>4.4</td>
<td>Discussion and future direction</td>
<td>127</td>
</tr>
</tbody>
</table>
4.5 Materials and Methods........................................................................................................ 130
  4.5.1 Sample collection and properties ............................................................................. 130
  4.5.2 Sample preparation for mass spectrometry ......................................................... 131
  4.5.3 Target acquisition by selected reaction monitoring mass spectrometry .......... 131
  4.5.4 Phase 2 peptide selection ....................................................................................... 132
  4.5.5 Phase 2 peptide quantification and Phase 3 peptide selection ......................... 132
  4.5.6 Generation of response curves and determination of figures of merit................. 133
  4.5.7 Multiplexed, scheduled selected reaction monitoring mass spectrometry method.......................................................... 133
  4.5.8 Verification of Phase 3 peptides in EPS-urines ..................................................... 133
  4.5.9 Quantitative and statistical analyses .................................................................... 134

Chapter 5 Summary and future direction .......................................................................... 135
5 Summary and future direction .......................................................................................... 135
  5.1 Summary of key findings ............................................................................................ 135
    5.1.1 In-depth proteomic analyses characterizing the direct and urinary expressed prostatic secretion proteomes .......................................................... 135
    5.1.2 Discovery phase: identification of candidate biomarkers of extracapsular prostate cancer ......................................................................................... 136
    5.1.3 Verification phase: targeted quantification of selected prostate cancer biomarker candidates ................................................................. 136
    5.1.4 Additional proteomics analysis of prostate-proximal fluids ................................ 137
  5.2 Future directions .......................................................................................................... 138
  5.3 Concluding remarks .................................................................................................... 142

References .......................................................................................................................... 143
List of Tables

Table 1.1 Descriptors of disease biomarkers................................................................. 14
Table 1.2 Common sources of prostate cancer biomarkers.............................................. 15
Table 1.3 Summary of common labeled and label-free protein quantification techniques using MS......................................................................................................................... 29
Table 2.1 Clinical information of the direct-EPS fluids analyzed by MudPIT proteomics........ 39
Table 2.2 Clinical information for the prostate cancer (5 samples) and benign prostatic hyperplasia (6 samples) EPS-urines analyzed by MudPIT................................................................. 51
Table 2.3 Clinical information for the pooled urine and EPS-urine samples analyzed by MudPIT........................................................................................................................................ 58
Table 3.1 Clinical information of patients used for discovery proteomics.......................... 80
Table 3.2 Fourteen candidates selected by data mining scheme........................................ 89
Table 3.3 Clinical information of patient samples used for Western blotting verification......... 91
Table 3.4 Clinical information of patient samples used to generate EPS-urine pools............ 95
Table 4.1 Overview of clinical information from patients enrolled in Cohort A of EPS-urines. 114
Table 4.2 Limits of detection and quantification for 34 peptides determined by response curves generated using AQUA peptides spiked into urine................................................................. 118
Table 4.3 Overview of clinical information from patients enrolled in Cohort B of EPS-urines. 119
Table 4.4 Relative expression of Phase 3 proteins and peptides in EPS fluids using various quantitative methods.................................................................................................................. 126
List of Figures

Figure 1.1 The prostate gland. ........................................................................................................... 2

Figure 1.2 Cell types of the prostate gland. .................................................................................. 3

Figure 1.3 Progression of human prostate cancer. ......................................................................... 6

Figure 1.4 Receiver operating characteristic curve. ...................................................................... 13

Figure 1.5 Prostate specific antigen access to circulation. ............................................................. 17

Figure 1.6 Various strategies for novel molecular biomarker identification. ............................... 19

Figure 1.7 Schematic of multidimensional protein identification technology (MudPIT). ............... 25

Figure 1.8 Dynamic concentration range of human plasma proteins. ........................................... 27

Figure 1.9 Selected reaction monitoring mass spectrometry. ........................................................ 30

Figure 2.1 Proteomics analyses of direct expressed prostatic secretions. .................................... 41

Figure 2.2 Functional enrichment of proteins detected in EPS. ..................................................... 42

Figure 2.3 Integrative data mining and comparison to other databases. ....................................... 44

Figure 2.4 Comparison to other published proteomes. ................................................................. 46

Figure 2.5 Integration of multiple data resources with the EPS proteome. ................................. 48

Figure 2.6 Immunohistochemistry images as obtained from the Human Protein Atlas. ............... 50

Figure 2.7 Study workflow. Proteomic analysis of EPS-urine and urine samples. ....................... 53

Figure 2.8 Characterization of the EPS-urine proteome. ............................................................... 54

Figure 2.9 Functional enrichment of proteins detected in EPS-urine and comparison of the EPS-urine proteome to other related body fluids. ........................................................................... 56
Figure 2.10 Comparison between the two proteomics data of the current study

Figure 2.11 Data mining to identify prostate-enriched proteins

Figure 2.12 Additional annotations for prostate-enriched proteins

Figure 2.13 Validation of mass spectrometry data for seven selected proteins

Figure 3.1 Workflow of study design from the discovery of differentially expressed proteins to verification of candidates in prostatic fluids

Figure 3.2 Average number of unique proteins identified in the duplicate MudPIT runs for discovery direct-EPS samples

Figure 3.3 Total number of peptides (A) and total number of spectra (B) recorded for each individual direct-EPS analyzed in the discovery phase

Figure 3.4 Venn diagram depicting the total number of proteins exclusively expressed in the EC or OC groups

Figure 3.5 The current proteomics dataset compared to prostate-related proteomics data from various reports

Figure 3.6 Proportion of proteins with predicted SPs and TMDs

Figure 3.7 Gene Ontology enrichment of the EC and OC direct-EPS

Figure 3.8 Distribution of all 624 proteins identified in the current direct-EPS data set

Figure 3.9 Two significantly enriched pathways within the 133 differentially expressed proteins are involved in hemostatic and immunological processes

Figure 3.10 Additional modules enriched by pathway analysis

Figure 3.11 Protein prioritization scheme for the selection of candidates for verification

Figure 3.12 Measurement of PSA and PAP in direct-EPS

Figure 3.13 Differential expression of verified candidates
Figure 3.14 Differential expression of candidates measured by immunoassays. .......................... 93

Figure 3.15 Standard curves generated for each peptide used for SRM-MS measurements. .... 97

Figure 3.16 Verification of differential expression of candidates in EPS-urines pooled from patients with EC or OC prostate tumors. ........................................................................................................ 98

Figure 3.17 PSA measurements in individual EPS-urines from patients with EC and OC prostate tumors by ELISA .................................................................................................................. 99

Figure 4.1 Study workflow outlining the three major phases of the current project. ............... 113

Figure 4.2 Peptides in EPS-urine. ............................................................................................... 115

Figure 4.3 Phase 3 peptides. ....................................................................................................... 117

Figure 4.4 Cohort B SRM-MS data quality. .................................................................................. 120

Figure 4.5 Absolute quantification of Phase 3 peptides in EPS-urine. ........................................ 122

Figure 4.6 Comparative quantification of candidates in different subgroups of EPS. ............. 125

Figure 5.1 Overview of all proteomics datasets generated for various prostate-proximal fluids. ........................................................................................................................................ 138
List of Appendices

Appendix Table 1.1  Definitions of the American Joint Committee on Cancer for prostate cancer tumor staging

Appendix Table 2.1  Proteins confidently identified by MudPIT proteomics in 9 direct-EPS samples

Appendix Table 2.2  Proteins confidently identified by MudPIT proteomics in 9 direct-EPS samples that could be mapped to the Human Gene Atlas mRNA microarray

Appendix Table 2.3  Proteins confidently identified by MudPIT proteomics in 9 direct-EPS samples that could be mapped to an antibody in the Human Protein Atlas

Appendix Table 2.4  Proteins confidently identified by MudPIT proteomics in 9 direct-EPS samples and 22Rv1 prostate cancer culture medium

Appendix Table 2.5  Proteins confidently identified by MudPIT proteomics in 9 direct-EPS samples and PC3 prostate cancer cell culture medium

Appendix Table 2.6  Proteins confidently identified by MudPIT proteomics in 9 direct-EPS samples and LNCaP prostate cancer cell culture medium

Appendix Table 2.7  Proteins confidently identified by MudPIT proteomics in 9 direct-EPS samples and seminal plasma

Appendix Table 2.8  Proteins confidently identified by MudPIT proteomics in 9 direct-EPS samples and normal urine

Appendix Table 2.9  Proteins confidently identified by MudPIT proteomics in 9 direct-EPS samples and $\geq$ 10-fold above median in prostate tissue on the Human Gene Atlas microarray

Appendix Table 2.10  Proteins identified in either the individual EPS-urines or the pooled samples pre- and post-DRE
Appendix Table 2.11  All 49 prostate-enriched proteins identified in the current study including additional annotations

Appendix Table 3.1  Protein report

Appendix Table 3.2  Peptide report

Appendix Table 3.3  Pathway enrichment analysis for major network modules identified by clustering algorithm

Appendix Table 3.4  Peptide sequences used for SRM-MS

Appendix Table 3.5  SRM-MS relative quantification of peptides

Appendix Figure 3.1  Transitions and integrated quantifier ions

Appendix Table 4.1  Information for the Phase 1 (232) peptides selected for SRM-MS evaluation

Appendix Table 4.2  Fold changes of diagnostic and prognostic peptides in Cohort A measured by SRM-MS

Appendix Table 4.3  Figures of merit for 34 peptides selected for verification

Appendix Table 4.4  Cohort A patient information

Appendix Table 4.5  Cohort B patient information

Appendix Table 4.6  Functions of Phase 3 protein candidates

Appendix Figure 4.1  Response curves for each of the 34 peptides analyzed in the verification phase

Appendix Figure 4.2  Correlation between the 34 Phase 3 peptides from EPS-urine of Cohorts A and B
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMACR</td>
<td>alpha-methylacyl-coenzyme A racemase</td>
</tr>
<tr>
<td>AR</td>
<td>androgen receptor</td>
</tr>
<tr>
<td>AUC</td>
<td>area under the curve</td>
</tr>
<tr>
<td>BCR</td>
<td>biochemical recurrence</td>
</tr>
<tr>
<td>BPH</td>
<td>benign prostatic hyperplasia</td>
</tr>
<tr>
<td>CA125</td>
<td>cancer antigen 125</td>
</tr>
<tr>
<td>CID</td>
<td>collision induced dissociation</td>
</tr>
<tr>
<td>CNA</td>
<td>copy number alteration</td>
</tr>
<tr>
<td>CV</td>
<td>coefficient of variation</td>
</tr>
<tr>
<td>DIA</td>
<td>data independent acquisition</td>
</tr>
<tr>
<td>DRE</td>
<td>digital rectal examination</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EC</td>
<td>extracapsular</td>
</tr>
<tr>
<td>ELISA</td>
<td>enzyme linked immunosorbent assay</td>
</tr>
<tr>
<td>EPS</td>
<td>expressed prostatic secretions</td>
</tr>
<tr>
<td>ESI</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>FDR</td>
<td>false discovery rate</td>
</tr>
<tr>
<td>FFPE</td>
<td>formalin fixed paraffin embedded</td>
</tr>
<tr>
<td>FI</td>
<td>functional interaction</td>
</tr>
<tr>
<td>FT-ICR</td>
<td>Fourier transform ion cyclotron resonance</td>
</tr>
<tr>
<td>GO</td>
<td>Gene Ontology</td>
</tr>
<tr>
<td>HGPN</td>
<td>high grade intraepithelial neoplasia</td>
</tr>
<tr>
<td>HPA</td>
<td>Human Protein Atlas</td>
</tr>
<tr>
<td>ICAT</td>
<td>isotope coded affinity tags</td>
</tr>
<tr>
<td>IL-6sR</td>
<td>interleukin 6 receptor</td>
</tr>
<tr>
<td>iTRAQ</td>
<td>isobaric tags for relative and absolute quantification</td>
</tr>
<tr>
<td>KLK3</td>
<td>kallikrein 3</td>
</tr>
<tr>
<td>LC</td>
<td>liquid chromatography</td>
</tr>
<tr>
<td>LOD</td>
<td>limit of detection</td>
</tr>
<tr>
<td>LOQ</td>
<td>limit of quantification</td>
</tr>
<tr>
<td>m/z</td>
<td>mass-to-charge ratio</td>
</tr>
<tr>
<td>MALDI</td>
<td>matrix assisted laser desorption ionization</td>
</tr>
<tr>
<td>miRNA</td>
<td>microRNA</td>
</tr>
<tr>
<td>MME</td>
<td>membrane metalloendopeptidase</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>MS/MS</td>
<td>tandem mass spectra</td>
</tr>
<tr>
<td>MSMB</td>
<td>microseminoprotein-β</td>
</tr>
<tr>
<td>MudPIT</td>
<td>multidimensional protein identification technology</td>
</tr>
<tr>
<td>NC</td>
<td>noncancer</td>
</tr>
</tbody>
</table>
Chapter 1
Introduction and thesis overview

Portions of this chapter have been reformatted from the original publication: Kim, Y., Kislinger T. (2013). Novel approaches for the identification of biomarkers of aggressive prostate cancer. Genome Medicine 5:56

1 Introduction and thesis overview

1.1 The prostate gland

1.1.1 Anatomy and histology of the prostate gland

The prostate is a secretory gland that sits at the base of the bladder in the male urogenital tract (Figure 1.1). It surrounds the urethra and functions to produce fluid that consists of factors that nourish sperm and aids in fertilization.

Three major regions of the prostate that differ histologically and biologically exist: the peripheral, central, and transition zones (1). The transition zone is commonly the site of benign prostatic hyperplasia, whereas carcinoma occurs primarily in the peripheral zone of the gland (1, 2). Surrounding the prostate is an integral fibromuscular band known as the prostatic capsule.

Within the epithelium of the prostate, three distinct cell types exist (Figure 1.2): secretory luminal cells, basal cells, and neuroendocrine cells (2). Secretory luminal cells predominate the epithelium and function to produce prostatic proteins, including prostate specific antigen (PSA), that are secreted into the lumen of the prostatic duct. They are characterized by the expression of cytokeratin 8 and 18, CD57, and androgen receptor (2-7). The basal cells form a continuous layer, separating luminal cells from the basement membrane and stroma (1). They do not produce secretory proteins and are characterized by the expression of cytokeratins 5, 14, and CD44 (3-8). Lastly, a small population of neuroendocrine cells are dispersed throughout the basal cell layer (2). These cells are believed to support the growth of the secretory luminal cells (9, 10), and can, in rare cases (1% of all prostate cancers), become neoplastic, producing highly aggressive small cell carcinoma of the prostate (2).
Figure 1.1 The prostate gland. The secretory gland is located at the base of the bladder in the male urogenital tract. The prostatic urethra runs through the organ, carrying urine through it. Three major zones of the prostate exist: the central, transition, and peripheral zones.
3

Figure 1.2 Cell types of the prostate gland. Secretory luminal cells predominate the prostate epithelium, releasing proteins into the lumen of the prostatic acini. Basal cells separate the luminal cells from the basement membrane and stroma. A small population of neuroendocrine cells are dispersed throughout the basal cell layer.

1.2 Prostate cancer

1.2.1 Prostate cancer epidemiology and statistics

Prostate cancer is the most common cancer among Canadian men (excluding non-melanoma skin cancer), occurring in 1 of every 8 men and killing every 1 in 28. In 2015 alone, 24,000 new cases and 4100 deaths are estimated in Canadian men (11). Most patients are diagnosed after age 65 and most men over 85 years of age have histological prostate cancer (12). The world-wide incidence rate of this disease can vary considerably, with the highest rates occurring among black American men and the lowest in India (13). These differences can be attributed to genetic
susceptibility, exposure to external risk factors, or the accuracy of cancer registration across countries (14). Despite these daunting statistics, only 15-20% of men will present with aggressive disease (15, 16), while the majority of patients remain asymptomatic for many years.

The incidence of prostate cancer is climbing in both high and low-risk populations around the world for different reasons (17). For instance, in the US, the wide-spread use of PSA screening has led to the identification of asymptomatic men who harbor the disease (18), while in Sweden, where screening is much less common, the frequent use of transurethral resection of the prostate and random biopsies has led to a staggering 100% increase in prostate cancer detection over the last three decades (18). These practices cannot explain, however, the rising incidence of prostate cancer in developing countries, where unidentified external risk factors likely play a role (19).

Epidemiological studies have demonstrated a stronger familial clustering of prostate cancer amongst patients when compared to age and race-matched controls (20). For instance, brothers of young prostate cancer patients (age at diagnosis <63 years) have a four-fold increased risk of the disease compared to their brothers-in-law and the general population (21). Brothers and fathers of prostate cancer patients have a 76% higher risk of prostate cancer than the male first-degree relatives of control subjects (22). Interestingly, the risk of prostate cancer increases for first-degree relatives of prostate cancer patients with decreasing age at diagnosis in the patients; thus, suggesting that screening in men with a family history of prostate cancer should begin at an earlier age (23, 24). Polymorphisms in genes regulating androgen metabolism and metabolites of oxidative stress related to apoptosis (25-32) have been linked to prostate cancer. Androgens play a vital role in the development of the prostate, and medical castration (withdrawal of testosterone) is a method of treatment for patients with metastatic prostate cancer. Animal studies have demonstrated that testosterone and dihydrotestosterone can induce the formation of prostate tumors; however, a definitive study in humans correlating androgen concentrations and prostate cancer has yet to surface (14). Dietary factors such as high intakes of alpha-linolenic acid and calcium, as well as red meat cooked at high temperatures, may be linked to increased risk of disease (14). Other factors such as smoking and alcohol consumption have also been investigated with varying results.
1.2.2 Histopathology and clinical progression of prostate cancer

Prostate tumors predominantly arise from the malignant transformation of luminal epithelial cells (adenocarcinomas), accounting for >95% of all cancers of the prostate. Other, much less prevalent categories of prostate cancer are ductal adenocarcinoma, mucinous carcinoma, signet ring carcinoma, and neuroendocrine. Primary prostate cancer consists of multiple independent and genetically distinct foci (33-37), while different metastases within an individual may be clonally related (38, 39). This suggests that metastatic cancer may arise from the selective advantage of individual clones during cancer progression (40). Age is the greatest risk factor for prostate cancer; even still, prostates from healthy men between the ages of 20 and 40 frequently bear histologic foci of prostate cancer (41-43). This suggests that, although multiple neoplastic transformation events may occur in the prostate, most of them only give rise to latent disease and do not progress to clinical presentation. Furthermore, latent prostate cancer may arise from a different pathogenic program from clinical prostate cancer, whereby activating events are mitigated or there is sufficient tumor suppression to maintain the foci in a subclinical state (40).

High grade prostatic intraepithelial neoplasia (HGPIN), which occurs in 0.7-7.1% of needle biopsies (44), is thought to be an immediate precursor of early invasive prostate cancer (2, 45, 46), although not a prerequisite for carcinomas (Figure 1.3). This is supported by their existence in proximity to carcinomas within the peripheral zone and their resemblance to early invasive carcinoma including multifocality, similarity in chromosomal abnormalities and reduction in the expression of the differentiation markers E-cadherin and vimentin (2, 47-53). In contrast, neoplasms can be distinguished from early invasive cancer by the disruption of the basal layer, which is completely lost in cancer (54). Prostatic intraepithelial neoplasia is histologically characterized by the presence of luminal epithelial hyperplasia, reduction in basal cells, enlargement of nuclei and nucleoli, cytoplasmic hyperchromasia, nuclear atypia, and a marked elevation of cellular proliferation markers (45).
Figure 1.3 Progression of human prostate cancer. Adapted Shen and Abate-Shen, Genes Dev. 2010 © Cold Spring Harbor Laboratory Press. Prostate cancer arises as a result of genetic events that culminate in the transformation of the secretory epithelial cells (adenocarcinoma). Prostatic intraepithelial neoplasia is thought to be a non-obligatory precursor for prostate cancer.

Secondary metastasis of prostate cancer most commonly occurs in bone, where it forms characteristic osteoblastic lesions (55, 56). Although the mechanisms by which prostate cancer cells form metastases in bone are poorly understood, tumor cells from patients with localized and advanced disease can be detected in circulation as well as in bone marrow (57-59).

1.2.3 Current concepts of insignificant versus significant prostate cancer

Insignificant and indolent are terms used interchangeably to prostate describe cancer that is low-grade, small volume, organ-confined (OC), and thus not likely to progress to clinical significance in the absence of treatment (60). The Epstein preoperative criteria was developed in an attempt to clinically and pathologically define subgroups of patients with insignificant disease who may be spared treatment (61). Insignificant disease is defined as: PSA <10 ng/ml, organ-confinement without extraprostatic extension, no Gleason pattern 4/5, and <33% positive biopsies (60-64). Extraprostatic or extracapsular (EC) extension refers to the presence of tumor beyond the confines of the prostate, and is defined as carcinoma mixed with periprostatic adipose tissue, or bulging out beyond the contours of the gland. The presence of extraprostatic extension is associated with an increased risk of biochemical recurrence (BCR) (65). The reported incidence of clinically insignificant tumors is approximately 40% (62); however, about 30% of patients are misclassified based on the criteria and present significant disease upon inspection of their radical
prostatectomy (RP) specimens (60). Furthermore, even insignificant tumor patients have a 10% 5-year risk of BCR after treatment (66). Therefore, there is a clear and urgent need for identifying truly biologically significant prostate cancer.

Between 27% and 53% of patients who receive definitive treatment for prostate cancer develop BCR within 10 years of treatment, as indicated by two consecutive PSA values of >0.2 ng/ml (67, 68). However, only about 30% of patients with PSA relapse develop clinical recurrence and 6% of them die from it (69, 70). Interestingly, the rate of recurrence appears to be correlated with the risk of disease. Kapadia et al., demonstrated that in patients treated with radiotherapy, BCR occurred in 8%, 15%, and 36% of low-risk, intermediate-risk, and high-risk patients, respectively (71). Time to post-RP BCR, PSA velocity and doubling time, pathological stage and Gleason score are important clinical parameters for stratifying patients into risk groups for prostate cancer-specific mortality (72). Approximately half of the patients develop local disease, while the other half develops distant metastatic disease with or without local disease. Patients may undergo diagnostic imaging and biopsies to evaluate clinical recurrence and may take a course of salvage therapy (radiotherapy, hormone therapy, salvage RP), with varying outcomes depending on factors such as age, pre-treatment PSA, and the presence of distant metastatic disease.

1.2.4 Genetic alterations and molecular subtypes of prostate cancer

Genetic mutations found in prostate cancers occur in abrupt periodic bursts, resulting in large scale reshuffling of DNA leading to cancer (73). This phenomenon, known as “chromoplexy”, is a process by which distant genomic regions are disrupted in a coordinated fashion that eliminates cancer-suppressor genes, leaving genes that may drive cancer progression. Common mutations in prostate cancers are somatic copy number alterations (CNA), point mutations, structural rearrangements, and aneuploidy (74). Copy number alterations are present in 90% of prostate cancers, and are more common in prostate cancer than other types of cancers (75).

Highly altered primary tumors are associated with increase risk of relapse (76). Deletions can be seen on chromosomes 6q, 8p, 10q, and 13q and include tumor suppressing genes such as NKGX3-1, pTEN, BRCA2, and RB1. Meanwhile, frequent amplification of chromosomes X, 7, 8q, and 9q can be observed, which include the androgen receptor (AR) and MYC oncogenes (74). Structural rearrangements are more commonly observed in high-risk prostate tumors, of which
the TMPRSS2-ERG and ETS fusions are present in 50% of cases (77). Interestingly, recurrent mutations in SPOP (involved in JNK and Hedgehog signaling pathways) may define a new molecular subtype of prostate cancer since this mutation is only found in ETS fusion-negative tumors and exhibit a pattern of distinct alterations in 5q21 and 6q21 (78).

Like breast, renal, and ovarian cancers, primary prostate cancer has a somatic mutation rate of 1 x 10^-6 to 2 x 10^-6 (79-81), leading to thousands of point mutations that can exist in the prostate tumor genome; however, only about 20 per genome may have influence on protein stability and function (74). In prostate cancer, point mutations have been observed in TP53, PTEN, RB1, and PIK3CA tumor suppressors (78, 81-83), and activating mutations have been reported in the oncogenes KRAS and BRAF.

1.2.5 Current methods of prostate cancer and diagnosis

In the absence of symptoms, early screening of prostate cancer begins with a digital rectal examination (DRE) and a PSA blood test. The DRE is a physical examination in which the physician examines the prostate for nodules through the rectum. Prostate specific antigen, also known as kallikrein 3 (KLK3), is a glycoprotein and serine protease that is produced by the secretory epithelial cells of the prostate. Elevated levels of PSA in the bloodstream (the most widely accepted cutoff for PSA is 4.0 ng/ml) is associated with prostate cancer. However, PSA increases with benign prostatic conditions such as benign prostatic hyperplasia (BPH) and prostatitis, as well as transient causes such as acute urinary retention and ejaculation within 48 hours of testing (84).

The estimated sensitivity and specificity of the PSA test at a cutoff of 4.0 ng/ml was ~21% and 91%, respectively, for detecting prostate cancer (85). The positive predictive value (PPV; i.e., the proportion of men with PSA >4.0 ng/ml who have prostate cancer) is ~30% (86-88), thus about 1 man in 3 with an elevated PSA will have prostate cancer upon biopsy; with PSA >10 ng/ml, the PPV increases to ~40-60% (87, 89). The negative predictive value (NPV; i.e., the proportion of men with a PSA ≤4ng/ml who do not have prostate cancer) is 85% (90). Prostate specific antigen screening has been shown to reduce prostate cancer-specific mortality by 20% but the number of men needed to be treated in order to prevent one death is estimated to be 18-29 (91).
The confirmation of a prostate cancer diagnosis can only be established by examining biopsy samples. The biopsy is performed with transrectal ultrasound (TRUS) imaging to help guide the biopsy needles in the prostate. Typically, 6 or more cores are taken to sample different areas of the organ. Prostate cancer from biopsy specimens is marked by the absence of p63 and cytokeratin 5 and 14, markers for basal cells (92, 93), while α-methylacyl-CoA racemase is upregulated in cancerous lesions (93-95). If cancer is identified, the pathologist will determine the grade and stage of the cancer.

Disease prognosis is defined as the prediction of malignancy in the absence of or post therapy (96). Prognostic factors are those that, independent of treatment, predict relapse or progression (96). Currently, pathologic Gleason grading is considered to be the best predictor of outcome (97). Devised by Dr. Donald Gleason, this procedure involves the histologic evaluation of the extent of glandular differentiation and pattern of arrangement of carcinoma cells in hematoxylin and eosin (H&E) stained tissue sections (98). Five grades, from 1 (well differentiated) to 5 (most poorly differentiated), are assigned to the most predominant growth pattern (primary grade) and the second most common pattern (secondary grade), which are summed to generate a Gleason score ranging from 2 to 10.

Patients with Gleason scores 7 or higher are at increased risk of extraprostatic extension and recurrence after therapy (99, 100). Furthermore, individuals with Gleason 4+3 tumors may be at greater risk of prostate cancer-specific mortality than Gleason 3+4 patients (101). Poorer prognosis is associated with Gleason 4 or 5 cancers, while patients with tumors composed of Gleason 3 pattern are unlikely to progress and die of the disease (101, 102). The defining difference between Gleason 3 and 4 is a focal loss of integrity of discrete glandular units of the tumor. Gleason 3 cancers consist of acini that have a smooth and rounded edge with an intact basement membrane, while Gleason 4 tumors are composed of poorly differentiated, fused glands, frequently with irregular borders that invade the stroma (103). Gleason 5 pattern is represented by sheets of cancerous cells due to the complete loss of any rounded glandular shape (54).

Gleason grading has also been linked to various clinical end points such as clinical stage, response to therapy, biochemical failure, progression to metastatic disease, and cancer-specific and overall survival (104-106). Needle biopsy Gleason grade is typically combined with serum
PSA and clinical stage in Partin tables, which predict risk for extraprostatic extension, seminal vesicle invasion, lymph node metastasis, and aid in treatment planning (107). Needle biopsy Gleason grading provides some idea of the pathologic stage, but is not absolutely accurate; patients with low-grade prostate cancer may still progress and develop tumors that spread outside of the prostate, while some high grade tumors may never extend outside of the gland. Although Gleason grading allows for the assignment of two separate patterns in an individual sample, thus controlling for some degree of tumor heterogeneity within the gland, an average of 2.7 Gleason grade patterns are observed in whole prostates (33), with some even having four different patterns (108). Even tertiary patterns of high grade Gleason pattern 4 or 5 that make up <5% of the tumor have been found to impact stage and progression (109), so high grade tertiary patterns are now recommended to be included in reports (107). In addition, it has been recommended that the most predominant pattern and the highest grade should be summed as the Gleason score, in cases where the highest grade is not the most predominant nor the second most common pattern observed (110).

Staging is performed in order to assess the extent of spread of the cancer within the prostate or to other parts of the body and to facilitate treatment planning. Various tests are performed in order to gather sufficient information for an accurate assessment of the disease, including imaging and biopsies. The American Joint Committee on Cancer TNM cancer staging system (Appendix Table 1.1) is most widely used and is based on five factors: 1) the extent of the primary tumor (T category); 2) the involvement of lymph nodes (N category); 3) the presence of metastasis (M category); 4) serum PSA at diagnosis; and 5) the Gleason score (obtained from biopsy or prostatectomy). Nomograms are multivariate prognostic algorithms that combine different prognostic factors and facilitate treatment decision-making. Preoperative nomograms include information such as ethnicity, prostate weight, serum PSA, clinical stage, and Gleason score. Postoperative nomograms incorporate additional information such as surgical margins, seminal vesicle invasion, and regional lymph node involvement.
1.3 Prostate cancer biomarkers

1.3.1 Introduction to cancer biomarkers

Accurate and timely assessment of prostate cancer prognosis remains one of the most significant clinical challenges in prostate cancer management. Rapid advances in molecular technologies are likely to lead to significant progress in the foreseeable future. Despite these technological strides, prostate cancer is still over-diagnosed and many patients are unnecessarily treated. Possible reasons are the multifocal and heterogeneous nature of this disease leading to frequent misclassification of patients, intra-institutional variability, and patient variability; all of which contribute to the lack of well-defined and validated prognostic biomarkers.

A biomarker is a measureable biological indicator that can provide information about the presence or progression of a disease, or the effects of a given treatment. A clinically useful biomarker should have high sensitivity and specificity, as well as high PPV and NPV, and facilitate clinical decisions to administer optimal care (111) (Table 1.1).

According to the Early Detection Research Network (112), a biomarker should undergo five major phases of development before it can be confidently utilized under clinical settings and benefit the population. These phases are: i) preclinical exploratory studies, where tumor/aggressive disease-associated samples are compared to non-tumor/indolent disease specimens in order to identify molecular characteristics that distinguish both cohorts and can be further explored; ii) clinical assay development and validation, whereby an assay that can accurately measure the biomarker and reliably segregate tumor from non-tumor specimens is developed; iii) retrospective longitudinal studies that utilize specimens from individuals who were monitored over time for the development or progression of disease (e.g., patients who progress from indolent to aggressive prostate cancer) are compared to individuals who do not develop disease or do not progress; iv) prospective screening studies that are performed using the assay in order to evaluate the extent of disease at the time of detection; and v) randomized control studies that are performed in order to determine the reduction of disease burden in the population, as a result of performing the assay.
1.3.2 Clinical and analytical considerations for cancer biomarker assay development

Once a putative cancer biomarker has been determined, a clinically useful assay must be developed. Clinical validation involves determining the sensitivity and specificity, which measure the diagnostic accuracy of a test (113) (Table 1.1). Clinical sensitivity refers to the ability of a test to correctly classify individuals affected by the disease. Specificity is the ability of the test to correctly identify individuals without the disease. These parameters depend on the decision threshold used for a test; that is, a predetermined threshold that, when exceeded, indicates disease or state (113). Receiver operating characteristic (ROC) analysis can be used to evaluate the performance of a clinical assay and is used for determining the threshold (Figure 1.4). Various methods of calculating the optimal decision threshold exist. The most common method is to find the point on the curve closest $x = 0$, $y = 1$ coordinates (114, 115). The area under the curve (AUC) of the ROC curve is a measure of the sensitivity and specificity of a test over all threshold values. An ideal assay would have an AUC of 1.0 (100% sensitivity and specificity), while an assay with no discriminatory power would have an AUC of 0.5.
Figure 1.4 Receiver operating characteristic curve. A biomarker with good discriminatory power has an AUC closer to 1.0 (e.g., marker A). A biomarker with no discriminatory power has an AUC of 0.5 (e.g., marker B).

The sensitivity and specificity of a biomarker tell us about the accuracy of the test, but do not provide information about the probability of disease. The clinical prevalence of the disease in question also plays an important role in the performance of an assay. If the prevalence of the disease is high in the population, it is more likely that an individual who tests positive for a disease will truly have the disease than if the prevalence of the disease is low. The parameters that need to be evaluated are the PPV and the NPV. The PPV indicates the likelihood that, given a positive test result, the patient actually has the disease; the NPV indicates the likelihood that, given a negative test result, the patient is free of the disease in question. Although an assay may have high accuracy, low prevalence of the disease may limit its adoption into clinical use (113). For instance, the ovarian cancer biomarker, cancer antigen 125 (CA-125), has a PPV of 4%; hence, only 4 out of 100 positive test will indicate actual disease (113, 116).
Table 1.1 Descriptors of disease biomarkers.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic biomarker</td>
<td>A measureable factor that is a surrogate manifestation of the disease in question.</td>
</tr>
<tr>
<td>Prognostic biomarker</td>
<td>A measurable factor that predicts the probable course of disease in the absence of treatment.</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>The proportion of individuals who test positive for the biomarker and have the disease. Sensitivity = True Positives / (True Positives + False Negatives)</td>
</tr>
<tr>
<td>Specificity</td>
<td>The proportion of individuals who test negative for the biomarker and do not have the disease. Specificity = True Negatives / (False Positives + True Negatives)</td>
</tr>
<tr>
<td>Positive predictive value (PPV)</td>
<td>The probability that individuals with a positive test result truly have the disease. PPV = True Positives / (True Positives + False Positives)</td>
</tr>
<tr>
<td>Negative predictive value (NPV)</td>
<td>The probability that individuals with a negative test result truly do not have the disease. NPV = True Negatives / (False Negatives + True Negatives)</td>
</tr>
</tbody>
</table>

Assays targeting a single or a multitude of biomarker candidates must also meet rigorous analytical standards for reliable quantification and reproducibility. Parameters that should be accounted for include precision, limit of detection (LOD), limit of quantification (LOQ), linearity and working range, and specificity (113). These so called Figures of Merit (117) may be affected by differences in sample matrix and external factors that affect the assay’s sensitivity (117). Therefore, a calibration curve, represented by the response versus the concentration of a given analyte (117), should be generated for each analyte for which an assay is being developed. Calibration is defined as the procedure for determining the relationship between the assay’s generated signal for a given analyte and the analyte’s concentration (113). Preferably in the sample matrix that the assay will be used in, a series of concentrations of the analyte (spanning at least two orders of concentration magnitude) is introduced into the matrix. The data is plotted as the measured concentration versus the spiked-in concentration. The slope of the curve represents the analytical sensitivity of the assay for the analyte. A slope of 1 and an intercept of 0 is ideal, signifying a perfect correlation between the theoretical and measured concentrations of the analyte.

The calibration curve is also used to define the linear range of detection, the LOD and LOQ, and the reproducibility of the assay. The lowest concentration at which an analyte can be reliably distinguished from noise is represented by the LOD (113). The lower LOQ is the lowest analyte concentration at which a quantitative measurement can be reliably made, while the upper LOQ is described as the highest concentration of the analyte that falls within the linear range of the assay (117). The calibration curve can also provide information about possible interference or endogenous signals from the sample matrix, determined by measurements of the blank sample.
Precision, usually expressed as standard deviation (SD), variance, or the coefficient of variation (CV), is determined by the closeness of agreement between a series of measurements for an analyte in a sample under specific conditions. Within-run precision (repeatability) refers to repeated measurements performed under the same conditions. Within-laboratory precision (reproducibility) is determined by measurements made across laboratories, under different conditions such as analysis by different technicians (113). Low analyte concentrations may hamper precision as the detection limit of the assay is reached. Biological variation – i.e., the fluctuation of an analyte concentration around a homeostatic set point – can greatly affect precision and thus should be taken into consideration.

1.3.3 Biological sources of biomarkers

When faced with the task of identifying disease biomarkers, “gold standard” samples would be those in which cases and controls differ absolutely and exclusively for the disease of concern (118). This would require that cases and controls be identical in all respects except for the condition for which the biomarker is being identified. Model systems, including cell lines and genetically homogeneous animals that reflect human conditions may be useful in this respect (118). These systems allow for ample longitudinal sample collection, have uniform genetic background, and are typically subjected to highly controlled environments that greatly abrogate the vast variation seen amongst human subjects. However, models can only provide insight into potential biomarkers candidates, and should not be wholly relied upon for subsequent analyses.

Table 1.2 Common sources of prostate cancer biomarkers.

<table>
<thead>
<tr>
<th>Source</th>
<th>Available markers</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tissue</td>
<td>Histological patterns, cellular proliferation, molecular factors, genetic aberrations</td>
<td>Information about protein expression and prostate biology</td>
<td>Invasive, snap-frozen/FFPE, highly complex, intra-patient variation</td>
</tr>
<tr>
<td>Blood</td>
<td>Circulating tumor cells, circulating tumor DNA, microRNA, soluble proteins</td>
<td>Non-invasive, large quantities</td>
<td>Large dynamic range, dilution of markers, complex, distal from tumor site</td>
</tr>
<tr>
<td>Tissue proximal fluids (e.g., EPS, urine)</td>
<td>microRNA, RNA (e.g., PCA3), soluble proteins, exosomes</td>
<td>Proximal to tumor, higher concentration of factors, non-invasive, routine collection</td>
<td>Non-standardized, intra-patient variation</td>
</tr>
</tbody>
</table>
Human tissues represent an essential class of biomarker sources (Table 1.2). Indeed, tissue can provide histopathological features, molecular factors, and genetic aberrations or patterns that may be indicative of disease, patient outcome, and facilitate treatment design. Allowing for direct evaluation of protein or gene expression, tissue samples can provide important insights into the biological mechanisms of a given condition (119). However, tissues are obtained through invasive means that are associated with risks for the patients and limit repeat measurements. Blood (i.e., serum or plasma) is an attractive alternative to tissue specimens that is already in routine clinical use. Factors such as circulating tumor cells and DNA, microRNA, metabolites, and soluble proteins can be non-invasively and repeatedly collected with ease. Unfortunately, the immense complexity, wide dynamic range, and dilution of specific disease markers make it extremely difficult to discover protein biomarkers in this fluid. To facilitate this goal, extensive fractionation strategies and affinity approaches that remove some of the most abundant blood constituents can be utilized, but introduces additional sample preparation steps.

1.3.3.1 Tissue proximal fluids

Tissue proximal fluids are a class of biofluids that are in direct contact with the disease site. It is rationalized that local bathing of the site results in enrichment of protein biomarkers compared to distal circulating fluids (120, 121). In the context of prostate cancer, prostate proximal fluids include urine, seminal fluid, semen, and expressed prostatic secretions (EPS). Expressed prostatic secretions can be collected directly from the prostate just prior to RP. Although this fluid is presumably abundant in prostatic proteins, it is not possible to collect more than once.

Urine has been shown to be a source of microRNAs (122), RNAs (123), soluble proteins (124), and exosomes (125), and can be considered as a prostate proximal fluid since the prostatic urethra carries it through the prostate. In the context of a DRE, urine containing EPS (EPS-urine) can be routinely collected in voided urine. Indeed, EPS-urine has previously been interrogated for prostate cancer-related proteins and prostasomes (126, 127). A major advantage of urine over serum or plasma, with regards to protein biomarker detection, is that its contents remain relatively stable and do not undergo massive proteolytic degradation (128). Nevertheless, variations in collection volume may result in varying protein concentrations, underscoring the need for standardized collection protocols.
1.3.3.2 Biomarkers in biological fluids

Fluctuations in the levels of molecules in biological fluids, such as blood and tissue proximal fluids, can serve as cancer biomarkers. As cancer progresses, tumor angiogenesis and tissue invasion causes the destruction of the tissue architecture. This results in the release of factors into the interstitial fluid and delivery into the circulation by way of the lymphatics. For instance, PSA, which is normally abundantly expressed by the secretory epithelial cells of the prostate diffuses through the basement membrane, stromal layer, and the walls of blood and lymphatic capillaries into the circulation, and can be seen in the serum of healthy men in the range of 0.5-2 ng/ml (129) (Figure 1.5). In prostatic carcinoma, the destruction of tissue architecture through the loss of basal cells and degradation of the basement membrane, angiogenesis, and disruption of cell attachment, causes PSA to leak out of the glandular lumen and into circulation, reaching serum PSA levels beyond 4 ng/ml (129).

Figure 1.5 Prostate specific antigen access to circulation. Adapted from Kulasingam and Diamandis Nature Clinical Practice. 2008 © Macmillan Publishers Limited. Normally, small amounts of PSA enter the circulation by way of the lymphatics. Under diseased states, the architectural integrity of the prostate is compromised. This leads to PSA leakage out of the gland and subsequent elevation in circulating PSA levels.
Changes in protein abundances may also be explained by aberrant secretion or shedding of secreted or membrane-bound proteins. Indeed, mutations in the signal peptide (SP) domains of proteins causes impairment of the SP function and cellular localization of human proteins (130). Furthermore, a loss of epithelial cell polarity in cancer may result in irregular trafficking of plasma membrane proteins to apical and basolateral domains causing misdirectional release of proteins into the extracellular domain (131). Quantitative changes in molecules may also be explained by fluctuations in gene expression. Alterations in gene or chromosome copy numbers and changes in transcriptional activity have been described for prostate cancer and are highlighted in the following sections.

1.3.4 Adjunct prostate cancer biomarkers

In broad terms, current and proposed alternative or adjunct markers for prostate cancer can be divided into clinical-pathological features and molecular factors. These include the classic Gleason grading and more recently discovered molecular features that might offer insight into disease progression and prognosis. Various methodologies can be utilized to identify these factors (Figure 1.6), and some of them are highlighted in the following sections.
Figure 1.6 Various strategies for novel molecular biomarker identification. Adapted from Kulasingam and Diamandis, Nature Clinical Practice. 2008 © Macmillan Publishers Limited. Advancing technologies have paved the way for disease biomarker discovery. The integration of multiple discovery modalities holds promise for novel molecular biomarkers.

### 1.3.4.1 Cellular markers

Ki-67 is a nuclear protein that is associated with cellular proliferation (132). Its immunohistochemical staining index has been correlated with outcome in treated patients (133-136). Heterogeneous immunohistochemical staining for α-methylacyl-coenzyme A racemase (AMACR) has been correlated with Gleason score (137), and low AMACR gene expression in localized prostate cancer has been linked to recurrence and metastasis (138). Prostate specific membrane antigen (PSMA) is a transmembrane protein expressed in all types of prostatic tissue that is used in the diagnosis of prostate cancer (139). Its over-expression is associated with higher tumor grade, stage, PSA recurrence, and metastatic disease (140, 141).
1.3.4.2 Genetic aberrations as prognostic biomarkers

Focusing on a specific pathway or a group of interrelated genes involved in fundamental tumor biology has also proven useful. Cuzick et al. (142) focused on genes involved in cell cycle progression and measured mRNA expression of 126 genes in formalin-fixed paraffin-embedded prostate cancer tissues. A 31-gene signature was generated on the basis of their correlation with the mean expression of the entire panel of 126 genes. When used to retrospectively score patients who underwent prostatectomy and patients with localized disease, this signature was shown to predict recurrence after surgery and risk of death in conservatively managed patients, independently of Gleason score and other clinical factors. Using comparative transcriptomic analyses, Ding et al. (143) identified the robust activation of the Tgfβ/Bmp-Smad4 signaling pathway in indolent Pten-null mouse prostate tumors. Deletion of Smad4 in the Pten-null mouse prostate led to highly proliferative, invasive, metastatic and lethal tumors. Combined with key molecular players, cyclin D1 and osteopontin, their four-gene signature (PTEN, SMAD4, cyclin D1 and osteopontin) could predict BCR and supplement the Gleason score in predicting lethal metastasis of prostate cancer in patients.

Genomic instability, caused by variations such as CNAs, may also be useful prognosticators of prostate cancer. Recently, Lalonde et al., identified a 100-loci DNA signature that could predict 5-year biochemical relapse in patients treated with radiotherapy or RP; its effect augmented by tumor hypoxia (144). In a comprehensive genomic analysis of prostate cancer, Taylor and colleagues analyzed CNAs in primary prostate tumors and found distinct patient clusters with varying degrees of relapse, with no association with Gleason score (76). Penney and colleagues (145) constructed a 157-gene signature based on the comparison of Gleason ≤6 and Gleason ≥8 patients. When applied to patients with Gleason 7 scores, their signature improved the prediction of lethality, compared to Gleason score alone. DNA methylation patterns in prostate cancer may also provide insight into prostate cancer outcome. Cottrell et al. (146) performed a genome-wide scan in patients with early recurrence, high Gleason score or advanced stage and identified 25 methylation markers that were significantly different between low and high Gleason score patients. Furthermore, the methylations of 3 markers (GPR7, ABHD9, Chr3-EST) were found to be significantly increased in patients with recurrence, as measured by elevated post-prostatectomy PSA levels.
1.3.4.3 MicroRNAs

MicroRNAs (miRNAs) are a class of small, non-coding RNA molecules, involved in the negative regulation of gene expression. Porkka and colleagues (147) demonstrated distinct miRNA expression profiles of BPH, untreated prostate cancers, and hormone-refractory prostate cancers, suggesting a potential prognostic role for miRNAs. Mitchell et al. (148) demonstrated that tumor-derived miRNAs are present in plasma and could show that miR-141 was significantly elevated in the sera of prostate cancer patients, demonstrating the utility of miRNAs as blood-based cancer biomarkers. Khan et al. (149) analyzed localized prostate tumors and adjacent normal tissues, as well as samples from advanced cases using isobaric tags for relative and absolute quantification (iTRAQ) followed by mass spectrometry (MS). Integrating their findings with a cancer microarray database, the authors identified differentially expressed proteins that are targets of miR-128, which was further supported by in vitro experiments demonstrating a role for miR-128 in prostate cancer invasion.

1.3.4.4 Circulating biomarkers

Urokinase plasminogen activator (uPA) and its inhibitor, PAI-1, have been associated with aggressive prostate cancer exhibiting extraprostatic extension and seminal vesicle invasion, and post-prostatectomy recurrence in patients with aggressive disease (150). Preoperative plasma levels of transforming growth factor beta 1 (TGF-β1) has been shown to be a predictor of BCR (151) and, in conjunction with preoperative plasma levels of interleukin 6 receptor (IL-6sR), has been associated with metastasis and progression (152).

Disseminated tumor cells in the bone marrow, a common site of prostate cancer metastasis, have been shown to have an association with metastatic disease and high Gleason score (153, 154). Although disseminated tumor cells may be a prognostic marker of unfavorable outcome in patients with localized disease at diagnosis, attention has shifted to tumor cells that have entered the peripheral blood as these are more easily accessible. The number of circulating tumor cells can be determined at the time of diagnosis and elevated numbers, as indicated by reverse transcriptase polymerase chain reaction for PSA, have been associated with advanced stage and increased Gleason score (155). Goodman et al. (156) determined that prior to treatment, a cut-off
value of 4 circulating tumor cells per 7.5 ml of blood or more was negatively correlated with survival and could predict metastasis.

1.3.4.5 Urinary biomarkers

Prostate cancer antigen 3 (PCA3) is a prostate specific non-coding RNA that was first identified in a comparative transcriptomics study in tumor and adjacent normal tissues (157). Subsequently, a RT-PCR based test was developed for its detection in urinary EPS (158). A ratio of the PCA3:PSA RNA, known as the PCA3 score, is used in combination with other clinical information to guide decisions for repeat biopsy in men who are 50 years of age or older and who previously have had at least one negative prostate biopsies. Interestingly, Nakanishi et al. (159) reported the mean PCA3 score to be significantly lower in patients with low volume and low-grade prostate tumors, compared to patients with advanced tumors. However, the ability of the PCA3 test to predict aggressive prostate cancers is under debate (159-161). Tomlins et al. (77) first reported the occurrence of a recurrent TMPRSS2:ERG (transmembrane protease serine 2 gene fusion with E twenty-six (ETS) transcription factors) fusion transcript. These fusions were detectable in 42% of urinary EPS samples from men with prostate cancer (162), although their presence in urinary sediment was not correlated with biopsy Gleason scores (163). Telomerase is a ribonucleoprotein involved in telomere synthesis and repair (164). Its activity, which can be measured in urinary EPS using the telomeric repeat amplification protocol assay (165, 166), was found to be increased in prostate cancer and has been shown to be associated with Gleason score (165). Urinary annexin A3 and various matrix metalloproteinases have also been shown to have diagnostic/prognostic potential for prostate cancer (167-170).

Approximately 3% of the total urinary protein content is comprised of exosomal proteins (171), thus representing a sub-fraction for discovery of prostate cancer biomarkers (172, 173). Exosomes are small vesicles (40-100 nm) that are secreted by various normal and tumor cells (173), containing protein, RNA and lipids (174). Wang et al. (175) used shotgun proteomics to generate the largest catalogue of urinary exosome proteins to date. In their study, over 3000 unique proteins were identified from samples derived from nine healthy individuals. Exosome secretion is elevated in biofluids of cancer patients, including prostate cancer (176) and exosomes have been shown to be enriched in tumor cell-specific transcripts (177, 178). MiRNA and mRNA are transferable between cells via exosomes and have been shown to be functional in
their new location (179). Nilsson et al. (173), in a proof-of-concept study, showed that urinary exosomes derived from prostate cancer patients contain two known biomarkers (PCA3 and TMPRSS2:ERG) and thus may be used as disease biomarkers.

1.4 Clinical proteomics for biomarker discovery

Proteins are the effector molecules of the cell and are likely to be the most affected in disease (118). Proteomes are dynamic and dictated by cellular mechanisms that regulate gene expression, post-transcriptional events, and post-translational modifications. Furthermore, significant changes occur in proteomes in response to physiological and environmental stimuli. Hence, cataloguing the proteome – the entirety of proteins expressed by a particular cell, tissue, or organism under specific conditions (180) – may provide important indicators of disease.

Clinical proteomics is defined as the analysis of the identity, quality and the quantity of proteins in patient samples (181). Mass spectrometry-based proteomics is the most widely applied tool in clinical proteomics and holds promise to identify new disease biomarkers, drug targets, and aberrant molecular pathways in disease. A typical cancer biomarker discovery workflow consists of a discovery phase, whereby a comprehensive comparative catalogue of proteins under different disease states is generated. This is followed by verification of putative biomarkers using targeted methods of quantification, and finally, validation and clinical assay development (118).

1.4.1 Discovery proteomics using shotgun, bottom-up proteomics

Proteomic analysis of intact proteins (known as top-down proteomics) largely remains limited by sensitivity and throughput (182). Therefore, bottom-up proteomics – based on the analysis of proteolysed peptides emerging from proteins (183) – has been widely adopted in proteomic investigations. Shotgun proteomics refers to the application of bottom-up proteomics to protein mixtures (184-186).

A mass spectrometer consists of an ion source that ionizes analytes, a mass analyzer that measures the mass-to-charge ratio (m/z) of the ionized analytes, and a detector that measures the number of ions of each m/z value (187). Most commonly, proteins and peptides are volatized and ionized using the soft ionization approaches: electro spray ionization (ESI) (188) or Matrix-Assisted Laser Desorption Ionization (MALDI) (189). Because ESI ionizes analytes out of
solution, it is amenable to liquid-based separation such as liquid chromatography (LC), and is the preferred method for the analysis of complex samples (187).

As ionized peptides enter the mass spectrometer, they are guided and manipulated by electric or magnetic fields in a vacuum system. The basic types of mass analyzers used for proteomics are: time-of-flight (TOF), quadrupole, ion trap, Fourier transform ion cyclotron resonance (FT-ICR), and more recently, Orbitrap. These can also exist as hybrids, such as triple quadrupoles, TOF-TOF, and quadrupole TOF. Time-of-flight instruments determine the m/z of an ion based on the amount of time it takes to travel through a flight tube and hit a detector (190-194); lighter ions fly faster than heavier ions, and thus reach the detector sooner. Quadrupoles are essentially mass filters, allowing the selective passage of ions of a defined m/z by applying a particular potential that stabilizes their trajectory down the quadrupole, while filtering out ions of other m/z values (195, 196). A mass spectrum is generated by the step-wise application of different potentials and detecting the ions that pass through at each m/z value over the m/z range (196). Ion traps capture and accumulate incoming ions in a dynamic electric field; subsequently ejecting them to the detector according to their m/z. Orbitrap mass analyzers trap ions around a central spindle electrode (197-199). The oscillation frequency of trapped ions is mass dependent and can be converted from the recorded time-domain signal into a m/z spectrum using the Fourier transformation algorithm (200). Similarly, FT-ICR mass analyzers resolve the masses of ions by recording the rotation frequency of the ions in a magnetic field (201). Detectors (e.g. electron multiplier) detect ions that pass through the mass analyzer and generate an electric current that is proportional to their abundance.

The workflow of a typical shotgun experiment begins with the enzymatic digestion (usually with the protease trypsin) of a mixture of proteins in a gel or in solution into peptides, followed by LC separation. The generation of high-quality spectra heavily depends on the sufficient fractionation of immensely complex samples (for instance, a typical serum digest contains 400,000 – 600,000 peptides (202)). First reported by Washburn and colleagues (184, 203), multidimensional protein identification technology (MudPIT) is a technique of fractionation, yielding comprehensive proteome coverage (204) (Figure 1.7). Here, peptides are loaded on a biphasic LC column consisting of a strong cationic exchange (SCX) and a C18 reverse phase (RP). The protein mixture is separated by means of SCX using iterative steps of increasing salt concentration (“salt
bumps”) followed by RP chromatography. Peptides are thus separated in two orthogonal dimensions prior to MS analysis.

**Figure 1.7 Schematic of multidimensional protein identification technology (MudPIT).** In this setup, a peptide mixture sample (S) is automatically loaded onto a biphasic pre-column consisting of a strong cation exchange (SCX) and a reverse phase (RP) resin, coupled to an analytical RP column. The mobile phase is water with increasing concentrations of an organic solvent (e.g., acetonitrile). The eluting peptides are subsequently introduced to the mass spectrometer by way of ESI. A series of increasing concentrations of ammonium acetate steps (pink to red circles) further separate increasingly basic (charged) peptides.

In shotgun proteomics, intact peptide ions (known as precursor/parent ions) travel through the mass spectrometer, where their m/z values are recorded to generate a mass spectrum referred to as a MS or survey scan (196, 206). Following each survey scan, precursor ions are isolated based on abundance in a data-dependent manner and subsequently fragmented. This is known as a MS/MS or product ion scan and generates tandem mass spectra. The MS spectrum provides information about the exact mass and charge of a peptide and the MS/MS spectrum provides information about the amino acid sequence of that peptide. This is compared to theoretical, in
silico tandem mass spectra in a protein database and peptides are assigned to protein sequences in a process known as protein inference.

The process of data dependent acquisition collects MS/MS data for only 3-10 of the most abundant precursor ions that exceed a predetermined intensity threshold (207). In highly complex samples, the number of co-eluting ions can significantly exceed the number of MS/MS spectra the mass spectrometer can record (202). Therefore, high abundance peptides are more likely to be selected for fragmentation, while low abundance species will be sampled by chance. This “random sampling” of peptides results in the masking of low abundance peptides by co-eluting peptides of higher abundance, and is affected by sample complexity, chromatographic resolution, ionization efficiency, ion suppression, and MS cycling time (202).

The dynamic concentration range of proteins, especially in clinical samples, remains a significant challenge in proteome analyses by current LC-MS systems. For instance, the plasma proteome concentration range spans up to 12 orders of magnitude (Figure 1.8). The issue is further amplified by the presence of highly abundant proteins such as serum albumin (35-30 mg/ml) and immunoglobulins (5-18 mg/ml), which mask lower abundance proteins below one microgram per milliliter (208, 209). To circumvent this, fractionation of the proteome can be performed in order to reduce sample complexity and increase sensitivity of detection. However, this approach is significantly more time consuming, labour intensive, and introduces another level of variability due to the introduction of more steps in sample processing (208). Alternatively, high abundance proteins, such as albumin in plasma, can be selectively removed by immunodepletion, thereby enriching for less abundant proteins (210). Yet another approach exploits the principles of Michaelis-Menten kinetics of protein digestion in order to remove high-abundance proteins (211). In this method, a limited amount of protease, trypsin, is introduced to a lysate, whereby high-abundance proteins are digested before lower-abundance proteins, following Michaelis-Menten kinetics. The resulting peptides are then removed via a molecular weight cut-off spin filter and the remaining proteome is digested under standard digestion conditions and analyzed by LC-MS/MS. This method has been shown to improve the identification and sequence coverage of low-abundance proteins by 3-fold. Other avenues of tackling the wide dynamic range and the immense heterogeneity of complex proteomes is to enrich for sub-proteomes (for example, N-glycoproteins account for the majority of the current clinically used biomarkers and
drug targets, as they are often secreted by cells or found at the cell surface (212)) and tissue-proximal fluids that may be more abundant in relevant biomarker candidates.

Figure 1.8 Dynamic concentration range of human plasma proteins. Adapted from Schiess et al., Molecular Oncology. 2008 © Federation of European Biochemical Societies. The dotted line represents the concentration range of various plasma proteins. Classic plasma proteins, such as albumin, are in excess of 5 orders of magnitude higher in abundance than tissue leakage products and other potentially disease-related proteins.

1.4.2 Protein quantification using mass spectrometry

Mass spectrometry-based quantification strategies can be categorized as label-free or labeled (Table 1.3). Label-free quantification includes spectral counting and extracted ion chromatograms. Spectral counting (202, 203, 213) is an estimate of the protein abundance based on the frequency of MS/MS spectra (spectral counts) generated from a given protein divided by the number of theoretically observable peptides for that protein (known as the protein abundance index) (214). It is based on the assumption that higher abundance proteins produce more tryptic peptides, resulting in higher spectral counts. Indeed, a linear relationship between spectral counts and abundance has been demonstrated over at least two orders of magnitude (202, 215); however, this method biases against low abundance species.
The intensity-based approach uses the measured peptide chromatographic peak areas of peptide precursor ions that belong to a given protein (202, 213, 216, 217). The key to this approach is the alignment of chromatographic peaks in order to achieve accuracy in comparative quantitative data. High quality peaks have aligned retention times, matching precursor ions, charge state, and fragment ions (218). The AUC of the aligned peptides from each sample are derived and compared for estimating relative protein abundances. The AUC has been shown to directly correlate with the concentration of peptides in the range of 10 fmol – 100 pmol (216, 219). One obvious factor that must be carefully taken into consideration when using this approach is the presence of co-eluting peptides and shifts in retention time. It has been demonstrated that both spectral counting and extracted ion chromatograms correlate with protein abundances in complex samples (217, 220-223).

Isotope labels can be resolved either at the MS or MS/MS level, depending on the type of labeling method adopted. These include stable isotope labeling with amino acids in cell culture (SILAC) (224), using mass tags such as isotope-coded affinity tags (ICAT) (225), isobaric tags for relative and absolute quantification (iTRAQ) (226) and tandem mass tags (TMT) (227). Although labeling methods allow for more accurate quantification, additional preparation steps and costs should be considered.

Absolute quantification of proteins by MS is based on stable isotope dilution (SID) (228), which relies on the introduction of predetermined amounts of stable heavy isotope labeled standards into the sample. Isotope dilution has been used in the quantification of small molecules for decades (229), and has been adopted for the absolute quantification of peptides in biological material in recent years. Peptide labeling methods include chemical synthesis (AQUA) (230), biologically synthesized artificial quantification concatemer (QconCAT) peptides (231), and intact isotope labeled protein standard absolute quantification (232-234).

Absolute quantification (AQUA) peptides are highly purified, chemically synthesized peptides with 13C and 15N isotope labels at lysine and arginine residues. These standards have identical sequences and thus the same physicochemical properties and chromatographic elution times to endogenous peptides, but they have a known mass difference that is resolvable by MS. Quantification using this approach is very straightforward – known amounts of AQUA internal standards are spiked into the sample prior to the LC-MS step and the resulting recorded
abundance of the standard is compared to the endogenous peptide. Although a highly attractive method, some caveats exist: due to limitations of chemical synthesis, certain peptides such as those above 15 amino acids in length and those containing chemically reactive residues (e.g., oxidation of tryptophan, methionine, cysteine, etc.) or sequences (such as aspartate-glycine, N-terminal glutamine, etc.) may not be compatible. AQUA peptides are very expensive, due to the labour intensive process of production, purification, and quantification. Furthermore, since AQUA peptides are usually fully tryptic peptides, complete tryptic digestion of the target peptides must be ensured in order to prevent quantification of only the digested fraction (235).

**Table 1.3 Summary of common labeled and label-free protein quantification techniques using MS.**

<table>
<thead>
<tr>
<th>Class</th>
<th>Methods</th>
<th>Strategy</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labeled</td>
<td>Stable isotope labeling (e.g., SID-SRM)</td>
<td>Utilizes synthesized non-radioactive isotope labeled analogs of the analyte as standards</td>
<td>Better quantification accuracy; internally controlled</td>
<td>Requires extra processing steps</td>
</tr>
<tr>
<td></td>
<td>Chemical tagging (e.g., ICAT, iTRAQ)</td>
<td>Introduction of stable isotope labels via a chemical tag</td>
<td>Allows for direct comparisons of multiple samples</td>
<td>Incomplete labeling; difficult to compare across experiments for iTRAQ due to MS/MS level quantification</td>
</tr>
<tr>
<td></td>
<td>Metabolic labeling (e.g., SILAC)</td>
<td>Metabolic incorporation of stable isotopes</td>
<td>Live cells; 100% efficiency of labeling; label introduced before protein purification</td>
<td>Limited to 3 conditions per experiment</td>
</tr>
<tr>
<td>Label-free</td>
<td>Spectral counting</td>
<td>Protein abundance estimated by the number of MS/MS spectra recorded</td>
<td>Correlates linearly with protein abundance over two orders of magnitude; no extra processing steps necessary; can be performed in any type of sample</td>
<td>Semi-quantitative; prone to quantification error</td>
</tr>
<tr>
<td></td>
<td>Peak integration</td>
<td>Intensities of same peptides observed in separate runs compared to determine relative abundance; quantification is at MS level</td>
<td>No labeling steps required</td>
<td>Requires high resolution for accurate quantification; requires peptides to be selected for fragmentation in every run</td>
</tr>
</tbody>
</table>

**1.4.3 Targeted proteomics**

Discovery proteomics does not require a priori knowledge about the contents of the sample being analyzed, but the stochastic nature of shotgun proteomics limits the reproducible detection and quantification of proteins between different experiments. Targeted proteomics approaches selectively analyze only a subset of predefined peptides within a proteome, and hence, offer
much higher sensitivity, selectivity, and accuracy as compared to quantification by discovery methods.

1.4.3.1 Selected reaction monitoring mass spectrometry

Selected reaction monitoring mass spectrometry (SRM-MS) is the most widely used targeted MS technique and has been used for reliably quantifying small molecules (236). In more recent years, it has been adopted as a robust, sensitive, reproducible and specific assay for protein quantification (237-239) and validation of cancer biomarkers from various sources (240-244).

Selected reaction monitoring mass spectrometry allows for the detection and quantification of predefined sets of peptide targets by employing the use of triple quadrupole (QqQ) mass spectrometers that allow for two-stage mass filtering. In SRM-MS mode, a precursor ion of a specific m/z is selected in a front-end mass analyzer (Q1) (Figure 1.9). The ion is then fragmented by collision-induced dissociation (CID) in a collision cell (Q2), and its resulting co-eluting fragment ions (transitions) (206) are selectively measured as a function of time in a second mass analyzer (Q3).

Figure 1.9 Selected reaction monitoring mass spectrometry. By using two stages of mass filtering in a triple quadrupole mass spectrometer, a specific peptide can be isolated, fragmented, and subsequently analyzed. This method offers a high level of sensitivity and specificity and can be coupled to stable isotope dilution for accurate quantification of proteins in complex mixtures.

In order to develop a SRM-MS assay, a priori selection of proteotypic peptides (peptides that are unique to the proteins of interest and are reliably detected by MS) and the determination of their MS coordinates (peptide elution time, most intense fragments, and relative intensities of fragments) is required. Only a small fraction of the millions of peptides resulting from the enzymatic digestion of a cellular proteome are proteotypic (245), and thus, the ability to target
these greatly facilitate peptide detection and alleviate redundant peptide sampling and analysis time (206). Computational tools to guide the selection of proteotypic peptides have been developed based on physicochemical properties of frequently observed peptides from large-scale shotgun experiments (245). Tools such as SRM-MS Collider (246) can identify redundant transitions of a peptide within a proteome background of choice that can help determine potentially interfering transitions. Spectral libraries from previous experiments (either available through online repositories such as the Global Proteome Machine, SRMAtlas, or through own experimental data) can also be used to help select frequently observed proteotypic peptides. Alternatively, recombinant proteins of candidates can be digested and analyzed by MS to generate a spectral library. The acquired spectra can provide information for the peptide SRM-MS assay: m/z of the precursor and its fragments, the preferred ionization charge state, the relative intensity of the fragment ions signals, and the chromatographic retention time (206).

To achieve sufficient sensitivity and precise quantification, the SRM-MS cycle time and dwell time should also be considered. The dwell time refers to the amount of time the instrument spends monitoring a transition. A higher dwell time, that is, the longer the instrument spends measuring a given transition, the higher the sensitivity of the assay for that transition. The cycle time is the sum of all the dwell times of each transition to be monitored in the assay. A longer cycle time results in fewer data points being measured over the chromatographic elution time of each peptide, which can affect both the detection and quantification of peptides. Under conditions where too many transitions are monitored in a single run, the dwell time may be reduced to sufficiently compromise sensitivity and quantitative accuracy. In such cases a scheduled approach, whereby the instrument scans for transitions only during a small window of time centered around a peptide’s expected elution time, can significantly augment the number of monitored transitions in a single experiment. For less abundant peptides, the dwell time can also be increased during the elution window in order to increase sensitivity.

1.4.3.2 Application of selected reaction monitoring mass spectrometry to biomarker research

Validation of proposed biomarker candidates in a major bottleneck in protein biomarker development. The gold standard for clinical protein biomarker quantification is enzyme linked immunosorbent assay (ELISA), which relies on the use of specific antibodies. Although ELISAs
have excellent specificity and sensitivity, the limited number of antibodies and the cost of developing a clinically validated ELISA assay (which runs up to $1 million (247, 248)) severely limit their use for biomarker validation. Protein quantification by SRM-MS has been shown to have good correlation with ELISAs (249, 250), although its sensitivity has been shown to range from 34-60% lower depending on the analyte (251). Considering the vast numbers of proposed biomarkers, the capability of SRM-MS to simultaneously quantify a multitude of analytes in a matter of hours compared to the usually single analyte analysis of ELISA, significantly facilitates the timely appraisal of putative candidates. Selected reaction monitoring mass spectrometry assays can be used to rapidly evaluate putative biomarker candidates, of which the most promising candidates would be carried forward for ELISA assay development. Therefore, SRM-MS can serve as an attractive complementary technique to ELISA in the biomarker validation pipeline.

Selected reaction monitoring mass spectrometry application to the quantification of proteins in complex biological systems has made considerable headway in recent years. Picotti et. al (252) demonstrated the application of SRM-MS to target proteins in a complex biological system (total S. cerevisiae extract) and its ability to cover a dynamic range of over 4-5 orders of magnitude. In their work, low-abundance proteins that were previously unidentifiable by MS were successfully detected and quantified. Hüttenhain et al. (253) developed a high-throughput workflow for the quantification of cancer-associated proteins in human urine and plasma. Their study, which utilized SRM-MS, tracked 408 urinary proteins detectable by SRM-MS. Interestingly, 169 of these were previously undetected in the datasets from Human Protein Atlas and a urinary proteome dataset from Adachi et al. (124). Furthermore, using SRM-MS assays in ovarian cancer and benign ovarian tumor patient plasma, the authors demonstrate the reproducible differential expression of a number of candidates. In another study, Cima and colleagues (240) focused their analyses on the N-glycoproteome of Pten-null mouse serum and prostate. Label-free comparative analysis of the Pten-null animals and age-matched wild-type mice revealed 111 and 12 candidates that were significantly differentially expressed in tissue and sera, respectively. Next, the authors utilized SRM-MS assays to reliably quantify the 39 protein orthologs (selected based on consistent quantification) in sera of prostate cancer patients and controls, and used the resulting profiles to build predictive regression models for diagnosis and grading of prostate cancer. In our own study (described in Chapter 3), I identified a number of putative prognostic
markers of prostate cancer and verified the differential expression of these in prostatic fluids using SRM-MS (244).

1.5 Rationale and objectives

1.5.1 Rationale

There exists a wide discrepancy between the number of men with prostate cancer and those dying from it. Although treatment efficacy has improved, the discrepancy is, at least in part, attributable to the detection of cancers that are non-life threatening within the patient’s lifetime (254). Considering the potentially serious hazards associated with prostate biopsies, and even more so with treatment procedures, the current modalities of diagnosis and prognosis are failing to accurately classify patients based on the significance of their disease and subjecting them to unnecessary risks. Establishing non-invasive molecular features capable of distinguishing insignificant prostate tumors from significant ones would greatly alleviate such burdens. In the current thesis, I have utilized proteomics methodologies to investigate the molecular factors present in prostatic fluids in order to delineate differential protein signatures of prostate cancer subtypes.

1.5.2 Hypotheses and aims

(1) Discovery proteomic exploration of EPS and identification of prostate-enriched proteins

I hypothesized that the EPS proteome is a rich source of prostatic proteins. The following aims were carried out to test the hypothesis:

a) Catalogue the direct-EPS and EPS-urine proteomes using discovery proteomics techniques;

b) Integrate the proteomics data with other available resources to characterize the EPS proteomes;

c) Comparatively analyze the proteomes of pre- and post-DRE urines for the identification of prostate-enriched proteins;
d) Verify the expression of a number of enriched proteins.

(2) Comparative proteomic analyses of organ-confined and extracapsular EPS fluids

I hypothesized that comparative analyses of the proteomes of EPS from patients with EC and OC prostate cancers would reveal putative prognostic biomarkers for prostate cancer. The following aims were carried out to test the hypothesis:

a) Characterize the proteomes of EPS derived from prostate cancer patients with EC and OC tumors;

b) Implement a data mining strategy to prioritize proteins based on their biomarker potential utilizing pathway analysis, publicly available datasets, and literature review;

c) Verify the differential expression of candidates using immunoassays;

d) Evaluate the application of targeted proteomics using SRM-MS in EPS-urine;

e) Develop SRM-MS assays and quantify candidates in EPS-urine.

(3) Systematic large-scale development of targeted proteomics assays for biomarker verification in prostate-proximal fluids

I hypothesized that protein signatures of prostate cancer could be generated and accurately quantified in EPS-urine using a multiplexed proteomics assay. The following aims were carried out to test the hypothesis:

a) Pinpoint putative protein biomarkers and their proteotypic peptides based on discovery proteomics data from Aim 2;

b) Design SRM-MS assays for peptides using synthetic peptide standards in a urine background;

c) Apply SRM-MS assays to a cohort of EPS-urines from patients with EC, OC prostate cancer, as well as noncancer individuals for relative quantification of candidates;
d) Develop a single SRM-MS assay capable of quantifying multiple biomarker candidates;

e) Apply SRM-MS assay to an independent cohort of EPS-urines for absolute quantification and verification;

f) Utilize bioinformatics and statistics to identify a protein signature to distinguish different patient groups.
Chapter 2
In-depth proteomic analyses of prostate proximal tissue fluids

This chapter has been reformatted from the following original publications:


2 In-depth proteomic analyses of prostate proximal fluids

2.1 Abstract

It is expected that clinically obtainable fluids that are proximal to organs contain a repertoire of secreted proteins and shed cells reflective of the physiological state of that tissue, and thus represent potential sources for biomarker discovery, investigation of tissue-specific biology, and assay development. The prostate gland secretes many proteins into a prostatic fluid that combines with seminal vesicle fluids to promote sperm activation and function. Proximal fluids of the prostate that can be collected clinically are seminal plasma and expressed prostatic secretions (EPS). In this chapter, I present two separate proteomic investigations into direct-EPS and EPS-urine. In the first study, multidimensional protein identification technology (MudPIT)-based proteomics was applied to direct-EPS obtained from nine men with prostate cancer, resulting in the confident identification of 916 unique proteins. Bioinformatics analyses using publicly available microarray data of 21 human tissues (Human Gene Atlas), the Human Protein
Atlas database, and other published proteomics data of shed/secreted proteins were performed to systematically characterize this comprehensive proteome.

In a second study, 1022 unique proteins in a heterogeneous cohort of 11 EPS-urines derived from biopsy-negative noncancer diagnoses, benign prostatic hyperplasia (BPH), and low grade prostate cancer were identified. Furthermore, MudPIT was used to generate and compare the differential proteome from a subset of pooled urines and EPS-urines (pre- and post-digital rectal examination urines, DRE, respectively) from noncancer and prostate cancer patients. The direct proteomic comparison of these highly controlled patient sample pools enabled the generation of a list of prostate-enriched proteins detectable in EPS-urine that is distinguishable from the complex urine protein background. A combinatorial analysis of both proteomics datasets and the systematic integration with publicly available proteomics data of related body fluids, human tissue transcriptomic information, and immunohistochemistry images from the Human Protein Atlas database allowed the demarcation a robust panel of 49 prostate-derived proteins in EPS-urine. Finally, the expression of seven of these proteins was validated using Western blotting, supporting the likelihood that they originate from the prostate.

2.2 Introduction

Proximal fluids are found adjacent to a given tissue or organ and contain a repertoire of secreted proteins and shed cells reflective of the physiological state of that tissue (255). Hence, tissue-proximal fluids are rapidly emerging as discovery sources of proteins, metabolites, and genetic biomarkers for cancers. Seminal plasma and EPS are prostate-proximal fluids that can be collected in the clinic. Seminal plasma consists of fluids originating not only from the prostate, but also from several other male accessory glands such as the seminal vesicles, epididymis, and Cowper’s gland. In aging men with prostatic disease, obtaining seminal fluids for diagnostic assays and screening purposes is highly dependent on patient age and disease status, and thus not an attractive assay medium.

The exocrine compartment of the prostate is composed of differentiated epithelial cells, that actively secrete proteins such as prostate specific antigen (PSA), prostatic acid phosphatase (PAP), and prostaglandins into the glandular lumen (256), and hence, these proteins can be accessed by the expulsion of prostatic fluids. Expressed prostatic secretions can be readily
obtained in the clinic by forced (or expressed) ejection of prostatic fluids into the urethra following prostate massage. For evaluation of some types of prostatitis, a vigorous massage is done to force fluid through the urethra to collect a few drops from the tip of the penis for subsequent use in microorganism cultures. If a more gentle massage is performed in the context of a DRE of the prostate, the EPS is then immediately collected in voided urine post-exam (EPS-urine). Currently, a commercial genetic assay for prostate cancer detection, based on the presence of a non-coding RNA (PCA3), uses post-DRE EPS-urines as a source of prostate-derived genetic material (257-259).

In the context of prostate cancer, undiluted EPS (direct-EPS) can be collected while the patient is under anesthesia, just prior to prostatectomy. In comparison to the clinical DRE massages, a more vigorous digital rectal massage can be done to force prostatic fluids into the urethra with subsequent collection from the penis. For prostate cancer biomarker discovery purposes, direct-EPS has several ideal clinical features beyond its nature as a proximal fluid. Since direct-EPS is collected just prior to prostatectomy, clinical information that led to the decision to perform a prostatectomy is readily available (255). Post-prostatectomy pathology staging and grading can also be obtained for comparison with the pre-surgical assessments. When linked with clinical outcome data, these direct-EPS fluids thus have the potential to be used for discovery of both diagnostic and prognostic biomarkers for management of prostate cancer.

A MudPIT-based proteomics approach was employed to obtain a detailed inventory of proteins present in direct-EPS and EPS-urine samples. By using a high mass accuracy Orbitrap mass spectrometer and rigorous identification criteria, a high quality account these fluids was obtained to a degree not previously reported. To select for proteins likely secreted from the prostate in the complex EPS-urine background, a highly controlled quantitative comparison of pooled samples of urine collected prior to DRE and EPS-urine was performed. Importantly, these pre- and post-samples were collected from the same patients, enabling direct quantitative intra-patient comparisons to identify proteins that are upregulated in the post-DRE samples. The combination of these data and their integration with publicly available proteomic (124, 260), microarray data from the BioGPS portal (http://biogps.org/) (261), and immunohistochemistry images from the Human Protein Atlas (HPA; http://www.proteinatlas.org/) (262, 263) allowed the generation of a panel of 49 proteins likely secreted by the prostate and detectable in EPS-urine. Additionally, the
entire proteomics data was deposited to the Tranche server (www.proteomecommons.com) for others to utilize.

2.3 Results

2.3.1 In-depth proteomic analyses of direct expressed prostatic secretions

2.3.1.1 MudPIT profiling of direct expressed prostatic secretions from prostate cancer patients

Nine individual direct-EPS samples from men with prostate cancer (Gleason 6-7, stages pT1c-pT2b) were selected for MudPIT-based proteomic analysis (Table 2.1). To minimize variability introduced by sample handling, direct-EPS samples were directly digested in-solution and analyzed by a 9-step MudPIT on a LTQ-Orbitrap XL mass spectrometer.

Table 2.1 Clinical information of the direct-EPS fluids analyzed by MudPIT proteomics. a Serum PSA values, ng/ml. b Gleason score determined by post-prostatectomy pathology. c Determined by post-prostatectomy pathology. d Determined pre-prostatectomy by transurethral ultrasound measurements.

<table>
<thead>
<tr>
<th></th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>64</td>
<td>53</td>
<td>67</td>
<td>65</td>
<td>70</td>
<td>50</td>
<td>61</td>
<td>69</td>
<td>59</td>
</tr>
<tr>
<td>PSAa</td>
<td>37.5</td>
<td>6.7</td>
<td>7.3</td>
<td>4.8</td>
<td>5.1</td>
<td>4.1</td>
<td>3.5</td>
<td>8.7</td>
<td>6.2</td>
</tr>
<tr>
<td><strong>Pathological stage</strong></td>
<td>T2b</td>
<td>T1c</td>
<td>T1c</td>
<td>T1c</td>
<td>T1c</td>
<td>T1c</td>
<td>T1c</td>
<td>T1c</td>
<td>T1c</td>
</tr>
<tr>
<td><strong>Gleasonb</strong></td>
<td>4+3</td>
<td>3+4</td>
<td>3+4</td>
<td>3+3</td>
<td>3+4</td>
<td>3+3</td>
<td>3+4</td>
<td>3+4</td>
<td>3+3</td>
</tr>
<tr>
<td><strong>Percent tumorc</strong></td>
<td>75</td>
<td>25</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>20</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td><strong>Prostate volumed</strong></td>
<td>53</td>
<td>41</td>
<td>56</td>
<td>31</td>
<td>96</td>
<td>48</td>
<td>70</td>
<td>42</td>
<td>73</td>
</tr>
</tbody>
</table>

Rigorous protein identification criteria resulted in the identification of 916 unique proteins with high confidence (Figure 2.1A and B), with 230 to 550 unique proteins identified per individual cancer patient sample (Figure 2.1B). The list of 916 proteins (Appendix Table 2.1) obtained from
the direct-EPS cancer samples were compared to those reported in a recent publication by Li and colleagues (264) who collected several EPS fluids using a more vigorous prostate massage procedure (Figure 2.1C). Only twenty proteins were unique to the study of Li et al. (264), consisting primarily of blood-derived proteins. The vast majority of proteins of the 916 proteins were uniquely identified in the current study with 816 unique protein clusters and 94 protein clusters that were shared with Li et al. Based on the very low levels of hemoglobin and seminogelin proteins detected in the analysis, it can be concluded that direct-EPS samples have minimal contamination with blood and seminal vesicle derived proteins. In relation to known prostate derived proteins, the next most abundant proteins accounting for an additional 25% total abundance were lactoferrin, PSA, PAP, zinc-α2 glycoprotein and aminopeptidase N. Functional annotation of the direct-EPS proteome using Gene Ontology, predicted transmembrane domains (TMD; TMAP algorithm) and signaling peptides (SP; PrediSi algorithm) revealed a wide variety of functional categories (Figure 2.2A-C). A significant number of detected proteins were membrane proteins and members of the extracellular region, including 34 CD molecules (Appendix Table 2.1). This was further supported by the fact that ~ 40% of the detected proteins had at least one predicted TMD and contained a predicted SP, indicative of targeting to the secretory pathway (Figure 2.2D and E). A smaller number of proteins could be assigned to other intracellular organelles.
Figure 2.1 Proteomics analyses of direct expressed prostatic secretions. (A) EPS from nine men with confirmed prostate cancer were in solution digested and individually analyzed by online MudPIT on an LTQ-Orbitrap XL mass spectrometer. Rigorous peptide and protein identification criteria resulted in the identification of 916 unique proteins. (B) Number of unique proteins detected in each of the analyzed EPS samples. (C) Comparison of our EPS proteome to a previously published EPS dataset.
Figure 2.2 Functional enrichment of proteins detected in EPS. (A-C) Shown are selected Gene Ontology (GO-Slim) categories of the EPS proteome. (D) Distribution of predicted transmembrane domain containing proteins (TMAP algorithm). (E) Distribution of predicted signal peptides (PrediSi algorithm).
2.3.2 Integrative data mining and comparison to other public databases

One goal of the current study was to provide the research community with a high quality, well-annotated resource of proteins expressed in EPS. This could improve our understanding of general prostate biology and guide the discovery of novel prostate cancer biomarkers. To accomplish this goal, the data was mapped to two publicly available resources: the Human Gene Atlas (265), which is based on the mRNA microarray expression profile of >50 human tissues, and the Human Protein Atlas (263, 266), a large resource of immunohistochemistry images and available human antibodies (Figure 2.3). The rationale for these comparisons was to identify proteins detected in EPS fluids that are also enriched in prostate tissue relative to other human tissues included in the microarray database. Also, integration of EPS proteins with the Human Protein Atlas database can catalogue the availability of validated antibodies, which could provide researchers with an important tool to further study the biological function of selected proteins.

Using only mRNAs with ≥3-fold above median expression in prostate tissue compared to 20 other human tissue types in the Human Gene Atlas, 279 direct-EPS gene products were identified, implying a potential function specifically related to prostate biology. In the Human Protein Atlas, a total of 513 proteins detected in the direct-EPS proteome currently have antibodies with varying staining intensities in human prostate tissue. The identities of the EPS proteins mapped to the two public databases are available in Appendix Tables 2.2 and 2.3.
Figure 2.3 Integrative data mining and comparison to other databases. (A) The EPS proteome was mapped against the Human Gene Atlas (mRNA microarray) and Human Protein Atlas (antibodies and immunohistochemistry images). (B) 279 unique gene products detected in our EPS proteome showed ≥3-fold above median expression in normal prostate tissue on the the Human Gene Atlas microarray (detected by 317 probe IDs), as compared to 20 other tissues.
Mapping against available Human Protein Atlas (HPA) antibodies and CD molecules is provided.

2.3.2.1 Comparison of the direct-EPS proteome to other published proteomic datasets

Next, the direct-EPS proteome was integrated with three other published proteomic datasets (124, 267, 268). This included a recent publication of proteins secreted into the cell culture media by three established prostate cancer cell lines (PC3, LNCaP, and 22Rv1) (268) and two human body fluid datasets (urine and seminal plasma) directly related to EPS (267, 268). The rationale for the comparison to the cell secretions was to distinguish between proteins readily secreted by prostate cancer cells from proteins being secreted by other cell types of the prostate. This is not an ideal comparison, however, as the cell lines are more reflective of metastatic prostate cancers and the direct-EPS samples were from lower grade Gleason 6 and 7 cancers. An additional caveat is that these cells have numerous genomic alterations and have been grown under culture conditions for many passages distant from the primary isolates, and might not truly reflect an in vivo setting. Approximately 300 proteins (Figure 2.4) were shared between the EPS proteome and each of the three prostate cancer cell lines (see Appendix Tables 2.4-2.6 for detailed mappings). Likewise, approximately 500 proteins were shared between the EPS proteome and normal human urine and seminal plasma (see Appendix Tables 2.7 and 2.8). Several well-known prostate markers (i.e. PSA, PAP) were shared among the direct-EPS proteome and the described prostate cancer cell lines. The data provided in this study could therefore be used to mechanistically investigate proteins identified in vivo (EPS) using well-established prostate cancer cell lines. Interestingly, several well known prostate markers (i.e. MSMB, TMPRSS2) were only identified in the EPS proteome and the other two available body fluids (see below), strengthening the rationale of using in vivo obtainable fluids to study prostate/prostate cancer biology. It is likely that proteins exclusively detected in the EPS proteome, as compared to proteins detected in healthy human urine and seminal plasma, are specifically enriched in secretions of the prostate under pathological conditions, and were therefore absent or at a very low concentration, in the other body fluids.
Comparison to other published proteomes. The EPS proteome was compared to proteins detected in two other published, related body fluid proteomes Pilch et al. seminal plasma and Adachi et al. urine) and proteins secreted into the condition medium by three established prostate cancer cell lines (Sardana et al.).

### 2.3.2.2 Multi-dataset integration

To fully utilize the power of multi dataset comparison, all publicly available data resources were integrated with the current direct-EPS proteome (Figure 2.5), as an example of a potential data mining strategy. Gene products strongly enriched in normal prostate tissue using the described microarray dataset (10-fold above median expression) that were also detected in the direct-EPS proteome and had antibodies through the Human Protein Atlas database were focused on. This integration resulted in 54 unique gene products satisfying all described integration criteria.
(Figure 2.5 and Appendix Table 2.9). A heat map display of this integration is presented in Figure 2.5, including available mappings to other public proteome datasets. Although all gene products are strongly expressed in prostate tissue, most showed additional expression in several other tissues. Known prostate markers (vertical line), such as PSA, PAP, and microseminoprotein-β (MSMB) were readily identified. Comparison to other published proteomics datasets is also indicated in the presented heat map, and this data could be used to further guide potential follow up analyses. As all 54 proteins displayed in the heat map have available immunohistochemistry images at the Human Protein Atlas, the available images were systematically screened for both normal prostate and prostate cancer tissues. Several proteins with intriguing staining patterns are discussed below, demonstrating an example of how the current data could be used by others.
Figure 2.5 Integration of multiple data resources with the EPS proteome. Heat map of 54 unique gene products detected in the EPS proteome with available antibodies in the Human
Proteome Atlas database. Selected gene products were enriched in normal prostate tissue as compared to 20 other human tissues using the Human Gene Atlas microarray database (≥10-fold above median expression in prostate). Mappings to other available proteome datasets are indicated. Well known prostate markers are highlighted by a vertical line.

Several proteins were identified which displayed a trend towards differential expression in normal prostate compared to prostate cancer tissue using available IHC images, although patient variability and image quality differences were also observed (Figure 2.6). Neprilysin (MME, CD10) showed a trend towards decreased expression in prostate cancer tissue. Alternatively, GDF15 and FASN showed a trend towards increased expression in prostate cancer tissues (Figure 2.6). As the entire dataset has been linked to available mRNA microarray expression, Human Protein Atlas antibody staining, and other published, related proteomics datasets, this data will likely provide an important resource for the prostate biology/cancer community to guide the investigation of other proteins in hypothesis driven experiments.
Figure 2.6 Immunohistochemistry images as obtained from the Human Protein Atlas. Selected examples of proteins found in the Human Protein Atlas database that showed trends towards differential expression in normal compared to cancer prostate tissues.

2.3.3 Identification of prostate-enriched proteins by in-depth proteomic analyses of expressed prostatic secretions in urine

2.3.3.1 MudPIT profiling of expressed prostatic secretions in urine

The first aim of this study was to provide the first in-depth proteome catalogue of this clinically useful proximal tissue fluid (Figure 2.7, left panels). In particular, individual EPS-urine samples from men diagnosed with low grade prostate cancer (n = 5; Gleason score total 6-7) and biopsy negative benign conditions (n = 6, BPH) were selected for MudPIT-based proteomic analyses (Table 2.2).
Table 2.2 Clinical information for the prostate cancer (5 samples) and benign prostatic hyperplasia (6 samples) EPS-urines analyzed by MudPIT. Serum PSA values (reported in ng/ml) are from the time of initial diagnosis. The treatment column indicates the clinical course followed for each cancer patient (DVP – DaVinci prostatectomy; Cryo, cryoablation therapy; AS, active surveillance.

<table>
<thead>
<tr>
<th>Sample #</th>
<th>Age</th>
<th>Serum PSA</th>
<th>Gleason</th>
<th>Pathological stage</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPS-U1</td>
<td>59</td>
<td>9.2</td>
<td>3+3</td>
<td>T1c</td>
<td>DVP</td>
</tr>
<tr>
<td>EPS-U2</td>
<td>66</td>
<td>1.9</td>
<td>3+3</td>
<td>T1c</td>
<td>AS</td>
</tr>
<tr>
<td>EPS-U3</td>
<td>77</td>
<td>2.7</td>
<td>3+4</td>
<td>T2b</td>
<td>Cryo</td>
</tr>
<tr>
<td>EPS-U4</td>
<td>73</td>
<td>5.9</td>
<td>3+3</td>
<td>T1c</td>
<td>DVP</td>
</tr>
<tr>
<td>EPS-U5</td>
<td>53</td>
<td>6.0</td>
<td>3+3</td>
<td>T1c</td>
<td>AS</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>65.5 ± 9.8</td>
<td>5.1 ± 2.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample #</th>
<th>Age</th>
<th>Serum PSA</th>
<th>Gleason</th>
<th>Pathological stage</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPS-U6</td>
<td>76</td>
<td>4.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPS-U7</td>
<td>63</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPS-U8</td>
<td>57</td>
<td>3.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPS-U9</td>
<td>74</td>
<td>2.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPS-U10</td>
<td>54</td>
<td>4.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPS-U11</td>
<td>61</td>
<td>4.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>64.2 ± 8.9</td>
<td>3.9 ± 0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The proteomic characterization of the EPS-urine was performed on this heterogeneous group of 11 samples as a representation of the different types of patient origin, reflective of the most common benign and prostate cancer conditions presenting in urology clinics. For each sample, 150 µg of total protein was directly digested in solution and analyzed in triplicate by a 9-step MudPIT on a LTQ-Ion Trap mass spectrometer, as previously described (269, 270). A total of 1022 unique proteins were identified in the EPS-urines (Appendix Table 2.10) by at least two unique peptides (Figure 2.7, left panels), ranging from 178 to 667 unique proteins determined per individual MudPIT run (Figure 2.8A). Although the same amount of total protein was digested and each sample was analyzed in triplicate with a relatively low overall standard deviation (average standard deviation of 27), there was still large variation in total proteins detected for each sample. This high variability is reflective of the biological inter-sample variations among human specimens, which complicates data analyses. Furthermore, dynamic metabolic changes within each individual are manifested in urine protein content (271), and highlight some of the
general problems of proteomic analyses of proximal body fluids (272). These results could also reflect variation in sample collection, as the DRE procedure required to “express” the prostatic fluids will be different for each individual (255). Other known limitations are related to the DRE collection procedure and can be attributed to the physician (e.g. size of hand, ability to reach to the prostate, etc.) or to some patient physical parameters (e.g. orientation on table during examination, overweight status, etc.). Therefore, standard collection protocols as well as internal standards are required to ensure proper collection and to circumvent the introduction of sample variability resulting from the collection procedure (255).
Figure 2.7 Study workflow. Proteomic analysis of EPS-urine and urine samples. The EPS-urine proteome was defined by MudPIT analyses of 11 individual heterogeneous samples (prostate cancer, PCa and benign prostatic hyperplasia, BPH) of EPS-urine. A similar analysis was performed on pooled urine and EPS-urine samples (PCa and noncancer, NC). The comparison of the two datasets using bioinformatics data mining lead to the identification of some putative prostate-enriched candidates within the complex EPS-urine proteome.
Figure 2.8 Characterization of the EPS-urine proteome. Number of unique proteins detected per MudPIT analyses in the (A) individual EPS-urine samples and in the (B) pooled urine and EPS-urines. The black dots in (A) (from EPS-U1 to EPS-U5) represent the EPS-urines from individual prostate cancer patients and the white dots in (from EPS-U6 to EPS-U11) represent the EPS-urines from individual benign prostatic hyperplasia patients. Every sample was analyzed in triplicate. (EPS-U: EPS-urine; U: urine; PCa: prostate cancer; NC: noncancer).

To obtain a systematic overview of the functional categories of proteins expressed in EPS-urine, Gene Ontology (GO) (273) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analyses (274) on the 1022 identified proteins (Figure 2.9) were performed. The comparison was limited to several representative GO terms in order to obtain a high-level functional overview. In the biological processes category, a large proportion of identified proteins had functional roles
involved in proteolytic activity, cellular adhesion and motion, and immune responses (Figure 2.9A). In the cellular component category (Figure 2.9B), GO terms with an annotation to the extracellular region were overrepresented; this was further supported by the evidence that ~ 50% of the detected proteins were classified in the SwissProt database as proteins with signal peptide sequences, and more than 30% were classified as secreted proteins (Figure 2.9E). Notably, 12% of the EPS-urine proteins contained the annotation of vesicle localization by GO (Figure 2.9B); this could explain why the lysosomal pathway is the most highly over-represented among the KEGG pathways in the EPS-urine dataset (Figure 2.9D). In the molecular function category, a large proportion of detected proteins had calcium ion binding activity and were involved in the regulation of peptidase activity (Figure 2.9C). Based on the global GO term analyses, EPS-urine proteins belong to a large variety of functional categories and, as expected from a secreted fluid fraction, a large number of these proteins are extracellular and contain a signal peptide sequence.

To further characterize the EPS-urine proteome in the context of other prostate-related fluids, the current dataset was integrated with normal human urine (124) and direct-EPS (260), which showed that 455 proteins are shared between these fluids by proteomics analyses (Figure 2.9F). Present among these shared proteins were several previously proposed prostate cancer biomarkers (i.e. PSA, PAP, MSMB, PSMA, TMPRSS2), strengthening the rationale of using in vivo obtainable fluids to study the prostate and prostate cancer biology. On the other hand, these candidates were detected in all three fluids, which demonstrates the dilemma in detecting prostate-enriched proteins in a complex protein background of general urine proteins.
Figure 2.9 Functional enrichment of proteins detected in EPS-urine and comparison of the EPS-urine proteome to other related body fluids. (A-E) The list of the 1022 unique proteins identified in the EPS-urine samples was compared to the human UniProt Database. The reported graphs show significantly over-represented ($p < 0.001$) Gene Ontology (GO) terms (A-C), KEGG pathways (D) and SwissProt entries (E) in the EPS-urine dataset (several representative annotation terms are shown). (F) The current EPS-urine dataset was compared to previously published urine (124) and direct-EPS (from Section 2.3.1) (260) proteomic datasets. Proteins are clustered in homogeneous groups based on 98% similarity (i.e. cluster anchors).
Focusing on prostatic secretions, 181 proteins were shared between EPS-urine and direct-EPS (Figure 2.9F), suggesting that these proteins are specifically enriched in prostatic secretions and absent or at a very low concentration in urine. This could represent a useful protein dataset for selecting new prostate specific biomarkers with diagnostic and prognostic capacities for prostatic diseases. Interestingly, 113 proteins were unique to EPS-urine (absent from direct-EPS and urine). This could be explained by different mass spectrometry (MS) conditions used between our group and the study by Adachi et al (124), as well the phenomenon of random sampling (202) and biological variability, since neither study was likely successful in detecting the entire body fluid proteome.

This study provides a detailed description of EPS-urine and extensively expands our knowledge of the EPS-urine proteome. The dataset of 1022 proteins (Appendix Table 2.10) can be used to implement diagnostic test platforms and improve current screening procedures for prostatic diseases.

2.3.3.2 Identification of prostate-enriched proteins

A second aim of this study was to provide the first direct comparison of urine samples pre- and post-DRE using pooled samples from both normal and prostate cancer patients (Figure 2.7, right panels). These valuable samples enabled the identification of proteins likely released by the prostate as a result of the DRE. For this purpose, five patients with prostate cancer and five noncancer individuals were screened once prior to DRE and once after DRE, in order to obtain internally controlled urines and EPS-urines, respectively (Table 2.3). Each sample pool was analyzed by a 5-step MudPIT on a LTQ-Orbitrap XL, leading to the identification of 444 unique proteins (Figure 2.7, right panels and Appendix Table 2.10), with a range of 141 to 258 unique proteins determined per individual MudPIT run (Figure 2.8B).
Table 2.3 Clinical information for the pooled urine and EPS-urine samples analyzed by MudPIT. Urine and EPS-urine sample pairs were collected from the same patient and pooled as described in Materials and Methods (serum PSA reported in ng/ml).

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Serum PSA</th>
<th>Gleason</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Noncancer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>3.8</td>
<td>BPH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>10.8</td>
<td>BPH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>5.9</td>
<td>BPH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>3.3</td>
<td>Normal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>6.4</td>
<td>Normal</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean ± SD</strong></td>
<td>65.0 ± 10.0</td>
<td>6.0 ± 2.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Prostate cancer</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>2.6</td>
<td>3+3</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>11.7</td>
<td>3+3</td>
<td>Intermediate</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>2.3</td>
<td>3+4</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>6.4</td>
<td>3+3</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>5.7</td>
<td>3+3</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td><strong>Mean ± SD</strong></td>
<td>72.8 ± 8.6</td>
<td>5.7 ± 3.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In order to highlight signatures of proteins enriched in the EPS-urine, a semi-quantitative comparative analysis of the EPS-urine and urine data was performed based on the QSpec algorithm (275). Proteins were considered to be prostate-enriched if they had a ≥2-fold change in spectral abundance factors (202, 275) with a false discovery rate ≤0.05 by QSpec analysis. Applying these criteria to the entire list of 444 proteins, a panel of 49 significantly enriched proteins was generated (Appendix Table 2.11). Interestingly, direct comparison of both proteomic datasets (1,022 versus 444) showed an overlap of 406 proteins that included the 49 protein signature of “prostate-enriched proteins” (Figure 2.10).
Figure 2.10 Comparison between the two proteomics data of the current study. The Venn diagram shows a comparison between the EPS-urine dataset (1022 proteins) and the dataset containing proteins identified in both pooled urine and EPS-urine samples (444 proteins). The overlapping area of 406 proteins includes a short list of proteins enriched in the EPS-urine samples compared to the urine samples (termed: prostate-enriched proteins).

The prostate tissue selectivity of the 49 identified protein signature was also evaluated at the transcriptome level (Figure 2.11A). The proteins were mapped in the BioGPS database, and the gene expression profiles for normal human prostate tissue were compared to 24 other major normal human tissues, under the assumption that transcript levels measured by microarrays can be correlated with protein abundance (270). Although it is well known that expression levels of both biomolecules are not always concordant (269, 276, 277), our own experience suggests reasonable correlation if the data is locked on the protein level (i.e. using only transcripts with available proteins evidence) (269). This type of comparison can provide additional evidence for prostate-enriched expression levels, similar to previous investigations (260). Genes with prostate expression levels 2-fold above the median across all 25 tissues were considered to be prostate-enriched. The distribution of this analysis is shown in Figure 2.11A and represents the expression level of each gene coding for the 49 selected proteins in prostate tissue. Among the 49 selected gene products identified several well-known prostate biomarkers, such as PSA (KLK3), MSMB
and PAP (ACPP) (highlighted in green), supporting the hypothesis that direct comparison of urine and EPS-urine by semi-quantitative proteomics results in the identification of prostate-enriched proteins suitable for further investigation in biomarker studies. According to this analysis, 32 of the 49 informative genes are enriched in the prostatic tissue. The genes indicated with red bars in the graph correspond to the seven proteins for which antibodies were obtained and used to verify the MS-based proteomics data.

Next, a random sampling of the dataset was examined to address the question of chance selection in arriving at the 49-protein candidate list. Proteins were randomly selected 10,000 times from the entire list of 1022 proteins detected in EPS-urine in order to evaluate if a random selection from a larger dataset of EPS-urine proteins yielded the same probability of selecting prostate-enriched proteins. The boxplot shown in Figure 2.11B reports the comparison between the two datasets. As expected, the shortlist of 49 selected proteins is enriched in prostate-specific proteins when compared with a random assortment of 10,000 selections from the same EPS-urine database. The median values of the two distributions tend to be lower in the random sampling analysis (median = 0.6) compared to the selection (median = 2.2) with a highly significant P-value ≤0.001.
Figure 2.11 Data mining to identify prostate-enriched proteins. (A) Comparison of the 49 prostate-enriched proteins to published human tissue transcriptomic data from the BioGPS gene expression portal (261). On the X-axis gene names are shown and the Y-axis represents a scale indicating the gene expression level in prostate as compared to 24 additional human tissues. The Y-values were obtained as log$_2$ [fold change] ratio of prostate gene expression level versus the median value calculated for 25 major organs (listed in Materials and Methods). Green bars highlight protein biomarkers already suggested as candidates for prostate cancer prognosis and diagnosis (used as internal positive controls). Red bars highlight the 7 proteins that have been validated in this study. The red line demarcates the 2-fold increased expression level in normal prostate as compared to all other tissues. (B) Box plot shows prostate-enriched proteins in the 49 protein candidate list versus a random sampling of the entire 1022 EPS-urine protein dataset. The Y-axis has the same log$_2$ scale reported in (A), rectangles are bounded by the lower and upper quartiles, the solid lines in the rectangles are the medians, the box whiskers extend to the minimum to the maximum data point of the rectangle, and the circles represent outliers beyond this range. *** p-value < 0.001.

2.3.3.3 Characterization of prostate-enriched proteins

To better characterize the list of prostate-enriched proteins, information available at the Human Protein Atlas (262, 263) and UniProt databases (278) were used. As shown in Figure 2.12A, 38 proteins (78%) of all 49 have at least one antibody ID available at the Human Protein Atlas (262, 263) (Appendix Table 2.11), which was validated and used to generate tissue expression profiles. In Figure 2.12A, the protein expression patterns based on the immunohistochemistry images available in the database are indicated. The evaluation of these images, based on staining intensity and protein distribution in the human normal prostate tissue, suggested that 45% of the proteins positivity was found in the glandular epithelium, only 13% in the stromal cells and 29% showed equal staining between the two compartments.
Figure 2.12 Additional annotations for prostate-enriched proteins. (A) Availability of antibodies reported in the Human Protein Atlas database, for the 49 selected proteins. The immunohistochemistry images were manually screened and protein distribution was annotated (glandular cells or stroma). (B) Subcellular location of the 49 prostate-enriched proteins according to the UniProt database.

Together, protein tissue distribution and cellular localization are parameters that can provide important insights into the function of a protein. As expected, the respective subcellular patterns reported from the UniProt database (278) were correlated (Figure 2.12B) with additional information collected from other available resources (i.e. ProteinCenter, signal peptide annotations) (Appendix Table 2.11). The majority (33%) of the proteins in the analyzed EPS-urine samples were secreted (without overlap with the other categories) (Figure 2.12B). This localization is consistent with the biological expectation that urine contains, by definition, many extracellular proteins. Another significant percentage of proteins (22.5%) were localized to the...
membrane compartment. Interestingly, although the EPS-urine is not enriched in intracellular proteins (18.4%), almost half (44%) of this category consist of lysosomal proteins. This probably points out that exosome formation is the dominant excretion pathway in urine and reflects the biological and physiological role of these proximal fluids, through the presence of specific transport pathways for lysosomal proteins.

2.3.3.4 Generation of candidate shortlist

The analyses suggest an enrichment of prostate-specific proteins in the EPS-urines compared to the urine samples. In particular, the list of 49 proteins (Appendix Table 2.11) encompasses a number of prostate specific markers that are currently used or have been previously considered as potential candidates for prostate cancer screening; among these are: MSMB (279), PSA (280, 281), PAP (reported in 1938 as the first serum biomarker for prostate cancer (282, 283), TMPRSS2 (284), and PSMA (285).

A small verification set, aimed at validating the findings from the proteomic investigations, was selected out of the 49 differentially expressed proteins. Selection was largely based on antibody availability; but other unbiased considerations were tissue expression patterns of candidates across prostate cancer and normal prostate tissues available through the Human Protein Atlas (262, 263), tissue specificity based on mRNA microarray data (see above), subcellular localization by GO analysis, and in-house proteomic datasets generated from urines and prostatic fluids from various conditions. This led to a shortlist of 7 proteins that were assayed in the verification stage.

2.3.3.5 Verification of proteomic data

In order to verify the proteomic data, Western blot analyses were carried out for each of the 7 candidates in pooled EPS-urines and urines from prostate cancer and noncancer patients. The panel in Figure 2.13A shows that each of the shortlisted candidates are more abundant in EPS-urine samples compared to the matching urines, supporting results from the MudPIT analyses. In particular, comparing noncancer and prostate cancer EPS-urines showed the downregulation of TGM4, LTF, ANPEP, MME and TIMP1, and upregulation of PARK7 and 14-3-3σ (Figure 2.13A), which correlated with the proteomic data analyses. These preliminary results also follow a similar trend when looking at immunohistochemical staining patterns in normal and neoplastic
prostatic tissues via Human Protein Atlas (262, 263), but will require further verification in large unrelated EPS-urine cohorts in the future (Figure 2.13B).

**Figure 2.13 Validation of mass spectrometry data for seven selected proteins.** (A) Western blot analyses of the seven candidate proteins confirmed the proteomic data (based on SpC) showing an enrichment of each selected protein in the EPS-urine compared to the urine samples. (B) Immunohistochemistry images obtained from Human Protein Atlas database showing the differential expression in normal and cancerous prostate glandular tissue of the seven selected proteins. The same trend was obtained by MS and Western blot analysis. (U_NC:

2.4 Discussion

The cumulative list of identified proteins in the EPS samples currently represents the largest compilation resource of proteins present in prostatic fluid secretions from men with confirmed prostate cancers. In comparison to other proteomics studies aimed at discovering prostate function or prostate cancer biomarkers using cell secretions (268), plasma (250, 286), or tissue (287, 288), the studies presented in this chapter used a relevant organ-proximal body fluid for the detection of secreted proteins in vivo. One goal of the studies presented in this chapter was to provide a high-quality, well-annotated resource of proteins present in EPS. This could improve our understanding of general prostate biology and guide the discovery of novel prostate cancer biomarkers. While commercial genetic assays have been developed using EPS-urines as a source of prostate-derived genetic material (257, 289, 290), an in-depth proteomic analysis of this fluid has not been reported. In contrast to direct-EPS (260), which is likely to contain prostate-secreted proteins at a higher concentration, EPS-urine is highly diluted by a dynamic and variable urine background. Nevertheless, EPS-urine samples are clinically more relevant since they can be obtained by routine DRE examination, and can be collected repeatedly for longitudinal sample collection relevant to active surveillance monitoring of men with prostate cancer. By directly comparing pre- and post-DRE samples, a number of proteins that are enriched in EPS-urines compared pre-DRE urines were identified.

Current early screening for prostate cancer relies on PSA detection in serum (291). Over time, this screening has dramatically reduced the presentation of high grade disease, yet other studies have indicated that the majority of early detected prostate cancers are not lethal and may be considered clinically insignificant (292-294). Thus, current methods to establish the risk of progression and prognosis of prostate cancers remain suboptimal, and a large number of patients are over treated with a significant negative impacts on patients and on health care (292, 293). Thus, new diagnostics are needed that could predict disease course and discriminate indolent tumors from aggressive tumors, allowing for more appropriate treatment options to be offered to newly diagnosed prostate cancer patients. Targeting identification of potential new biomarkers in
proximal fluids of the prostate like direct-EPS and post-DRE EPS-urines could address these clinical needs.

With the availability of modern mass spectrometers and high performance peptide separations like MudPIT (203, 205), we are now able to routinely identify hundreds to thousands of proteins in complex biological samples. This can result in a significant challenge for the standard biomarker validation strategy, ELISA, which is time-consuming and costly to develop. However, MS-based assays, such as selected reaction monitoring mass spectrometry (SRM-MS), can facilitate the rapid verification of candidates prior to ELISA development. An additional challenge is the selection of the most likely candidate biomarkers from the large list of identified proteins. To guide the selection criteria of promising candidate proteins involved in prostate function and prostate cancer biomarker discovery, the EPS proteome was integrated with a large variety of publicly available datasets. To further streamline SRM-MS assay development, the entire MS data has been deposited to the Tranche database for others to mine, as previously reported (269, 295, 296). Cumulatively, this EPS proteome dataset should facilitate the selection and confirmation of individual biomarker candidates in follow up studies.

2.5 Materials and methods

2.5.1 Materials

Ultrapure-grade urea, ammonium acetate, ammonium bicarbonate, Tris, and calcium chloride were from BioShop Canada Inc. (Burlington, ON, Canada). Ultrapure-grade iodoacetamide, DTT and formic acid were obtained from Sigma-Aldrich. HPLC-grade solvents (methanol, acetonitrile and water) were obtained from Thermo Fisher Scientific (San Jose, CA). Trifluoroacetic acid was from J.T. Baker (Phillipsburg, NJ, USA). Mass spectrometry-grade trypsin gold was from Promega (Madison, WI, USA). Solid-phase extraction C18 MacroSpin™ Columns were from The Nest Group, Inc. (Southboro, MA, USA).

2.5.2 Sample collection and properties

All samples were collected from patients following informed consent and use of Institutional Review Board approved protocols at Urology of Virginia, Sentara Medical Group and Eastern Virginia Medical School between 2007 and 2009, along with the Research Ethics Board of the
University Health Network (Table 2.1). Tumor grades, staging, prostate volumes and tumor volumes (297) were determined using standard pathological procedures. Patient information was recorded, including demographics, medical history, pathology results, and risk factors, and stored in a Caisis database system. All personal information or identifiers beyond diagnosis and lab results were not available to the laboratory investigators.

Direct-EPS fluids were obtained by vigorous prostate massage from subjects under anaesthesia just prior to prostatectomy as a result of a prostate cancer diagnosis. Each sample (0.2-1 ml) was collected from the tip of the penis in a 10 ml tube, diluted with saline to 5 ml, and stored on ice until transport to the biorepository for processing (within 1 hour of collection). For processing, any particulate material was removed by low-speed centrifugation. Aliquots (1 ml) of the supernatant were stored at -80°C.

EPS-urine samples were collected by performing a gentle massage of the prostate gland during DRE prior to biopsy, as previously described (260). The massage consisted of three strokes on each side of the median sulcus of the prostate and the expressed fluid from the glandular network of the prostate was subsequently voided in urine.

To generate sample pools from noncancer and cancer patients, 10-20 ml of urine and EPS-urine were collected from the same individual an hour before the DRE massage, herein denoted as urine, and after DRE, herein denoted as EPS-urine. Samples were pooled in order to generate sufficient material for analyses. Urine and EPS-urine from a group of 5 patients with prostate cancer and 5 biopsy negative, noncancer individuals (Table 2.3), were pooled together to generate a sample panel comprising 4 different conditions: U_NC: urine noncancer; EPS-U_NC: EPS-urine noncancer; U_Ca: urine cancer; EPS-U_Ca: EPS-urine cancer. After collection, samples were stored on ice for no longer than 1 hour. Each sample was aliquoted and stored at -80°C until use.

Individual EPS-urines were obtained from an independent cohort of 11 different patients: 5 with low-grade prostate cancer and 6 with biopsy negative benign conditions (BPH) (Table 2.2). Following collection, 9 ml of EPS-urine was centrifuged at 14,000 g to remove the cell pellet/sediment. The supernatant was recovered and concentrated using an Amicon Ultra-15 Centrifugal Filter (3 kDa cutoff; Millipore, Billerica, MA, USA) according to the manufacturer’s
instructions. Approximately 500 μl of each concentrated EPS-urine sample was recovered from the filter device and stored at −80°C until use.

2.5.3 Protein digestion and sample preparation

Direct-EPS fluid corresponding to 100 μg of total protein (as measured by Bradford assay) were concentrated to ~50 μl using an Amicon Ultra spin filter (3kDa cut-off; Millipore) followed by digestion as stated below. Briefly, concentrated EPS corresponding to 100 μg of total protein was diluted with an equal volume of 8M urea, 4mM DTT, 50mM Tris-HCl, pH 8.5 and incubated at 37°C for 1 hour, followed by carbamidomethylation with 10mM iodoacetamide for 1 hour at 37°C in the dark. Samples were diluted with 50mM ammonium bicarbonate, pH 8.5 to ~1.5M urea. Calcium chloride was added to a final concentration of 1mM and the protein mixture digested with a 1:30 molar ratio of recombinant, proteomics-grade trypsin at 37°C overnight. The resulting peptide mixtures were solid phase-extracted with C18 MacroSpin Columns from The Nest Group, Inc. (Southboro, MA, USA). Samples were concentrated by speedvac and stored at -80°C until used for MudPIT analysis.

Urine and EPS-urine were quantified using a NanoDrop™ 2000 spectrophotometer (Thermo Fisher Scientific, San Jose, CA) and volumes corresponding to 100 μg of total protein for the pools and 150 μg of total protein for the individual EPS-urines were resuspended in 50 μl of 8 M urea, 2mM DTT, 50 mM Tris-HCl, pH 8.5, and incubated at 37 °C with constant shaking for 30 min. Carbamidomethylation was performed by incubating samples with 8 mM of iodoacetamide for 30 min at 37 °C in the dark. Samples were then diluted to approximately 1.5 M urea using 100 mM ammonium bicarbonate, pH 8.5. Calcium chloride was added to a final concentration of 2 mM and the protein mixture was digested with trypsin (1:40 trypsin to protein ratio) at 37 °C overnight. The digested peptide mixture was purified with C18 MacroSpin™ columns and concentrated by vacuum centrifugation and reconstituted to a volume of 40 μl with 0.1 % formic acid. Samples were stored at -80 °C until used for MudPIT analysis.

2.5.4 MudPIT analyses

Direct-EPS samples
A fully automated 9-cycle two-dimensional chromatography sequence was set up as previously described (205). Peptides were loaded on a 7 cm pre-column (150 µm i.d.) containing a Kasil frit packed with 3.5 cm 5 µ Magic C18 100 Å reversed phase material (Michrom Bioresources) followed by 3.5 cm Luna® 5µ SCX 100Å strong cation-exchange resin (Phenomenex, Torrance, CA). Samples were automatically loaded from a 96-well microplate autosampler using the EASY-nLC system (Proxeon Biosystems, Odense, Denmark) at 3 µl/minute. The pre-column was connected to an 8 cm fused silica analytical column (75 µm i.d.; 5µ Magic C18 100 Å reversed phase material (Michrom Bioresources)) via a microsplitter tee (Proxeon Biosystems) to which a distal 2.3 kV spray voltage was applied. The analytical column was pulled to a fine electrospray emitter using a laser puller. For peptide separation on the analytical column, a water/acetonitrile gradient was applied at an effective flow rate of 400 nl/minute, controlled by the EASY-nLC system. Ammonium acetate salt bumps (8µl) were applied at 0%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, and 100% using the 96-well microplate autosampler at a flow-rate of 3 µl/minute in a vented-column set-up.

All direct-EPS samples were analyzed on a LTQ-Orbitrap XL. The instrument method consisted of one MS full scan (400-1800 m/z) in the Orbitrap mass analyzer, an AGC target of 500,000 with a maximum ion injection of 500 ms, 1 µscan and a resolution of 60,000 and using the preview scan option. Five data-dependent MS/MS scans were performed in the linear ion trap using the five most intense ions at 35% normalized collision energy. The MS and MS/MS scans were obtained in parallel. AGC targets were 10,000 with a maximum ion injection time of 100 ms. A minimum ion intensity of 1,000 was required to trigger a MS/MS spectrum. The dynamic exclusion was applied using a maximum exclusion list of 500 with one repeat count with a repeat duration of 30 seconds and exclusion duration of 45 seconds.

Urine and EPS-urine samples

Individual EPS-urines (5 prostate cancer and 6 BPH) were analyzed in triplicate using the same 9-cycle MudPIT procedure as described above. A quaternary HPLC pump was interfaced with a linear ion-trap mass spectrometer (LTQ, Thermo Fisher Scientific, San Jose, CA) equipped with a nanoelectrospray source (Proxeon Biosystems, Odense, Denmark). The pooled urine and EPS-urine samples were analyzed in triplicate on a LTQ Orbitrap XL, using a modified 5-cycle MudPIT, as previously described (205).
2.5.5 Protein identification and data analyses

Direct-EPS samples

Raw data was converted to m/z XML using ReAdW and searched by X!Tandem against a locally installed version of the human UniProt complete human proteome protein sequence database (version 2009_11; 20,323 sequences). The search was performed with a fragment ion mass tolerance of 0.4 Da, a parent ion mass tolerance of ±10 ppm. Complete tryptic digest was assumed. Carbamidomethylation of cysteine was specified as fixed, and oxidation of methionine as variable modification. A target/decoy search was performed to experimentally estimate the false positive rate and only proteins identified with two unique high quality peptide identifications were considered as previously reported (297) (1 decoy protein identified; FDR <0.5%). An in-house protein grouping algorithm was applied to satisfy the principles of parsimony.

Urine and EPS-urine samples

The following parameters were applied according to the instrument used:

LTQ analyses of individual EPS-urines

The search was performed with a fragment ion mass tolerance of 0.4 Da and a parent ion mass tolerance of 4 Da. Complete tryptic digest was assumed. Carbamidomethylation of cysteine was specified as fixed and oxidation of methionine as a variable modification. Only proteins identified with two unique high-quality peptide identifications per triplicate were considered, as previously reported (269, 296, 298) (11 decoy proteins identified; FDR ~1%). Each sample (n = 11) was examined by 3 technical replicates (33 total MudPIT analyses).

LTQ-Orbitrap XL analyses of pooled urines and EPS-urines

The search was performed with a fragment ion mass tolerance of 0.4 Da and a parent ion mass tolerance of ±10 ppm. Complete tryptic digest was assumed. Carbamidomethylation of cysteine was specified as fixed and oxidation of methionine as variable modification. Only proteins identified with two unique high quality peptide identifications per analyzed sample were
considered, as previously reported (205, 296) (2 decoy proteins identified; FDR ~0.5%). Each sample pool was analyzed by ≥3 technical replicates (13 total MudPIT analyses).

Protein relative abundance was calculated using the QSpec algorithm (269, 296, 298). Proteins were considered to be up-regulated in the pooled EPS-urine samples versus the urine samples if they complied with the following parameters: false discovery rate (FDR) <0.05 and fold change (FC) ≥ 2, based on the QSpec algorithm (241, 295).

2.5.5.1 Gene Ontology annotation and data comparisons

Direct-EPS

Annotation mappings (i.e. GO, predicted transmembrane domains, signal peptides etc.) were performed using the ProteinCenter bioinformatics software (Proxeon Biosystems, Odense, Denmark). Comparison of direct-EPS detected proteins against other proteomic datasets (prostatic secretions (275), seminal plasma (275), urine (264), condition media (267)) was accomplished by ProteinCenter. Proteins were sequence aligned (i.e. BLASTed) against each other and only proteins with at least 95% sequence identity were considered to match (i.e. protein clusters). Comparison to the Human Gene Atlas (mRNA microarray) (124) and the Human Protein Atlas (268) was accomplished via gene mapping using an in-house relational database. For the comparison to data in the Human Protein Atlas, only proteins with available IHC staining in normal or prostate cancer tissues were included. The staining criteria used for antibody selection were directly adapted from the Human Protein Atlas website and are based on rigorous multi-step antibody quality control criteria established by this consortium. Only gene products with a medium to high IHC scoring intensity in either normal prostate or prostate cancer tissue were used. Detailed information regarding the quality assurance can be found at (http://www.proteinatlas.org/qc.php). Cluster analyses (Euclidean) were performed using the open-source tool Cluster 3.0 (http://bonsai.ims.u-tokyo.ac.jp/mdehoon/software/cluster/software.htm) and heat maps were displayed using TreeView (http://jtreeview.sourceforge.net).

Urine and EPS-urine
Functional annotations (Gene Ontology terms, KEGG pathways, and Swiss Prot entries) were assigned using the Database for Annotation, Visualization and Integrated Discovery (DAVID, bioinformatics resources v6.7; http://david.abcc.ncifcrf.gov/) (265). Unique proteins detected in the EPS-urine dataset were compared to the UniProt database and the top five significantly over-represented categories were reported (p-value <0.001). Comparisons of the present EPS-urine dataset to urine (266) and direct-EPS (299) datasets was accomplished using ProteinCenter (Proxeon Biosystems, Odense, Denmark). Proteins were sequence-aligned against each other and only proteins with at least 95% sequence identity were considered a match (i.e. protein clusters).

2.5.6 Prostate-enriched proteins characterization

The BioGPS portal (http://biogps.org/) (124) was used to map identified proteins against available mRNA microarray datasets. Twenty-five major organ systems among those available in BioGPS were selected and linked to proteins via gene accessions. The expression level for each gene was based on averaged probe intensities, and the significant enrichment in prostate tissue (>2 fold change) was calculated as a log₂ ratio compared to the other selected tissues. The random sampling analysis was carried out using the unpaired one-tailed Student’s t-test. A p-value ≤0.05 was considered statistically significant. The alphabetical order of the selected organs is as follows: bone marrow, colon, heart, hypothalamus, kidney, liver, lung, lymph node, ovary, pancreas, placenta, prostate, salivary gland, skeletal muscle, skin, small intestine, smooth muscle, spinal cord, testis, thalamus, thymus, thyroid, uterus, whole blood, and whole brain.

The UniProt database (260) was used to assign subcellular localization to the 49 proteins enriched in EPS-urine pooled samples. The annotations were manually reported and the 49 proteins were grouped into three main categories: secreted, membrane, and intracellular (which includes cytoplasmic, nuclear, and lysosomal). Identified proteins were also screened against the Human Protein Atlas database (260) for availability of antibodies and to examine their prostate tissue expression patterns.
2.5.6.1 SDS-PAGE and Western blot analysis on urine and EPS-urine pools

For Western blotting, 40 μg of total proteins were separated on 8 or 10% SDS-PAGE gels and blotted on PVDF membranes (0.2 μm; Bio-Rad Laboratories, Hercules, CA). Membranes were blocked with 5% milk in TBS-Tween (0.2%) for 1 hour at room temperature and subsequently incubated overnight at 4°C with the following primary antibodies: anti-Lactoferrin (1:1000 #ab10110; Abcam, Cambridge, UK), anti–CD10 (MME 1:1000 #ab951; Abcam, Cambridge, UK), anti–TIMP1 (1:2000 #RP1-TIMP1; Triple Point Biologics, Forest Grove, OR), anti–CD13 (ANPEP 1:500 #ab7417; Abcam, Cambridge, UK), anti–TGM4 (1:500 #sc55791; Santa Cruz Biotechnology, Santa Cruz, CA), anti–14-3-3σ (1:250 #ab14123; Abcam, Cambridge, UK), and anti–PARK7 (1:1000 #ab11251; Abcam, Cambridge, UK). After three 10-minute washes with TBS-Tween (0.2%), membranes were incubated with anti-mouse/anti-rabbit/anti-goat IgG-HRP secondary antibody (Invitrogen, Carlsbad, CA) at a dilution of 1:25,000 for 1 hour at room temperature, washed and visualized with the SuperSignal West Pico chemiluminescent substrate (Thermo Fisher Scientific, San Jose, CA).
Chapter 3
Identification of candidate biomarkers of aggressive prostate cancer


3 Identification of candidate biomarkers of aggressive prostate cancer

3.1 Abstract

Current protocols for the screening of prostate cancer cannot accurately discriminate clinically indolent tumors from more aggressive ones. One reliable indicator of outcome has been the determination of organ-confined (OC) versus non-OC disease but even this determination is often only made following prostatectomy. This underscores the need to explore alternate avenues to enhance outcome prediction of prostate cancer patients. Fluids that are proximal to the prostate, such as expressed prostatic secretions (EPS), are attractive sources of potential prostate cancer biomarkers as these fluids likely bathe the tumor and accumulate disease related factors. Direct-EPS samples from 16 individuals with extracapsular (EC, n = 8) or OC (n =8) prostate cancer were used as a discovery cohort, and analyzed in duplicate by 9-step multidimensional protein identification technology (MudPIT) on a LTQ-Orbitrap XL mass spectrometer. A total of 624 unique proteins were identified by at least two unique peptides with a 0.2% false discovery rate (FDR). A semi-quantitative spectral counting algorithm identified 133 significantly differentially expressed proteins in the discovery cohort. Integrative data mining prioritized 14 candidates, including two known prostate cancer biomarkers: prostate specific antigen (PSA) and prostatic acid phosphatase (PAP), which were significantly elevated in the direct-EPS from the
OC cancer group. These and five other candidates (SFN, MME, PARK7, TIMP1, and TGM4) were verified by Western blotting in an independent set of direct-EPS from patients with biochemical recurrence (BCR) (n = 5) versus patients with no evidence of recurrence upon follow-up (n = 10). Lastly, we performed proof-of-concept selected reaction monitoring mass spectrometry (SRM-MS)-based relative quantification of the five candidates using unpurified heavy isotope-labeled synthetic peptides spiked into pools of EPS-urines from men with EC and OC prostate tumors. This study represents the first efforts to define the direct-EPS proteome from two major subclasses of prostate cancer using shotgun proteomics and verification in EPS-urine by SRM-MS.

3.2 Introduction

Prostate cancer is the most common malignancy to affect men in the Western world, but only 15–20% of these men will present with aggressive, lethal disease (15, 300) whereas the majority of patients will die of other causes. Although the implementation of large-scale screening for prostate cancer using serum PSA has dramatically improved early detection of disease, unnecessary biopsies and patient overtreatment are becoming increasingly evident (16, 300). Consequently, there has been a shift in emphasis away from detection of prostate cancer and toward identification of lethal disease. Currently, Gleason grading is considered to be one of the best outcome predictors; however, patients with Gleason 7 tumors are in the clinical “gray zone,” whereby the predictive ability of Gleason grading is mixed (72, 145). A recent study constructed a 157-gene signature based on the comparison of Gleason score <6 and 8 patients, and could show that their panel could predict lethality in the cohort of Gleason 7 patients (145). Nonetheless, the development and large-scale implementation of prognostic markers of prostate cancer has been hampered by numerous factors owing, in part, to the heterogeneous and multifocal nature of the disease (2). Although the widely used Gleason grading system attempts to control for heterogeneity of the glands and multifocality of cancerous lesions by summing the 2–3 most commonly observed histological patterns via inspection of multiple (typically 8–12) core biopsies, cancerous foci are still often missed (2, 300) providing only partial information leading to imprecise diagnoses and prognoses. Pathologic staging remains the gold standard for disease staging and risk assessment (301, 302); however, this process lacks timeliness in discriminating OC from EC disease. Indeed, one-third of individuals with non-OC disease are
identified only after surgery (303). Furthermore, 35% of men treated with radical prostatectomy (RP) with curative intent subsequently develop BCR (304-307) and the mean time from surgery to recurrence is 3.5 years (72). Significant risk factors for time to prostate specific mortality following BCR after RP are PSA doubling time, pathological Gleason score, and time from surgery to BCR (72). Estimates place the percent of lethal cases at 20–25% of all patients that show BCR, suggesting that nearly 75–80% of patients in this group may be over treated (69).

There is an emerging trend toward recruitment of men with perceived low-risk disease to an “active surveillance” monitoring approach. This is based on the supposition that most prostate cancers are slow growing, and that the more aggressive forms can be identified during a period of observation with little increased risk of death. Although a consensus may not exist for defining the disease stage where active surveillance is warranted, there is considerable agreement that men who have a PSA level less than 10 ng/ml, impalpable disease (clinical stage T1c) and only 1 biopsy core out of 12 or more that show Gleason 6 cancer are most likely to harbor indolent disease (308). Even so, these candidates for active surveillance will still contain individuals who will have disease progression and die from their cancer. Thus, despite efforts to recruit individuals to active surveillance protocols, overtreatment of prostate cancer is fueled by the lack of reliable means to accurately discriminate between men with clinically indolent prostate cancer from those with more aggressive disease (309, 310). This inability to accurately predict prostate cancer aggressiveness based solely on standard clinicopathologic features clearly underscores the need to explore the ability of additional biomarkers to enhance outcome prediction for men with prostate cancer. Furthermore, it is important to acknowledge that a single biomarker alone is unlikely to have sufficient prognostic power; rather, the integration of a panel of biomarkers hold the promise for improved prostate cancer detection and prognosis (300).

Fluids that are proximal to the prostate are attractive sources of potential prostate cancer biomarkers (255, 300), as they house secreted proteins and sloughed cells that provide a presumably more comprehensive assessment of the organ and extent of disease. Further, fluids such as urine are clinically favorable for their ease of collection, the volume and frequency at which they can be obtained, and their adaptability to routine clinical assays. Prostate-proximal fluids include seminal fluid, semen, and EPS. Here, we focus on the analysis of EPS as our biological specimen, using direct-EPS samples for the discovery of candidate prognostic biomarkers and both direct-EPS and pooled EPS-urines derived from independent sets of patients
for candidate biomarker verification. Direct-EPS is a prostatic fluid that is collected from patients undergoing prostatectomy by massaging the organ and expelling 0.5–1 ml of the fluid just prior to surgical removal. It was chosen as our discovery fluid as it is expected to house prostate-secreted proteins at a higher concentration and purity, and we have developed a workflow for the in-depth proteomic analysis of this fluid (260). Following discovery proteomics in 16 clinically stratified direct-EPS samples, verification studies were performed using independent sample sets of direct-EPS. Next, we focused our attention on the verification and quantitative analysis of candidate proteins in pooled EPS-urines. Before EPS-urine collection, men undergo digital rectal examination (DRE), often as part of a routine procedure, which causes direct-EPS to be expelled from the prostate and subsequently voided in urine. Because EPS-urine can be collected with substantial ease and in greater volumes and frequencies than direct-EPS, much attention has been paid to this fluid as a valuable resource of prostate cancer biomarkers amenable to routine clinical analysis. Following the recent FDA approval of the EPS-urine assay for prostate cancer gene 3 (PCA3), standardized clinical collection protocols will be widely implemented and easier access to this fluid is expected. Moreover, we have recently identified a number of prostate-enriched proteins in EPS-urine by comparing its proteome to a urine background (127).

The present study used MudPIT coupled with bioinformatics to catalog and comparatively analyze the direct-EPS proteomes from a small cohort of patients with EC versus OC prostate cancers. A semi-quantitative algorithm based on spectral counts (QSpec) (275) and an integrative data mining strategy led to the selection of a number of putative biomarkers that were verified by Western blotting in direct-EPS. Lastly, to demonstrate accurate quantitative measurements of verified candidates in EPS-urine, a pilot study utilizing SRM-MS was undertaken as a proof-of-concept.

3.3 Results

3.3.1 Proteomic characterization of expressed prostatic secretions from organ-confined and extracapsular prostate cancers

An overview of the study approach is illustrated in Figure 2.1. Direct-EPS derived from 16 individuals who were clinically stratified as having EC (n = 8) or OC (n = 8) prostate tumors
(Table 3.1) were digested in-solution and analyzed by a 9-step MudPIT on a LTQ-Orbitrap XL mass spectrometer, in duplicate (32 MudPITs).

Figure 3.1 Workflow of study design from the discovery of differentially expressed proteins to verification of candidates in prostatic fluids.
Table 3.1 Clinical information of patients used for discovery proteomics.

<table>
<thead>
<tr>
<th>Sample #</th>
<th>Age</th>
<th>Serum PSA (ng/ml)</th>
<th>T-stage</th>
<th>Gleason</th>
<th>Prostate volume (cm²)</th>
<th>% cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extracapsular tumor group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>57</td>
<td>4.9</td>
<td>T3</td>
<td>5+4</td>
<td>42</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>57</td>
<td>5.6</td>
<td>T3a</td>
<td>4+3</td>
<td>32</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>67</td>
<td>8</td>
<td>T3a</td>
<td>4+4</td>
<td>29</td>
<td>N/A</td>
</tr>
<tr>
<td>4</td>
<td>58</td>
<td>3.2</td>
<td>T3b</td>
<td>4+4</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>61</td>
<td>6</td>
<td>T2c</td>
<td>4+3</td>
<td>28</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>5.5</td>
<td>T3b</td>
<td>3+4</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>56</td>
<td>4.1</td>
<td>T3b</td>
<td>4+3</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>8</td>
<td>66</td>
<td>5</td>
<td>T3a</td>
<td>3+4</td>
<td>26</td>
<td>5</td>
</tr>
<tr>
<td>Organ-confined tumor group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>65</td>
<td>7.9</td>
<td>T2a</td>
<td>3+3</td>
<td>58</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>74</td>
<td>2.9</td>
<td>T2c</td>
<td>3+3</td>
<td>58</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>3.6</td>
<td>T2c</td>
<td>3+4</td>
<td>54</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>58</td>
<td>2.6</td>
<td>T2c</td>
<td>3+3</td>
<td>46</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>66</td>
<td>1.5</td>
<td>T2c</td>
<td>3+3</td>
<td>46</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>57</td>
<td>2.1</td>
<td>T2c</td>
<td>3+3</td>
<td>33</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>52</td>
<td>6.2</td>
<td>T2a</td>
<td>3+3</td>
<td>N/A</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>59</td>
<td>9.2</td>
<td>T2a</td>
<td>3+3</td>
<td>75</td>
<td>2</td>
</tr>
</tbody>
</table>

A total of 624 unique proteins were identified by at least two unique peptides with a 0.2% FDR (1 total reverse proteins) (Figure 3.2, 3.3, Appendix Tables 3.1 and 3.2). Among these, 78 proteins (13%) were only identified in the EC group and 216 (34%) were only identified in the OC group, while the majority (330 proteins, 53%) were shared by both (Figure 3.4). A substantial overlap of our data set with other published prostate-proximal and urine proteomic data sets was observed (Figure 3.5). Furthermore, a large proportion of proteins had at least one predicted transmembrane domain (TMD) and signaling peptide (SP) sequences (Figure 3.6). Proteins in this fluid spanned a wide range of functional roles, as determined via Gene Ontology (GO) enrichment (Figure 3.7). A number of Gene Ontology terms were significantly enriched in both tumor groups. They were implicated in proteolysis, regulation of programmed cell death, and cellular proliferation in the biological process category, as well as peptidase activity and cytoskeletal protein binding in the molecular function category. Notably, many proteins were localized in the extracellular region, as indicated by their over-representation in Gene Ontology.
Figure 3.2 Average number of unique proteins identified in the duplicate MudPIT runs for discovery direct-EPS samples.

Figure 3.3 Total number of peptides (A) and total number of spectra (B) recorded for each individual direct-EPS analyzed in the discovery phase.
Figure 3.4 Venn diagram depicting the total number of proteins exclusively expressed in the EC or OC groups.

Figure 3.5 The current proteomics dataset compared to prostate-related proteomics data from various reports.
Figure 3.6 Proportion of proteins with predicted SPs and TMDs.

Figure 3.7 Gene Ontology enrichment of the EC and OC direct-EPS.
3.3.2 Semi-quantitative comparison of expressed prostatic secretions from extracapsular and organ-confined prostate cancers

To identify differentially expressed proteins in both patient groups we employed the QSpec algorithm (275). Only proteins with a FDR of < 0.05 and at least a two-fold change in normalized spectral abundance factor between the EC and OC groups for each protein were considered. This analysis resulted in 133 differentially expressed proteins for further consideration (Figure 3.8), and 100 (77%) of these were present in a functional interaction (FI) network.

Figure 3.8 Distribution of all 624 proteins identified in the current direct-EPS data set. Blue and red dots represent the 133 differentially expressed proteins based on a fold change of at least 2 and a FDR of ≤0.05 by QSpec analysis (red dots are proteins up-regulated in the EC tumor
group, blue dots are significantly down-regulated in the EC tumor group, with the 14 shortlisted candidates indicated by their names).

Following extraction of these proteins from the FI network and the addition of linker genes, a proteomics-associated prostate cancer network consisting of 161 genes, 61 of which are linkers, was obtained. To decompose this network into smaller independent pieces and run pathway annotation for every cluster separately, we used network community analysis to automatically identify network modules that contain genes and their co-regulators that are involved in common biological processes with high probability. We identified five modules of at least 10 proteins and pathway enrichment analysis was done for each cluster separately. The first module was significantly enriched in the “Glycolysis/ gluconeogenesis” pathway; the second one was mapped to the cell cycle-related process “G2/M transition”; the functional annotation of the third and fifth modules showed that they were significantly enriched in hemostasis processes like “Complement and coagulation cascades” and “Further platelet releasate” and others (Figure 3.9, 2.10, and Appendix Table 3.3). The fourth module was not significantly associated with any known processes or pathways.
Figure 3.9 Two significantly enriched pathways within the 133 differentially expressed proteins are involved in hemostatic and immunological processes. Pink circles represent proteins up-regulated in the EC tumor group; blue circles represent proteins down-regulated in the EC group; orange triangles represent linkers.
3.3.3 Generation of candidate shortlist

To select proteins for verification in a small, independent cohort of direct-EPS samples, we performed a systematic data mining strategy using in-house and publicly available resources to minimize the panel of 133 proteins to a refined set (Figure 3.11).
Proteins were ranked initially based on six features by assigning a +1 value to a given protein with the following annotations (0 if the term could not be assigned): 1) present by MS in six out of eight samples corresponding to the risk group that the candidate was found to be higher in by spectral counting; 2) overexpression of candidate genes in the prostate as demonstrated by at least a threefold above median mRNA expression in normal prostate tissue compared with 22 different normal human tissues; 3) presence of predicted TMDs; 4) presence of predicted SP sequences; 5) cellular localization assignment to “cell surface” (GO:0009986) and “extracellular” (GO:0005576) by Gene Ontology annotation; and 6) cancer-associated protein as available from the Human Protein Atlas database. The rationale for each of the key features is as follows: as individual biological specimens can have a high level of variation, we set the criteria...
that a candidate protein should have been detected in at least six out of eight samples in its associated risk group. This would minimize the inclusion of proteins that are over represented by large numbers of spectra in just a small number of samples. The integration of mRNA microarray profiles from normal prostate and other human tissues with our proteomic data set was performed with the rationale that the most promising biomarker candidates would be preferentially expressed in the prostate, and thus may have an important functional role in the organ that contributes to the outcome of the patient. Microarray data was therefore solely used to select proteins with a likely tissue-selective expression pattern. We also took into consideration the presence of predicted TMDs and SPs, as well as Gene Ontology annotations linked to the cell surface and extracellular region as they are characteristics of secreted and plasma membrane proteins, found in many currently known biomarkers. Sixty proteins had at least three out of six features; 32 of which are cancer-associated proteins (253). Because of limitations in antibody availability, a small subset of proteins (14 candidates, Table 3.2) were chosen for verification. Also included in this list were two well-defined prostate cancer biomarkers, PSA and PAP, their levels were also evaluated by ELISA in a different set of 14 direct-EPS samples. As shown in Figure 3.12, the PSA and PAP concentrations determined by ELISA and Western blotting yielded a similar distribution as the MS analysis, with an elevated level of expression being observed in the OC group compared with the EC group. Further verification of the expression of a subset of the short-listed candidate proteins was carried out in new sets of EPS fluids.

Table 3.2 Fourteen candidates selected by data mining scheme.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Direction</th>
<th>mRNA overexpression in prostate</th>
<th>TMD</th>
<th>SP</th>
<th>Present in 6 of 8</th>
<th>Cellular localization</th>
<th>Cb</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADAMTS1</td>
<td>down</td>
<td>✅</td>
<td></td>
<td>✅</td>
<td>✅</td>
<td>membrane, cytosol, extracellular</td>
<td>✅</td>
<td>5</td>
</tr>
<tr>
<td>MME</td>
<td>down</td>
<td>✅</td>
<td></td>
<td>✅</td>
<td>✅</td>
<td>membrane, extracellular</td>
<td>✅</td>
<td>5</td>
</tr>
<tr>
<td>PAP</td>
<td>down</td>
<td>✅</td>
<td></td>
<td>✅</td>
<td>✅</td>
<td>membrane, cytosol, extracellular, vacuole</td>
<td>✅</td>
<td>5</td>
</tr>
<tr>
<td>PSA</td>
<td>down</td>
<td>✅</td>
<td></td>
<td>✅</td>
<td>✅</td>
<td>cytosol, nucleus, extracellular, organelle lumen</td>
<td>✅</td>
<td>5</td>
</tr>
<tr>
<td>TGM4</td>
<td>down</td>
<td>✅</td>
<td></td>
<td>✅</td>
<td>✅</td>
<td>membrane</td>
<td>✅</td>
<td>5</td>
</tr>
<tr>
<td>PEDF</td>
<td>up</td>
<td>✅</td>
<td></td>
<td>✅</td>
<td>✅</td>
<td>cytosol, extracellular, mitochondria</td>
<td>✅</td>
<td>4</td>
</tr>
<tr>
<td>PTPRS</td>
<td>down</td>
<td>✅</td>
<td></td>
<td>✅</td>
<td>✅</td>
<td>membrane, extracellular</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>SFN</td>
<td>down</td>
<td>✅</td>
<td></td>
<td>✅</td>
<td>✅</td>
<td>cytosol, nucleus, extracellular</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>TF</td>
<td>up</td>
<td>✅</td>
<td></td>
<td>✅</td>
<td>✅</td>
<td>membrane, endosome, cytosol, extracellular, mitochondria</td>
<td>✅</td>
<td>4</td>
</tr>
<tr>
<td>ANXA1</td>
<td>down</td>
<td>✅</td>
<td></td>
<td>✅</td>
<td>✅</td>
<td>membrane, cytosol, nucleus, cytoskeleton</td>
<td>✅</td>
<td>3</td>
</tr>
<tr>
<td>KCRB</td>
<td>down</td>
<td>✅</td>
<td></td>
<td>✅</td>
<td>✅</td>
<td>membrane, cytosol, mitochondria</td>
<td>✅</td>
<td>3</td>
</tr>
<tr>
<td>MYH9</td>
<td>down</td>
<td>✅</td>
<td></td>
<td>✅</td>
<td>✅</td>
<td>membrane, cytosol, nucleus, cytoskeleton</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>PARK7</td>
<td>down</td>
<td>✅</td>
<td></td>
<td>✅</td>
<td>✅</td>
<td>cytosol, extracellular, mitochondria</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>TIMP1</td>
<td>down</td>
<td>✅</td>
<td></td>
<td></td>
<td></td>
<td>cytosol, extracellular, organelle lumen</td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>
Figure 3.12 Measurement of PSA and PAP in direct-EPS. Samples were obtained from patients with EC and OC prostate tumors in the discovery cohort by MS (reported as normalized spectral abundance factors, NSAF) and in a new set of samples by ELISA (reported as μg/ml; **p-value ≤0.01; *p-value ≤0.05).

3.3.4 Verification of differentially expressed candidates in direct-EPS

For the verification of the shortlisted candidates, we chose a small cohort of direct-EPS samples that could be linked with clinical outcome information related to whether the patient had BCR or
no evidence of recurrence within a two-year period post-prostatectomy (Table 3.3). Using these samples, the expression levels of candidates the verification set were evaluated by Western blotting.

**Table 3.3 Clinical information of patient samples used for Western blotting verification.**

<table>
<thead>
<tr>
<th>Sample #</th>
<th>Age</th>
<th>Pre-op serum PSA (ng/ml)</th>
<th>T-stage</th>
<th>Gleason</th>
<th>Prostate volume (cm²)</th>
<th>% cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Biochemical recurrence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>64</td>
<td>4.8</td>
<td>T1c</td>
<td>3+3</td>
<td>31</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>57</td>
<td>4.6</td>
<td>T2a</td>
<td>3+4</td>
<td>40</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>64</td>
<td>37.5</td>
<td>T2b</td>
<td>4+3</td>
<td>53</td>
<td>74</td>
</tr>
<tr>
<td>4</td>
<td>58</td>
<td>8.9</td>
<td>T2c</td>
<td>3+4</td>
<td>41</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>56</td>
<td>4.1</td>
<td>T2b</td>
<td>4+3</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No evidence of recurrence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>61</td>
<td>7.8</td>
<td>T2a</td>
<td>3+3</td>
<td>38</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>64</td>
<td>1.3</td>
<td>T2a</td>
<td>3+3</td>
<td>36</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>62</td>
<td>3.6</td>
<td>T1c</td>
<td>3+3</td>
<td>N/A</td>
<td>&lt;2</td>
</tr>
<tr>
<td>4</td>
<td>51</td>
<td>4.6</td>
<td>T1c</td>
<td>3+3</td>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>58</td>
<td>4.5</td>
<td>T1c</td>
<td>3+3</td>
<td>57</td>
<td>&lt;5</td>
</tr>
<tr>
<td>6</td>
<td>55</td>
<td>6.7</td>
<td>T1c</td>
<td>3+4</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>67</td>
<td>7.3</td>
<td>T1c</td>
<td>3+4</td>
<td>56</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>53</td>
<td>5.2</td>
<td>T1c</td>
<td>3+4</td>
<td>30</td>
<td>15</td>
</tr>
<tr>
<td>9</td>
<td>50</td>
<td>5.7</td>
<td>T2a</td>
<td>3+3</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>10</td>
<td>61</td>
<td>10.4</td>
<td>T1c</td>
<td>3+4</td>
<td>31</td>
<td>15</td>
</tr>
</tbody>
</table>

Five out of the 14 the candidates were successfully verified to correlate with the discovery data (Figure 3.13, 3.14). These are: stratifin (SFN), membrane metalloendopeptidase (MME), Parkinson protein 7 (PARK7), tissue inhibitor of metalloproteinase 1 (TIMP1), and transglutaminase 4 (TGM4). For SFN, QSpec analysis of the discovery data in direct-EPS revealed that it was three-fold higher in the OC group with a FDR of <0.002. In the BCR and nonrecurrent samples, the level of SFN expression was significantly higher in the nonrecurrent group ($p = 0.03$). For MME, it was found to be two-fold higher in the OC proteome with a FDR of <0.002. In the comparison between BCR and nonrecurrent samples, MME expression was found to be significantly higher in expression in the nonrecurrent group compared with BCR patients ($p = 0.05$). In addition, we analyzed EC ($n = 23$) and OC ($n = 19$) direct-EPS samples for the expression of MME using a commercially available ELISA, and found a similar trend in this new set of samples ($p = 0.07$) (Figure 3.14B). PARK7 had a threefold difference by QSpec with a FDR of <0.002, with higher expression being observed in the OC group. Similarly, this protein was found to be significantly higher in the nonrecurrent group compared with the BCR group ($p = 0.008$). Both proteins TIMP1 and TGM4 were three-fold different in expression in the
discovery sample set, whereby their expression was higher in the OC group with a FDR of <0.002. Upon subsequent analysis in BCR and nonrecurrent patient direct-EPS samples, a similar trend was observed with heightened expression in the nonrecurrent group for TIMP 1 (p = 0.099) but not for TGM4 (p = 0.80).

**Figure 3.13 Differential expression of verified candidates.** Analysis was completed by MS in EC and OC direct-EPS and Western blotting in BCR (R) and nonrecurrent (NR) patient direct-EPS (reported as densitometry values; **p-value ≤0.01; *p-value ≤0.05; x, p-value ≤0.1; NS denotes an insignificant p-value.**
Figure 3.14 Differential expression of candidates measured by immunoassays. (A) Western blots and stained membranes used to probe for candidates in direct-EPS from BCR and nonrecurrent groups of patients. (B) MME ELISA in direct-EPS (x denotes p-value ≤0.1).

3.3.5 Verification of differentially expressed candidates in EPS-urines

A major aim of the current study was directed toward the detection of our candidates in a clinically accessible fluid that can be collected with minimal invasiveness. EPS-urine samples obtained during a DRE meet these criteria. Therefore, EPS-urine samples were obtained from patients about to undergo a prostate biopsy procedure who subsequently were diagnosed with pathology-confirmed EC and OC prostate tumors. Pools of EPS-urines from both groups were prepared (Table 3.4) and used to assess the presence of the candidate proteins in this biofluid by Western blot and a multiplexed SRM-MS approach. All candidates were quantifiable in the EPS-urine with at least one proteotypic peptide by SRM-MS (Figure 3.15, Appendix Figure 3.1, Appendix Tables 3.4, 3.5). The overall expression levels of MME, PARK7, and TIMP1 coincided with the Western blot analyses in the same pools (Figure 3.16). For SFN, one of the two peptides representing SFN (NLLSVAYK) was found to be lower in the OC tumor group,
whereas the other (VLSSIEQK) was not quantifiable. Similarly, both TGM4 peptides were down-regulated in the OC tumor group, which is not in line with the Western blot analyses. Prostate specific antigen expression was found to be similar between the OC and EC tumor groups as measured by two peptides: peptide HSQPWQVLVASR was higher in the OC group by an average of 1.7-fold and LSEPAELTDAVK was by an average of 1.3-fold. This trend was consistent with Western blot analysis of these same pools. ELISA was also performed on the individual EPS-urines that comprise the pools, whereby a notable difference between the two groups were not apparent (Figure 3.17).
Table 3.4 Clinical information of patient samples used to generate EPS-urine pools.

<table>
<thead>
<tr>
<th>Sample #</th>
<th>Serum PSA (ng/ml)</th>
<th>EPS PSA (μg/ml)</th>
<th>T-stage</th>
<th>Gleason</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Extracapsular tumor group</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5.8</td>
<td>67</td>
<td>T2b</td>
<td>4+5</td>
</tr>
<tr>
<td>2</td>
<td>24.5</td>
<td>28</td>
<td>T2a</td>
<td>5+4</td>
</tr>
<tr>
<td>3</td>
<td>500</td>
<td>12</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>4</td>
<td>3.2</td>
<td>75</td>
<td>T2a</td>
<td>4+4</td>
</tr>
<tr>
<td>5</td>
<td>0.1</td>
<td>1</td>
<td>T3a</td>
<td>4+4</td>
</tr>
<tr>
<td>6</td>
<td>0.2</td>
<td>12</td>
<td>T2b</td>
<td>4+5</td>
</tr>
<tr>
<td>7</td>
<td>65.7</td>
<td>77</td>
<td>T3b</td>
<td>5+5</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
<td>T3b</td>
<td>5+5</td>
</tr>
<tr>
<td>9</td>
<td>4.6</td>
<td>150</td>
<td>T1c</td>
<td>3+5</td>
</tr>
<tr>
<td>10</td>
<td>7.2</td>
<td>160</td>
<td>T1c</td>
<td>4+4</td>
</tr>
<tr>
<td>11</td>
<td>3.1</td>
<td>120</td>
<td>T1c</td>
<td>4+4</td>
</tr>
<tr>
<td>12</td>
<td>9.7</td>
<td>24</td>
<td>T2a</td>
<td>3+5</td>
</tr>
<tr>
<td>13</td>
<td>3.8</td>
<td>14</td>
<td>T1c</td>
<td>3+5</td>
</tr>
<tr>
<td>14</td>
<td>6.3</td>
<td>100</td>
<td>T2b</td>
<td>4+5</td>
</tr>
<tr>
<td>15</td>
<td>7.6</td>
<td>120</td>
<td>N/A</td>
<td>4+5</td>
</tr>
<tr>
<td>16</td>
<td>69.7</td>
<td>2</td>
<td>T1c</td>
<td>4+5</td>
</tr>
<tr>
<td>17</td>
<td>59.1</td>
<td>11</td>
<td>T3b</td>
<td>5+4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Organ-confined tumor group</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4.6</td>
<td>200</td>
<td>T1c</td>
<td>3+3</td>
</tr>
<tr>
<td>2</td>
<td>3.6</td>
<td>19</td>
<td>T1c</td>
<td>3+3</td>
</tr>
<tr>
<td>3</td>
<td>5.2</td>
<td>165</td>
<td>T1c</td>
<td>3+3</td>
</tr>
<tr>
<td>4</td>
<td>8.5</td>
<td>33</td>
<td>T1c</td>
<td>N/A</td>
</tr>
<tr>
<td>5</td>
<td>5.7</td>
<td>5</td>
<td>T1c</td>
<td>3+3</td>
</tr>
<tr>
<td>6</td>
<td>12.7</td>
<td>25</td>
<td>T1c</td>
<td>N/A</td>
</tr>
<tr>
<td>7</td>
<td>6.6</td>
<td>90</td>
<td>T1c</td>
<td>3+3</td>
</tr>
<tr>
<td>8</td>
<td>4.3</td>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>T1c</td>
<td>3+3</td>
</tr>
<tr>
<td>10</td>
<td>5.9</td>
<td>160</td>
<td>T1c</td>
<td>3+3</td>
</tr>
<tr>
<td>11</td>
<td>2.7</td>
<td>39</td>
<td>T1c</td>
<td>3+3</td>
</tr>
<tr>
<td>12</td>
<td>9.8</td>
<td>115</td>
<td>T1c</td>
<td>3+3</td>
</tr>
<tr>
<td>13</td>
<td>6.6</td>
<td>4</td>
<td>N/A</td>
<td>3+3</td>
</tr>
<tr>
<td>14</td>
<td>10</td>
<td>63</td>
<td>T1c</td>
<td>3+3</td>
</tr>
<tr>
<td>15</td>
<td>3.6</td>
<td>140</td>
<td>T1c</td>
<td>3+3</td>
</tr>
<tr>
<td>16</td>
<td>3.3</td>
<td>35</td>
<td>T1c</td>
<td>3+3</td>
</tr>
<tr>
<td>17</td>
<td>5.6</td>
<td>30</td>
<td>T1c</td>
<td>3+3</td>
</tr>
</tbody>
</table>
Figure 3.15 Standard curves generated for each peptide used for SRM-MS measurements.

Six concentration points were evaluated by 3 technical replicates in EPS-urine digests.
Figure 3.16 Verification of differential expression of candidates in EPS-urines pooled from patients with EC or OC prostate tumors. Five candidates were measured by Western blot and SRM-MS using heavy isotope-labeled peptide standards. Relative quantitative values are shown for each technical replicate and as the average ratio of light/heavy peptide in the OC tumor group divided by the light/heavy peptide in the EC tumor group ± standard deviation for each peptide.
Figure 3.17 PSA measurements in individual EPS-urines from patients with EC and OC prostate tumors by ELISA (NS denotes insignificant p-value). The same patient samples were pooled and assayed by Western blot for PSA as well as SRM-MS.

3.4 Discussion

Emerging advances in proteomics and genomics has led to the identification of a vast number of biomarker candidates for a plethora of malignancies; however, few see clinical application (209, 311). The potential of serum as a source of protein biomarkers is hampered by its immense complexity (118); thus we and others have looked to organ-proximal fluids for the discovery and targeted verification of putative biomarkers of various conditions (260, 264, 267, 312-315). Here, we interrogated a prostate-proximal fluid, EPS, for the identification of candidate biomarkers of aggressive prostate cancer by analyzing a cohort of patients with EC versus OC disease phenotypes. We followed up with verification of these biomarkers to identify aggressive disease through additional analyses of a cohort of BCR and nonrecurrent patient direct-EPS. Efforts to adapt urine as a source of biomarkers for prostate cancer have been made using various approaches. For instance, DNA markers such as GSTP1 hypermethylation (316), RNA markers such as PCA3 (317), and protein markers such as ANXA3 (167), matrix metalloproteinases (169), and urinary versus serum PSA (318, 319). In light of the clinical relevance of urine samples, we accessed EPS-urine for verification and potential future assay development of our candidate biomarkers.
Our 624 EPS protein data set contained a number of previously proposed urinary protein prostate cancer biomarkers, including PSA, ANXA3, and matrix metalloproteinases (MMP7 and 9), all of which demonstrated differential expression between the EC and OC groups. Furthermore, a large proportion of the proteome of EPS was predicted to have TMDs and SP sequences that are characteristic of plasma membrane and secreted proteins, including our five candidates. A cell-surface glycoprotein, MME, has been proposed to have an inhibitory effect on prostate cancer cell growth and migration through its association with PTEN, PI-3 kinase (320) and its loss has been shown to be associated with PSA relapse (321). The observation in the current study was that this protein is significantly higher in expression in the OC and nonrecurrent groups compared with the EC and BCR groups in direct-EPS. In general, there was a trend of decreased expression of most proteins detected in the EC groups relative to the OC groups. However, at the protein network level, proteins in the complement and coagulation cascade were significantly higher in expression in the EC tumor group. There are clearly a high abundance of immune system-related proteins identified in the EPS fluids, which is consistent with the chronic inflammation associated with development and progression of prostate cancer (322). Lymphocytes and macrophages present in the prostate are most often associated with this chronic inflammation, and less frequently plasma cells, eosinophils, and neutrophils (323, 324). The presence of tumor-associated macrophages have been reported as potential biomarkers of poor prognosis in prostate cancers (324, 325), and we have evidence that these macrophages can be detected by flow cytometry of the low speed cell pellets of EPS urine samples in a subset of prostate cancer patients (data not shown). A better understanding of the role of the many detectable biomarkers of immune system activity, like the members of the complement pathway overexpressed in non-OC EPS samples, is critical for improving diagnostic targeting of aggressive prostate cancers.

The complexity of the EPS proteome could be reduced using depletion columns, specifically targeted for high-abundance proteins such as albumin, before digestion. This method may facilitate the quantification of lower-abundance proteins although sample recovery may be a caveat if starting concentrations are low. Advances in nanoparticle technology are also showing promise in their application in proteomics to deepen proteome coverage and enhance recovery of low-abundance proteins. Nanoparticles, functionalized using high-affinity baits to selectively capture desired classes of proteins, have been shown to enrich otherwise, low-abundance
proteins, by orders of magnitude (326-328). Yet other strategies focus on the analysis of subproteomes by extraction of desired organelles or exploiting post-translational modifications to selectively enrich for desired classes of proteins. For instance, glycosylation is the most common post-translational modification and N-linked glycosylation is particularly common in secreted or plasma membrane proteins (329). Enrichment of this class of proteins can be achieved by the selective capture of N-linked glycosites to a solid support via hydrazide chemistry followed by enzymatic release by peptide N-glycosidase F (330, 331), or various lectin affinity approaches (332-334). Quantification can be achieved by stable isotope labeling and tandem MS. Zhao and colleagues have extended the technique of stable isotope labeling by amino acids in cell culture, by generating stable isotope labeled secretome standards for the quantification of soluble proteins in direct-EPS (315). Alternatively, more elegant techniques such as SISCAPA® (335, 336) can selectively enrich for target peptides using anti-peptide antibodies and coupled to SRM-MS for targeted quantification, albeit the pipeline from generation of monoclonal antibodies and their validation may be a significant limiting factor. Analyzing urinary sediments, which accounts for 47.7% of the total protein content by weight (171), may offer interesting markers of prostate cancer that may be missed by solely investigating the soluble fraction. Mataija-Botelho et al. identified 60 proteins in the sediment of urine, many of which were also identified in the soluble fraction, suggesting that simply discarding the sediment may alter abundance ratios of different proteins (337). EPS fluids are also an abundant source of exosomes, a sub-fraction that is also being targeted for discovery of prostate cancer biomarkers (172, 173, 338). Initial pilot proteomic studies of EPS-urine exosome fractions have indicated readily detectable and abundant proteins by MS (126). This exosome fraction is a subcomponent of the proteins analyzed in the present study, and we are currently pursuing more detailed analyses of these EPS-derived exosomes from samples representative of the disease spectrum of prostate cancers.

Our pipeline attempts to bridge discovery proteomics in direct-EPS with validation in a more clinically obtainable fluid, EPS-urine. As such, we observed some discrepancies that can be attributed to differences in direct-EPS and EPS-urine samples, sample size, patient heterogeneity, the translation of shotgun proteomics to targeted SRM-MS, and inherent challenges in SRM-MS quantification. In the current study, discrepancies in SFN and TGM4 by Western blotting and SRM-MS are reported. Although a moderate-to-good correlation between antibody-based assays (namely, ELISA) and targeted proteomics approaches have been reported (250, 339, 340),
discordance between the two methods can be conceptualized by the differences in the targeting approaches. Therefore, it may be possible that the epitope for the SFN and TGM4 antibodies used in the current study may be modified thus resulting in inaccurate measurements, or the antibodies may be also targeting other isoforms of the proteins as both are members of larger 14-3-3 and transglutaminase protein families. Furthermore, proteoforms – that is the different molecular forms in which the protein product of a single gene can be found, including changes due to genetic variations, alternatively spiced RNA transcripts, and post-translational modification (341) – may confound accurate quantification. The quantification of SFN and TGM4 thus remains inconclusive at this point of time.

Recently, Hüttenhain and colleagues demonstrated a high-throughput workflow for the development of SRM-MS assays for a large number of cancer-associated proteins in human urine and plasma (253). In their study, 408 proteins were detectable by SRM-MS in urine, of which 169 were previously undetected in the data sets from Human Protein Atlas and Adachi et al. (124). In another study, the serum and prostate glycoproteome of Pten-null and wild-type mice were comparatively assessed using label-free quantitative proteomics followed by SRM-MS of 39 protein orthologs in sera of prostate cancer patients and controls (240), demonstrating a useful platform for biomarker discovery. Similarly, the current study demonstrates the adaptation of a discovery-driven global proteomics experiment in a prostate-proximal fluid to a targeted quantitative assay in EPS urine. Although a large proportion of the SRM-MS optimization process for peptides can be achieved in a simpler matrix, such as BSA, yeast, or cell lysate background, finer optimization in the highly complex clinical specimen is an important requirement. General caveats associated with protein biomarker studies in urine are low protein abundance, the presence of cellular material, high concentrations of salts and interfering substances such as urea, urobilin and other metabolites, and intra- and inter-individual variability (271). In our hands, initial optimization studies were performed in a BSA background, whereby lower concentrations of heavy isotope labeled peptide standards were easily identifiable by at least three transitions. However, once introduced to the EPS-urine background, we observed large amounts of ion suppression and interference of the heavy isotope-labeled peptide standards that were easily identifiable in BSA. Most peptides showed good linearity within 1–500 fmol on-column range but equimolar concentrations of peptides showed variable responses. Therefore, in the end we used a heavy peptide standard concentration of 500 fmol on-column, which allowed
all peptides to be readily detectable in EPS-urine. Using heavy isotope labeled peptide standards, we quantified PSA, SFN, MME, PARK7, TIMP1, and TGM4 in pooled EPS-urines of EC and organ-confined prostate cancers by 1 or 2 peptides. The SRM-MS experiments presented in this proof-of-concept study primarily serves to demonstrate the applicability of EPS-urines to SRM-MS assay platforms and to support its value as an important clinical biofluid. By no means do pooled EPS-urines represent the heterogeneity of the patient population. The only way to verify the relevance of our proposed candidates would be to quantitatively measure them in large cohorts of individual EPS-urines and thus future work will be dedicated to extending the verification phase of the current study.

3.5 Materials and methods

3.5.1 EPS Sample Properties

Direct-EPS samples were obtained from patients following informed consent and use of Institutional Review Board approved protocols at Urology of Virginia and Eastern Virginia Medical School between 2007 and 2009 and the Research Ethics Review Board at the University Health Network. Sample collection and handling has been previously described (260). Briefly, an aggressive prostate massage was performed in confirmed prostate cancer patients under anesthesia prior to undergoing prostatectomy to collect 0.2–1 ml of fluid. Samples were diluted with saline to 5 ml and stored on ice for a maximum of 1 h. Particulates were sedimented by low-speed centrifugation and the resulting supernatants were aliquoted to 1 ml and stored at -80 °C. The clinical data linked to the patients enrolled in the study are outlined in Tables 2.1 and 2.3. Organ-confined disease is defined as the condition of patients with tumors not beyond stage T2c that are margin, seminal vesicle and lymph node negative. Extracapsular disease is differentiated from OC by evidence of margin involvement, extraproststatic extension, or positive for tumor in seminal vesicles or lymph nodes. BCR is defined as patients with rising serum PSA values over 0.2 ng/ml at 3 months or longer post-prostatectomy. EPS-urines were collected from men reporting to the clinic for a prostate biopsy via prostate massage with three strokes on each side of the median sulcus of the prostate followed by collection of voided urine (10–20 ml), as previously described (255). Pools of EPS-urines were derived from 17 patients classified as having OC prostate cancer and 17 from non-OC, EC cancers using 1 ml per patient sample. Tumor grades, staging, prostate volumes and percent of cancer positive needle biopsy cores were
determined using standard pathological procedures. Patient information was recorded, including demographics, medical history, pathology results, and risk factors, and stored in a Caisis database system. All personal information or identifiers beyond diagnosis and laboratory results were not available to the laboratory investigators.

### 3.5.2 Protein digestion and sample preparation for mass spectrometry

Ultrapure-grade urea, ammonium acetate, and calcium chloride were from BioShop Canada, Inc. (Burlington, ON, Canada). Ultrapure-grade iodoacetamide, and dithiotreitol (DTT) were from Sigma-Aldrich. HPLC-grade solvents (methanol, acetonitrile, and water) and formic acid were from Fisher Scientific. Mass spectrometry-grade trypsin was from Promega (Madison, WI). Solid-phase extraction C18 MacroSpin™ Columns were from The Nest Group, Inc. (Southboro, MA). Amicon spin filter columns, 0.5 ml, 3 kDa MWCO, were from Millipore.

Protein concentrations of the direct-EPS samples used for discovery-based proteomics were quantified by Bradford assay and a volume corresponding to 100 μg of total protein were concentrated to 20 μl using a spin filter column. Concentrated direct-EPS was diluted with 50 μl of 8 M urea, 2 mM DTT, 50 mM Tris-HCl, pH 8.5, and incubated at 37 °C for 1 h. Carbamidomethylation was performed by incubating samples with 8 mM of iodoacetamide for 1 h at 37 °C in the dark. Samples were then diluted to 1.5 M urea using 100 mM ammonium bicarbonate, pH 8.5. Calcium chloride was added to a final concentration of 2 mM and the protein mixture was digested with 4 μg of trypsin at 37 °C overnight. The digested peptide mixture was purified with C18 MacroSpin™ columns and concentrated by vacuum centrifugation and reconstituted to a volume of 40 μl with 0.1% formic acid. Samples were stored at -80 °C until used for MudPIT analysis. A 9-step MudPIT sequence was used as previously described (203, 204, 260).

### 3.5.3 Protein identification and data analysis

Raw data was converted to m/z XML using ReAdW (version 1.1) and searched by X!Tandem (version 06.06.01) against a locally installed version of the human UniProt complete human proteome protein sequence database (release date 2009, 20,323 sequences). The search was performed with a fragment ion mass tolerance of 0.4 Da and a parent ion mass tolerance of 10 ppm, and two miscleavages were allowed. Carbamidomethylation of cysteine was specified as
fixed and oxidation of methionine as a variable modification. A target/decoy search was performed to experimentally estimate the false positive rate and only proteins identified with two unique high quality peptide identifications were considered as previously reported (269, 296, 298) (1 decoy protein identified; FDR ≤0.2%). An in-house protein grouping algorithm was applied to satisfy the principles of parsimony (296). A total of 624 protein groups were identified (Appendix Tables 3.1, 3.2). Proteins with significant differential expression levels were determined based on a hierarchical Bayesian statistical algorithm known as QSpec (275). Proteins were considered to be differentially expressed if normalized spectral abundance factors (342) derived from QSpec were at least two-fold different between EC and OC groups with a FDR ≤0.05.

### 3.5.4 Gene Ontology annotation, pathway enrichment and network-based analysis of EPS proteins

Analysis of differentially expressed proteins by Gene Ontology assignment was performed using DAVID v6.7 (299, 343). Proteins were uploaded into the DAVID functional annotation tool and compared with the human genome background. Enriched pathways and Gene Ontology FAT terms with a minimum of two-fold enrichment and a Fisher’s Exact test p-value ≤0.05 were considered and the top Gene Ontology terms that are shared between the EC and OC tumor groups are reported as the proportion of genes involved relative to the total number of genes involved in the whole human genome background.

To understand the relationship among differentially expressed proteins further, we performed a network-based analysis using a protein FI network (344). From the FI network, which contains close to 50% of the human proteome and more than 200,000 FIs, we extracted the subnetworks that involve the differentially expressed proteins. Using a minimum spanning tree algorithm, the minimum number of additional “linker” genes from the full FI network were added in order to form a single fully-connected subnetwork (345). The subnetwork was clustered using the edge-betweenness algorithm (346) in order to identify functionally related network modules. Clusters with size ≥10 were kept for further analysis. All network diagrams were drawn using Cytoscape v. 2.7 (347). For pathway enrichment analysis of network modules, the binomial test was used and the FDR was calculated with 1000 permutations across all genes in the FI network (Appendix Table 3.3).
3.5.5 Integrative data mining

Differentially expressed proteins were compared with transcriptomic profiles based on gene mapping using an in-house relational database with mRNA microarrays for 22 normal tissues, including the prostate, available through Human Gene Atlas (261). The median mRNA expression level of each gene product was calculated across all 22 tissues and compared with the level of expression in the prostate alone, as detected by a total of 218 probe IDs. In this way, genes that demonstrate a propensity toward elevated expression in the prostate compared with other tissues at the transcriptomic level were derived. To assess whether the differentially expressed proteins had predicted TMDs and SPs, their assigned UniProt IDs were uploaded onto the Protein Center bioinformatics software (Thermo Fisher Scientific), which uses TMAP (http://emboss.sourceforge.net/apps/release/6.0/emboss/apps/tmap.html) and PrediSi (http://www.predisi.de/) tools for predictions. The Human Protein Atlas (348) was used to determine whether a given candidate was previously proposed as a cancer biomarker (Cb; from the Plasma Proteome Institute) and to assess the immunoreactivity of antibodies directed against candidate proteins in normal (non-neoplastic and morphologically normal) and cancerous prostate tissue sections. Comparisons across various proteomics data sets (direct-EPS, seminal plasma, urine, cell lines) (124, 260, 264, 267, 268, 312) were performed using the Protein Center bioinformatics software.

3.5.6 Immunoassays in prostatic fluids

The Human Neprilysin (MME) DuoSet® ELISA Development System (R&D Systems, Minneapolis, MN) was used to measure MME in direct-EPS. Total protein concentration corresponding to 140 μg was obtained from each sample. Samples and standard were diluted in 1% BSA in PBS, pH 7.4 and assayed in duplicate following the protocol outlined by the manufacturer. PSA and PAP ELISAs were performed as previously described (255, 349).

For Western blotting, anti-SFN (ab14123), anti-MME (ab951), anti- PARK7 (ab11251), anti-PSA (ab46976) were obtained from Abcam (Cambridge, MA), anti-TGM4 (sc55791) was from Santa Cruz Bio-technology (Santa Cruz, CA), anti-TIMP1 (RP1T1) antibody was from Triple Point Biologics (Forest Grove, OR). Direct-EPS from patients with BCR and no evidence of recurrence were quantified by Bradford assay and a corresponding 30 μg of total protein were
resolved on Criterion 10–20% Tris-HCl SDS-PAGE gels (Bio-Rad) and transferred to PVDF membranes (Bio-Rad). The membranes were blocked with 5% milk in TBS-T for 1 h at room temperature. Primary antibodies were diluted in blocking solution and incubated at 4 °C overnight, washed with TBS-T, and incubated with secondary antibody diluted in blocking solution for 1 h at room temperature. Signals were developed using SuperSignal West Pico or Femto chemiluminescent substrate (Thermo Fisher Scientific). Pooled EPS-urines were concentrated from 15 ml to 500 μl by using a spin column (Amicon Ultra Spin Filter, 3 kDa MWCO, Millipore) and quantified using Bio-Rad DC assay. Concentrations corresponding to 50 μg of total protein were resolved on 10% SDS-PAGE, transferred to PVDF membranes, and assayed as described above.

3.5.7 Statistical analysis

Following candidate protein selection, we verified their differential expression by Western blot densitometry and/or ELISA. To determine if the protein levels differ between OC tumor groups and EC tumor groups, we applied the Mann-Whitney U test to assess if the difference is statistically significant. A p-value ≤0.05 was considered significant.

3.5.8 SRM-MS development

A spectral library was built using the Skyline software tool (version 0.5 and higher) (350) from the LTQ-Orbitrap XL data obtained from the 16 direct-EPS samples. All spectra had scores that passed a stringent peptide score as determined by X!Tandem target decoy search criteria (i.e. 0.5% decoy spectral matches). In addition, two publicly available consensus libraries were used: NIST_human_QTOF (v 3.0 05/24/2011) and GPM human_ipi_cmp_20. Protein sequences were converted to FASTA format and uploaded into Skyline for the prediction of signature peptides. Peptides were chosen based on previously reported specifications (241): predicted tryptic digests of 7–30 amino acids in length and containing zero possible missed cleavage sites or cysteine and methionine residues, and matching the spectral libraries. Four to 6 of the most intense y-ions with a precursor charge of +2 and a fragment ion charge of +1 were selected from the libraries, resulting in several possible peptides for each of the candidates. The GPM (http://www.gpm.org) was used to narrow down the list of peptides to those that have been observed experimentally with a high number of +2 charge states. Furthermore, no possible missed cleavage sites within 3
amino acids before or after the peptide, no acidic residues on cleavage sites, and uniqueness to the gene (241) were additional requirements. Other favorable conditions included no possible post-translational modification sites, no N-terminal Q, E, no W, and SSR hydrophobicity 5–30 (241, 351, 352). Based on these criteria, 21 signature peptides for six candidates (SFN, MME, PARK7, TIMP1, TGM4, and PSA) were chosen for SRM-MS experimentation (Appendix Table 3.4).

Lyophilized, unpurified 13C- and 15N-labeled arginine and lysine peptides (SpikeTides_L™, JPT Peptide Technologies, Berlin, Germany) were obtained in a 96-well format and solubilized in 80% 0.1 M ammonium bicarbonate and 20% acetonitrile. A heavy isotope labeled peptide standard stock containing all 21 SpikeTides_L™ at equal concentrations was made for subsequent experimentation.

For optimization studies, the heavy peptide standard stock corresponding to 500 fmol on-column was added to BSA digests and automatically loaded from a 96-well microplate autosampler using the EASY-nLC system. Analysis was performed on a TSQ Vantage triple quadrupole mass spectrometer (Thermo Fisher Scientific, San Jose, CA) interfaced with the EASY-nLC system (Proxeon Biosystems, Odense, Denmark), as described previously (204, 260). A 40 min HPLC gradient running at a flow rate of 400 nl/min was used. The gradient conditions were as follows: starting with 100% buffer A (water/0.1% formic acid) followed by an increase to 60% buffer B (acetonitrile/ 0.1% formic acid) for 36 min, followed by a steep increase to 95% buffer B for 2 min, and finally to 100% buffer B for 2 min. For compound optimization, the Thermo TSQ Tune software (Thermo Fisher Scientific, San Jose, CA) was used. The XCalibur data system (Thermo Fisher Scientific, San Jose, CA) was used to input the following SRM-MS method parameters: one scan event over a 40 min method time at a fixed scan width of 0.02 m/z, scan time of 0.010 s, 0.2 fwhm Q1 and 0.7 fwhm Q3 resolution, and data collection in profile mode. For each peptide, 3– 4 most intense transitions for each eluting peptide were selected and their accurate retention times were recorded using Skyline (350). These transitions were grouped into two separate methods to maintain a cycle time of 1 s.
3.5.9 Relative quantification by SRM-MS in EPS-urines

For quantification by SRM-MS, EPS-urine pools were prepared as follows: concentrated EPS-urines were precipitated in 100% methanol (1:20 v/v ratio) overnight at -20 °C. Protein pellets were recovered by centrifuging at 16,000 g for 30 min at 4 °C. Two additional washes with methanol were performed to obtain “clean” protein pellets. The pellets were subsequently resolubilized in 50 μl of 8 M urea, 2 mM DTT, 50 mM Tris-HCl, pH 8.5, and incubated at 37 °C for 1 h. Carboxymethylation was performed by incubating samples with 8 mM of iodoacetamide for 1 h at 37 °C in the dark. Samples were then diluted to 1.5 M urea using 100 mM ammonium bicarbonate, pH 8.5. Calcium chloride was added to a final concentration of 2 mM and the protein mixture was digested with 4 μg of trypsin at 37 °C overnight. The digested peptide mixture was purified with C18 Macrospin™ columns and concentrated by vacuum centrifugation and reconstituted to a volume of 100 μl with 5% acetonitrile, 0.1% formic acid. Samples were incubated at 37 °C for 30 min to allow peptides to go into solution and subsequently centrifuged at 16,000 g for 30 min at 4 °C to remove any remaining debris.

For quantification in EPS-urine pools, a total of 1 μg of peptides on-column (as determined by the micro BCA assay) were spiked with 500 fmol on-column of the heavy isotope-labeled peptide standard stock and assayed following the same conditions as above. Each sample was assayed in four technical replicates (Appendix Table 3.5). Total peak area values and area ratios (light/heavy peptide) were obtained from Skyline (350). Relative quantititative values are shown for each technical replicate and as the average ratio of light/heavy peptide in the OC tumor group divided by the light/heavy peptide in the EC tumor group ± standard deviation for each peptide.
Chapter 4
Systematic development of a quantitative assay for putative prostate cancer biomarkers in expressed prostatic secretions


4 Systematic large-scale development of targeted proteomics assays for biomarker verification in prostate-proximal fluids

4.1 Abstract

Current prostate cancer prognostic factors stratify patients into risk groups, but are inaccurate in predicting outcome, resulting in overtreatment and repeat biopsies of many men with indolent disease. Therefore, a pressing need in prostate cancer management is improved prognostic factors that enable follow-up of men with low risk disease without repeat needle biopsies. To address this, I systematically explored non-invasively collected prostate proximal fluids, known as expressed prostatic secretions (EPS), as a source of prostate biomarkers. Analysis of direct-EPS from extracapsular (EC) and organ-confined (OC) prostate cancers revealed differentially expressed proteins based on spectral counts (244). Selection criteria led to 147 proteotypic peptides suitable for verification using selected reaction monitoring mass spectrometry (SRM-MS). For each candidate, a corresponding crude heavy isotope labeled peptide was synthesized and used for SRM-MS method development. Next, relative quantification of endogenous peptide concentrations in defined subgroups of prostate cancer patients and control EPS-urines was performed, revealing 34 candidate peptides with high potential as prostate cancer biomarkers. A single scheduled SRM-MS assay was developed and used to verify the differential expression of
these candidate peptides by absolute quantification in an independent cohort of stratified EPS-urines.

4.2 Introduction

The worldwide incidence of prostate cancer has been steadily rising but many patients exhibit tumors of an indolent nature; that is, these tumors grow slowly and pose minimal threat to the life of the patient, in the absence of treatment. Once cancer has been confirmed, the optimal course of action is tailored on a patient-to-patient basis, ranging from active surveillance programs to surgery and radiotherapy. The goal of individualized patient management is to spare patients with indolent disease from unnecessary procedures, while identifying and treating those who would benefit from treatment intensification. Gleason grading is a strong prognosticator of prostate cancer, but requires surgical or biopsy specimens. The number of patients admitted to the hospital as a result of post-biopsy complications is sizable, posing a significant burden on health care and risking serious complications. Most men diagnosed with prostate cancer have localized disease with excellent prognosis (localized cancers have 100% 5 year survival, SEER Cancer Statistics http://seer.cancer.gov), but nearly 70% receive early intervention in the form of surgery or radiotherapy (353-357). Furthermore, even active surveillance relies on repeat prostate specific antigen (PSA) testing, digital rectal examination (DRE), and multiple prostate biopsies.

Liquid biopsies, such as circulating tumor cells and free DNA have been proposed as non-invasive prostate cancer biomarkers, but their detection and enrichment remains technically challenging. Cataloguing the secreted and soluble factors released into the interstitial fluid that bathes the organ of interest may provide a novel inventory of putative disease biomarkers. Indeed, I have interrogated the collection of proteins comprising a prostate-proximal fluid, extraprostatic secretions (EPS) (126, 127, 244, 260). Reproducible detection and quantification of a protein in the matrix of interest is a requirement of any potentially valuable disease biomarker, but verification of these candidates is a major bottleneck in the pipeline from discovery to clinical implementation.

Traditionally, immunoaffinity based assays, namely enzyme-linked immunosorbent assays (ELISAs), are used to validate a biomarker; but this approach is time-consuming, costly, and
relies on the existence of validated antibodies for a given target protein. Targeted mass spectrometry (MS) offers an alternative approach to the rapid verification of candidate biomarkers. Selected reaction monitoring mass spectrometry (SRM-MS) is the leading method for the targeted quantification of proteins using MS. It offers excellent selectivity, sensitivity, and throughput, without the need for validated antibodies. In the current chapter, I will describe my efforts that extend the previously published investigations into the EPS proteome (126, 127, 244, 260). Here, I systematically generated a multiplexed quantitative SRM-MS assay targeting candidates with prostate cancer prognostic and diagnostic potential using a clinically relevant prostate proximal fluid.

4.3 Results

4.3.1 Candidate selection for targeted quantification by selected reaction monitoring mass spectrometry

Utilizing a previously published shotgun proteomics dataset derived from direct-EPS comparing the proteomes of EC and OC prostatic tumors (described in Chapter 3), 232 proteotypic peptides (Phase 1 peptides, Appendix Table 4.1) were shortlisted for targeted quantification by SRM-MS in EPS-urines (Figure 4.1). In order to abrogate potential confounding influences due to the differing modes of data acquisition, instruments, and samples, Phase 1 peptides were first evaluated for reproducible detection by SRM-MS. For this, crude heavy isotope labeled synthetic versions of the target peptides (JPT Peptide Technologies GmbH) were spiked into a pool of EPS-urine, and the light (endogenous) and heavy (synthetic) peptides were monitored. The data was manually inspected in order to select peptides that had at least 3 fragments ions that aligned at the expected peptide elution time, had co-eluting light and heavy peptides, had minimal interference, and were reproducible. Out of the 232 peptides monitored, 147 (63%) peptides (Phase 2 peptides, Appendix Table 4.1) met these criteria, and were thus taken forward to an independent cohort of EPS-urine samples.
Figure 4.1 Study workflow outlining the three major phases of the current project. Discovery data from direct-EPS derived from patients with EC or OC prostatic tumors was used to identify candidates demonstrating differential expression (Chapter 3). Proteotypic peptides from these candidates were carefully selected and evaluated by SRM-MS for performance in an EPS-urine background. Next, peptides were analyzed in EPS-urines from individuals with benign prostatic hyperplasia (BPH) or no evidence of disease and compared to EC and OC EPS-urines. Peptide quantification by SRM-MS was performed and 34 promising candidates with diagnostic and prognostic potential were identified based on estimated abundance changes. The 34 candidates were accurately quantified in a different cohort of EPS-urines for verification. Finally, ongoing statistical analyses are being performed to generate a peptide signature with a high potential for prostate cancer diagnosis and prognosis.
4.3.2 Targeted relative quantification of candidates in a cohort of EPS-urines

For targeted quantification of Phase 2 peptides, a cohort of EPS-urines derived from 74 clinically stratified prostate cancer patients and noncancer controls (BPH and normal individuals) was utilized (Cohort A, Table 4.1, Appendix Table 4.4). The overlap of these 147 putative candidates with a previously published dataset composed of 661 cancer-associated proteins in urine (253) was determined. Of the 147 Phase 2 peptides, only 10 overlapped with urine (Figure 4.2A), suggesting that post-DRE urines are enriched in prostate-derived proteins. Indeed, in a previous study (Chapter 2), myself and colleagues have demonstrated the significant enrichment of proteins in EPS-urine (post-DRE urine) when compared to urine (matched pre-DRE urine) (127).

Furthermore, the estimated average concentration of Phase 2 peptides measured in this cohort of EPS-urine demonstrated a dynamic concentration range of peptides spanning ~5 orders of magnitude, with the exception of the two KLK3 (PSA) peptides HSQPWQVLVASR and LSEPAELTDAVK (Figure 4.2B).

Table 4.1 Overview of clinical information from patients enrolled in Cohort A of EPS-urines.

<table>
<thead>
<tr>
<th></th>
<th>Cohort A: EPS-urine overflow</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean age ± SEM</td>
<td>pT stage</td>
<td>GS</td>
<td>+ margins</td>
<td>BCR</td>
<td>Mets</td>
</tr>
<tr>
<td>Controls</td>
<td>24</td>
<td>59.2 ± 1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>11</td>
<td>57.3 ± 2.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPH</td>
<td>13</td>
<td>60.8 ± 2.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>50</td>
<td>58.9 ± 0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organ-confined</td>
<td>37</td>
<td>57.9 ± 1.0</td>
<td>pT2a – pT2c</td>
<td>6 - 8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Extracapsular</td>
<td>13</td>
<td>61.7 ± 1.7</td>
<td>pT2c – pT3b</td>
<td>6 - 9</td>
<td>9</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

SEM, standard error of the mean
pT stage, pathological T stage
GS, pathological Gleason score
+ margins, positive surgical margins
BCR, biochemically recurrent
Mets, evidence of metastasis
Figure 4.2 Peptides in EPS-urine. (A) Venn diagram depicting the number of Phase 2 peptides that were also detected in the urines of a previous publication by Hüttenhain et al. (253). (B) Estimated concentrations of Phase 2 peptides based on the spike-in concentration of the crude heavy isotope labeled peptide standards. Red dots represent Phase 3 peptides that were subsequently quantified by using AQUA peptides.
4.3.3 Selection of peptide candidates for verification

To derive a small set of peptide candidates for verification, a student’s T-test was performed to compare the ratios of peptide abundance between cancer and noncancer groups, as well as EC and OC groups. The first criteria used to select candidates as potential diagnostic and prognostic biomarkers were p-value cutoffs of 0.05 and 0.1, respectively. Next, refinements to the list of candidates were made by adding peptides representing the proteins IGJ and ANXA1 because they were the only candidates that were upregulated in the EC tumor group, although they did not meet the p-value cut-off (p-value = 0.25). Furthermore, KLK3 peptides were added to monitor PSA levels in EPS-urine. Finally, each of the peptides that met these criteria were manually inspected for quality of SRM-MS traces. Overall, 34 candidates (Phase 3 peptides) demonstrated a potential for classifying individuals based on prostatic disease status (Figure 4.1, Figure 4.3, Appendix Table 4.2, Appendix Table 4.6). Peptides with diagnostic potential were considered to be those that could discriminate prostate cancer from noncancer conditions. Twenty-four such peptides were selected, of which the majority were upregulated in cancer (Figure 4.3A and B). These candidates are grossly involved in functions associated with response to oxidation when characterized by Gene Ontology enrichment (Figure 4.3C). Of the 14 candidates that demonstrated a prognostic potential – that is, they could discriminate EC tumors from OC ones – only two candidates were upregulated in the EC tumor group (Figure 4.3A and B). Candidates that were upregulated in the OC group are largely associated with the immune response by Gene Ontology enrichment analysis (Figure 4.3C).

In order to generate a SRM-MS assay for accurate targeted quantification of Phase 3 peptides, each peptide was synthesized as a highly purified (>97% purity) heavy isotope labeled standard (AQUA peptides, Thermo Fisher Scientific) and calibration curves were generated for all candidates in a urine background (Appendix Figure 4.1). All peptides demonstrated excellent linearity ($R^2$ ranging from 0.90 – 0.99) and reproducibility (coefficient of variation ranging from 0.4 – 19%) (Table 4.2, Appendix Table 4.3). The limit of detection ranged from approximately 0.006 to 8 fmol/μg of total urine peptides; the lower limit of quantification ranged from approximately 0.02 to 25 fmol/μg of total urine peptides.
Figure 4.3 Phase 3 peptides. (A) Venn diagram depicting the number of peptides with diagnostic potential and the number of peptides with prognostic potential. (B) The number of peptides that the up- or down-regulated in the cancer group compared to the noncancer group (left graph), and the EC tumor group compared to the OC tumor group (right graph). (C) Gene Ontology terms that are enriched among the diagnostic and prognostic peptides.
Table 4.2. Limits of detection and quantification for 34 peptides determined by response curves generated using AQUA peptides spiked into urine.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Uniprot</th>
<th>Peptide sequence</th>
<th>RT</th>
<th>Quantifier</th>
<th>$R^2$</th>
<th>LOD (fmol/µg)</th>
<th>LLOQ (fmol/µg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANXA1</td>
<td>ANXA1</td>
<td>TPAQFDADLR</td>
<td>24</td>
<td>y7</td>
<td>0.969</td>
<td>0.006</td>
<td>0.019</td>
</tr>
<tr>
<td>ANXA3</td>
<td>ANXA3</td>
<td>LTDFEYR</td>
<td>23</td>
<td>y5</td>
<td>0.974</td>
<td>0.074</td>
<td>0.225</td>
</tr>
<tr>
<td>ANXA3</td>
<td>ANXA3</td>
<td>SEIDLDDLIR</td>
<td>31</td>
<td>y6</td>
<td>0.951</td>
<td>0.034</td>
<td>0.103</td>
</tr>
<tr>
<td>BTD</td>
<td>BTD</td>
<td>VDLITFDPFAGR</td>
<td>36</td>
<td>y9</td>
<td>0.942</td>
<td>3.563</td>
<td>10.796</td>
</tr>
<tr>
<td>C3</td>
<td>C3</td>
<td>TGLQEVYVK</td>
<td>20</td>
<td>y6</td>
<td>0.989</td>
<td>0.315</td>
<td>0.956</td>
</tr>
<tr>
<td>CFB</td>
<td>CFB</td>
<td>YGLTYTATYPK</td>
<td>26</td>
<td>y7</td>
<td>0.989</td>
<td>0.222</td>
<td>0.673</td>
</tr>
<tr>
<td>CKB</td>
<td>CKB</td>
<td>VLPPELAEYR</td>
<td>30</td>
<td>y8</td>
<td>0.896</td>
<td>0.035</td>
<td>0.107</td>
</tr>
<tr>
<td>CLIC1</td>
<td>CLIC1</td>
<td>GFTIPEAFR</td>
<td>30</td>
<td>y5</td>
<td>0.931</td>
<td>8.421</td>
<td>25.518</td>
</tr>
<tr>
<td>CNDP2</td>
<td>CNDP2</td>
<td>TGQEIPVNR</td>
<td>21</td>
<td>y5</td>
<td>0.976</td>
<td>0.016</td>
<td>0.05</td>
</tr>
<tr>
<td>DDT</td>
<td>DOPD</td>
<td>FFPLESWQIYGK</td>
<td>35</td>
<td>y9</td>
<td>0.937</td>
<td>0.06</td>
<td>0.182</td>
</tr>
<tr>
<td>EEF2</td>
<td>EEF2</td>
<td>FSFSVPVVR</td>
<td>23</td>
<td>y5</td>
<td>0.982</td>
<td>0.526</td>
<td>1.593</td>
</tr>
<tr>
<td>FGA</td>
<td>FIBA</td>
<td>NSLFEYQK</td>
<td>22</td>
<td>y5</td>
<td>0.998</td>
<td>0.308</td>
<td>0.933</td>
</tr>
<tr>
<td>FGG</td>
<td>FIBG</td>
<td>YLQEYNSNQQ</td>
<td>21</td>
<td>y7</td>
<td>0.99</td>
<td>0.007</td>
<td>0.02</td>
</tr>
<tr>
<td>GSN</td>
<td>GELS</td>
<td>AGALNNSDAFVK</td>
<td>28</td>
<td>y9</td>
<td>0.994</td>
<td>0.424</td>
<td>1.283</td>
</tr>
<tr>
<td>HEXB</td>
<td>HSBX</td>
<td>VLDIATINK</td>
<td>30</td>
<td>y8</td>
<td>0.975</td>
<td>2.454</td>
<td>7.436</td>
</tr>
<tr>
<td>IDH1</td>
<td>IDHC</td>
<td>ATDFVVPQPGK</td>
<td>23</td>
<td>y5</td>
<td>0.983</td>
<td>0.4</td>
<td>1.211</td>
</tr>
<tr>
<td>IDH1</td>
<td>IDHC</td>
<td>VEITYPSDGTQK</td>
<td>21</td>
<td>y7</td>
<td>0.997</td>
<td>0.024</td>
<td>0.071</td>
</tr>
<tr>
<td>IGJ</td>
<td>IGJ</td>
<td>SSEDPNEDIVER</td>
<td>19</td>
<td>y8</td>
<td>0.996</td>
<td>0.073</td>
<td>0.221</td>
</tr>
<tr>
<td>KLK3</td>
<td>KLK3</td>
<td>HSQPWQVLVASR</td>
<td>25</td>
<td>y9</td>
<td>0.968</td>
<td>0.127</td>
<td>0.386</td>
</tr>
<tr>
<td>KLK3</td>
<td>KLK3</td>
<td>LSEPAELTDVAK</td>
<td>25</td>
<td>y9</td>
<td>0.994</td>
<td>0.034</td>
<td>0.103</td>
</tr>
<tr>
<td>KRT8</td>
<td>K2C8</td>
<td>LVSESSDVLPK</td>
<td>22</td>
<td>y9</td>
<td>0.996</td>
<td>0.08</td>
<td>0.244</td>
</tr>
<tr>
<td>PARK7</td>
<td>PARK7</td>
<td>VTVAQLAGK</td>
<td>19</td>
<td>y7</td>
<td>0.996</td>
<td>0.249</td>
<td>0.754</td>
</tr>
<tr>
<td>PGLS</td>
<td>PGLS</td>
<td>VTLTLPVNAAR</td>
<td>32</td>
<td>y7</td>
<td>0.965</td>
<td>0.044</td>
<td>0.135</td>
</tr>
<tr>
<td>PHGDH</td>
<td>SERA</td>
<td>GGIVDEAGALLR</td>
<td>27</td>
<td>y7</td>
<td>0.982</td>
<td>0.294</td>
<td>0.891</td>
</tr>
<tr>
<td>PHGDH</td>
<td>SERA</td>
<td>TLGIQLGR</td>
<td>29</td>
<td>y7</td>
<td>0.895</td>
<td>0.461</td>
<td>1.398</td>
</tr>
<tr>
<td>PRDX6</td>
<td>PRDX6</td>
<td>LPFPIDDR</td>
<td>32</td>
<td>y6</td>
<td>0.885</td>
<td>0.109</td>
<td>0.33</td>
</tr>
<tr>
<td>PRDX6</td>
<td>PRDX6</td>
<td>LSILYPATTGR</td>
<td>27</td>
<td>y6</td>
<td>0.963</td>
<td>0.318</td>
<td>0.964</td>
</tr>
<tr>
<td>RNH1</td>
<td>RIN1</td>
<td>VNPAEALNLR</td>
<td>28</td>
<td>y6</td>
<td>0.956</td>
<td>0.433</td>
<td>1.312</td>
</tr>
<tr>
<td>SERPINF1</td>
<td>PEDF</td>
<td>LQSLFDSPDFS</td>
<td>30</td>
<td>y10</td>
<td>0.919</td>
<td>0.054</td>
<td>0.162</td>
</tr>
<tr>
<td>SERPINF1</td>
<td>PEDF</td>
<td>LSYXEYTVK</td>
<td>18</td>
<td>y7</td>
<td>0.976</td>
<td>0.063</td>
<td>0.189</td>
</tr>
<tr>
<td>SFN</td>
<td>SFN</td>
<td>NLLSSVAYK</td>
<td>24</td>
<td>y6</td>
<td>0.992</td>
<td>0.065</td>
<td>0.197</td>
</tr>
<tr>
<td>TF</td>
<td>TRFE</td>
<td>YLGEEYVK</td>
<td>23</td>
<td>y6</td>
<td>0.999</td>
<td>0.694</td>
<td>2.102</td>
</tr>
<tr>
<td>TGM4</td>
<td>TGM4</td>
<td>VFVFSEVNGDR</td>
<td>25</td>
<td>y8</td>
<td>0.973</td>
<td>0.21</td>
<td>0.636</td>
</tr>
<tr>
<td>TGM4</td>
<td>TGM4</td>
<td>LEFSTGPNPSIAK</td>
<td>25</td>
<td>y8</td>
<td>0.921</td>
<td>0.358</td>
<td>1.084</td>
</tr>
</tbody>
</table>
4.3.4 Absolute quantification for the verification of candidates in EPS-urines

Next, the 34 Phase 3 peptides were quantified for the verification of their differential expression in an independent cohort of 207 EPS-urines. Prostate cancer EPS-urines were further stratified into pathological stage pT3 and pT2 tumors. Noncancer specimens were derived from patients with BPH and normal individuals (Cohort B, Table 4.3, Appendix Table 4.5).

Table 4.3 Overview of clinical information from patients enrolled in Cohort B of EPS-urines.

<table>
<thead>
<tr>
<th>Cohort B: EPS-urine supernatant</th>
<th>( N )</th>
<th>Mean age ± SEM</th>
<th>GS</th>
<th>BCR</th>
<th>Mets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>117</td>
<td>60.1 ± 0.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>48</td>
<td>58.1 ± 1.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPH</td>
<td>69</td>
<td>61.4 ± 0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>90</td>
<td>59.9 ± 0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pT2</td>
<td>61</td>
<td>59.6 ± 0.8</td>
<td>6 - 8</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>pT3</td>
<td>29</td>
<td>60.3 ± 1.1</td>
<td>6 - 9</td>
<td>10</td>
<td>3</td>
</tr>
</tbody>
</table>

SEM, standard error of the mean
GS, pathological Gleason score
BCR, biochemically recurrent
Mets, evidence of metastasis

Since sufficient sample was available, each EPS-urine was measured in two technical replicates, using a single scheduled SRM-MS method. The duplicate measurements for each sample in Cohort B demonstrated strong reproducibility (Figure 4.4A). A Pearson correlation analysis between the duplicate SRM-MS runs revealed that the majority of peptides have Pearson correlation coefficients of 0.8 or above (Figure 4.4B). Peptides that have a Pearson correlation coefficient of 0.8 or lower were found to be generally low-abundance peptides. The analysis also demonstrated good chromatographic reproducibility (coefficient of variation <1.5%) with most peptides eluting between 20 and 30 minutes over the 40-minute chromatographic gradient (Figures 4.4C and D).
Figure 4.4 Cohort B SRM-MS data quality. (A) Peptide abundance correlation between both replicates analyzed by SRM-MS. (B) The fraction of peptides that fall into different Pearson correlation coefficient cut-offs in the subgroups of EPS-urine samples. (C) The average retention
time and coefficient of variation for each of the 34 peptides. (D) The recorded retention times of the 34 candidates; boxplots represent the interquartile range; whiskers represent the minimum and maximum recorded retention time.

Next, the absolute concentration of each peptide across all 207 samples in Cohort B was determined based on the amount of spike in heavy AQUA peptides (Figure 4.5). The 34 peptides measured in this set of EPS-urines spanned approximately 5 orders of magnitude, from the low fmol to nmol range per µg total protein. As expected, the two peptides representing KLK3 (HSQPWQVLVASR and LSEPAELTDAVK) were found to be the most abundantly expressed peptides in EPS-urine. The remaining peptides spanned approximately 3 orders of magnitude, with the peptide representing DOPD being the highest and one of the peptides representing PEDF being the lowest in this set of EPS-urines. All proteins that were represented by two peptides were found to be within approximately 1 order of magnitude in abundance, except for the peptides representing PEDF, which were approximately 2 orders of magnitude different in abundance.
Figure 4.5 Absolute quantification of Phase 3 peptides in EPS-urine. Box plots represent the median and interquartile range. Whiskers represent the 1-99 percentile. Outliers are represented by red dots and the mean is represented by ‘+’.

4.3.5 Comparative analysis between subgroups of patients

In order to determine candidates with diagnostic and prognostic potential, comparative analyses between subgroups of Cohort B patients were carried out using the SRMStats software tool (version 1.6) (358). A number of candidates were significantly differentially expressed between cancer and normal individuals, with all candidates being significantly upregulated in the cancer
group (Figure 4.6A). The proteins that were represented in this comparison were IDHC, ANXA3, ANXA1, KLK3, KCRB, SERA, and CNDP2. Interestingly, IDHC, ANXA3, and KLK3 were represented by the significant upregulation of two peptides for each protein in the cancer group. When comparing the expression of candidates between the cancer and BPH groups, ANXA3, SERA, ANXA1, and KLK3 were found to be significantly upregulated in the cancer group (Figure 4.6B). The protein ANXA3 was represented by two peptides that were significantly upregulated in cancer. Next, a comparison of pathological stage pT3 and pT2 samples revealed 10 peptides representing HEXB, 6PGL, DOPD, KLK3, SERA, PEDF, PRDX6, and BTD. Two proteins – KLK3 and PRDX6 – were represented by two peptides each (Figure 4.6C). When comparing cancer and noncancer (BPH and normal individuals), ANXA3, IDHC, SERA, ANXA1, KCRB, and KLK3 were significantly upregulated in cancer (Figure 4.6D). ANXA3 and KLK3 were represented by two peptides each. Lastly, a comparison between BPH and normal samples revealed very little alterations, with IDHC being the only protein that was significantly upregulated in BPH (Figure 4.6E). The relative expression of Phase 3 proteins and peptides in EPS fluids (127, 244) using different quantitative methods are summarized in Table 4.4.
Figure 4.6 Comparative quantification of candidates in different subgroups of EPS. Volcano plots representing the fold-changes of Phase 3 peptides in various subgroups of EPS-urines from Cohort B. Red dots represent peptides that are upregulated; blue dots represent peptides that are downregulated; dotted vertical lines represent the boundaries of a 1.5 fold change; dotted horizontal line represents a p-value of 0.05.
Table 4.4 Relative expression of Phase 3 proteins and peptides in EPS fluids using various quantitative methods.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Protein</th>
<th>Peptide</th>
<th>ΔEPS (IE/OC)</th>
<th>ΔEPS (WB/EC/OC)</th>
<th>ΔEPS (vrl/CA-NC)</th>
<th>ΔEPS (vrl/CA-NC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANXA1</td>
<td>ANXA1</td>
<td>TPAGFDDELIR</td>
<td>↓↓↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>ANXA3</td>
<td>ANXA3</td>
<td>LTFDEYR</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>ANXA3</td>
<td>ANXA3</td>
<td>SEIDLLDIR</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>BTD</td>
<td>BTD</td>
<td>VDLITFDTPFAKR</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>C3</td>
<td>C3</td>
<td>TGLQGVEVK</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
</tr>
<tr>
<td>CFB</td>
<td>CFB</td>
<td>YGLVTVYATYPK</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
</tr>
<tr>
<td>CKB</td>
<td>CKB</td>
<td>VLTPELYAELR</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>CLJC1</td>
<td>CLJC1</td>
<td>GFTIEPAFR</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>CNDP2</td>
<td>CNDP2</td>
<td>TQGEIPVNR</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>DDT</td>
<td>DDP</td>
<td>FFPLESWQGK</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>EEF2</td>
<td>EF2</td>
<td>FSVSPVVR</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
</tr>
<tr>
<td>FGA</td>
<td>FIBA</td>
<td>NLSFYEQK</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
</tr>
<tr>
<td>FGG</td>
<td>FIBG</td>
<td>YLQEYNSNNQK</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
</tr>
<tr>
<td>GSN</td>
<td>GELS</td>
<td>AGALINSDAVFLK</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
</tr>
<tr>
<td>HEBX</td>
<td>HEBX</td>
<td>VLDHATINK</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>IDH1</td>
<td>IDHC</td>
<td>ATDFVPDPGPK</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
</tr>
<tr>
<td>IDH1</td>
<td>IDHC</td>
<td>VETITYPSDGTQK</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>IGJ</td>
<td>IGJ</td>
<td>SSEDPNEDVER</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
</tr>
<tr>
<td>KLK3</td>
<td>KLK3</td>
<td>HSQFVWVLASR</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>KLK3</td>
<td>KLK3</td>
<td>LSEPAELTDVK</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>KRT8</td>
<td>K2C8</td>
<td>LVSESSDLPK</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>PARK7</td>
<td>PARK7</td>
<td>VTVAGLAGK</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>PGLS</td>
<td>PGL</td>
<td>VTLTLPVLSK</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>PHGDH</td>
<td>SERA</td>
<td>GGINDEGALLR</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>PHGDH</td>
<td>SERA</td>
<td>TLGILGLGR</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>PRDX6</td>
<td>PRDX6</td>
<td>LPFPIDDR</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>PRDX6</td>
<td>PRDX6</td>
<td>LSILYATTGR</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>RHI1</td>
<td>RNI</td>
<td>VNPALENLNR</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>SERPIN1</td>
<td>PDEF</td>
<td>LQSLFDSPDFS</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
</tr>
<tr>
<td>SERPIN1</td>
<td>PDEF</td>
<td>LSHEVETK</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
</tr>
<tr>
<td>SFN</td>
<td>1433S</td>
<td>NLSVAYK</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>TF</td>
<td>TF</td>
<td>YLGEEYVK</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
</tr>
<tr>
<td>TGM4</td>
<td>TGM4</td>
<td>FVFSVEVGDR</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
<tr>
<td>TGM4</td>
<td>TGM4</td>
<td>LEFSTGPNPSIAK</td>
<td>↓↓↓↓↓↓↓↓↓↓↓</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
<td>↑↑↑↑↑↑↑↑↑↑↑</td>
</tr>
</tbody>
</table>

*Direct-EPS measurements performed by spectral counting from Kim et al., 2012 MCP
EPS-urine measurements performed by SRM-MS from Cohort A of the current study
EPS-urine measurements performed by SRM-MS from Cohort B of the current study
EPS-urine measurements performed by Western blot from Kim et al., 2012 MCP
EPS-urine measurements performed by SRM-MS from Cohort A of the current study
EPS-urine measurements performed by SRM-MS from Cohort B of the current study
EPS-urine measurements performed by Western blot from Principe et al., 2012 JPR
4.4 Discussion and future direction

The timely verification of exhaustive lists of candidate disease biomarkers is becoming ever more relevant as high-throughput technologies enable in-depth analyses of complex biological systems exposing a plethora of putative candidates. Despite this, the implementation of biomarkers into clinical practice is largely lagging behind. This is, in part, attributed to the lack of validated methods of candidate testing and further evaluation. Targeted proteomics, specifically SRM-MS, has placed itself at the forefront of candidate protein verification, due to its relatively low cost and high-throughput assay development workflow. Therefore, the aim of the current study was to systematically develop a targeted assay and apply it for the quantitative verification of prostate cancer biomarkers in a sample that is routinely collected in the clinic.

Here, I developed a single multiplexed SRM-MS assay for 34 promising prostate cancer diagnostic and prognostic candidates in EPS-urine. This work is an extension of previously published shotgun proteomics data demonstrating the differential expression of a large number of protein candidates in direct-EPS from patients with EC and OC tumors (described in Chapter 3) (244). The feasibility of obtaining coordinates (peptide elution times, most intense transitions) from different modes of data acquisition and instrumentation was demonstrated by extracting relevant information from discovery data on a LTQ-Orbitrap in order to lay the foundation for targeted assay development using a triple quadrupole mass spectrometer.

By utilizing semi-quantitative data (spectral counting) and spectral libraries generated from Chapter 3, the extracted coordinates and most abundant transitions were identified and systematically evaluated to develop a refined SRM-MS assay for differentially expressed proteins. By performing multiple rounds of assay refinement (Phases 2 and 3), I was able to narrow the list of promising candidates from over 200 candidates to 34 peptides with high diagnostic and prognostic potential. By using heavy isotope-labeled peptides standards, I was able to estimate and later absolutely quantify the candidates in the EPS-urine background.

Interestingly, a very small number of peptides in EPS-urine overlapped with those that were previously explored in urine by Hüttenthal and colleagues (253). The cancer associated proteins that were evaluated in their study were gathered from reports based on protein and nucleic acid changes in human plasma and tissue (253, 359). In the current study, the evidence was directly
derived from comparing EC and OC tumor groups and from prostate-proximal fluids. I have demonstrated in Chapter 2 that in pre- and post-DRE urines from the same patients, post-DRE urines (EPS-urines) indeed contain a unique subset of proteins that are exclusive to this proteome (127).

Nonetheless, there were some potential caveats that were faced in the design of the current study. The discovery and subsequent Cohort A of samples were obtained from patients with either EC or OC tumors based on pathology. This is in contrast to the samples used in the verification phase (Cohort B) where patients were stratified based solely on their pathological stage (pT3 versus pT2). There are slight differences in their designations – the major difference being that the EC and OC designations were primarily determined on the basis of pathological stage and margin status (for instance, a pT2c tumor with positive margins and a pT3 tumor without positive margins would be considered as EC cases). In order to evaluate the effects of these differences, the expression profiles of candidate peptides in Cohorts A and B of samples were evaluated (Appendix Figure 4.2). For the majority of peptides, Cohorts A and B were highly concordant in terms of differential expression between pT3 versus pT2, and EC versus OC, with the exception of a few proteins (IDHC, ANXA1, K2C8, PEDF, FIBG, and CFAB). When Cohort A samples were converted to pT3 and pT2 designations, only GELS was affected. It is unclear whether this effect is due to the difference in patient stratification, patient-to-patient variability, or technical issues. However, the current evidence suggests that there is no major confounding effect of patient stratification based on EC versus OC or pT3 versus pT2.

The overall relative expression profiles of the Phase 3 candidates across different EPS datasets generated in our lab is summarized in Table 4.4. The first dataset was described in Chapter 3 and in Kim et al., (244), and was the basis for candidate selection in the current study (Figure 4.1, Discovery stage). Here, the semi-quantitative method of spectral counting was used to estimate the relative abundance of proteins in direct-EPS from EC and OC tumors. In the same study, immunoassays (i.e. Western blots) targeting a small number of proteins were performed in pools of EPS-urine from EC and OC tumors. This data was compared to the expression changes of the peptide candidates in Cohorts A and B of EPS-urines analyzed in the current study. Encouragingly, the majority of candidates showed similar changes in direct-EPS and EPS-urine, supporting the workflow that was used for candidate selection. Nonetheless, some candidates in the EPS-urine did not reflect the trends observed in direct-EPS. This could be a result of patient
heterogeneity, instrumentation, and quantitative methods. Even when comparing the expression levels of the candidates between EPS-urines from Cohorts A and B, there are some differences. Interestingly, some pairs of peptides that represent a single protein demonstrate opposing trends. For instance, the protein ANXA3 has one peptide (LTFDEYR) that is upregulated in the EC group of EPS-urines in Cohort A, while the other peptide (SEIDLLDIR) is downregulated in the EC group. This could be due to inaccurate quantification due to the use of crude peptide standards of varying purities, to the presence of proteoforms (341) and modifications, or variations in proteolytic digestion efficiencies of endogenous proteins (235, 360).

An interesting finding from the work presented in this chapter is the trend of lower abundance of the majority of candidates with advancing disease. For instance, elevated serum levels of PSA are indicative of prostate cancer; however, EPS levels of PSA have been consistently lower in the EC group compared to the OC group, and in the pT3 group compared to the pT2 group (244). Similarly, a trend of decreased PSA in EPS-urine from cancer patients compared to noncancer individuals has also been noted by Drake et al. (255). This may be indicative of PSA leakage out of the prostate gland and into the circulation, as described in Chapter 1 (129). Similarly, it can be hypothesized that other proteins are leaking from the prostate and into the circulation with the deteriorating structural integrity of advanced prostate cancers. It would be beneficial to measure both EPS and serum levels of such proteins from the same patient, in order to test this hypothesis.

In the comparison between BPH and normal individuals from Cohort B, only one protein (IDHC) was found to demonstrate a significant difference in abundance. The remaining peptides were found to show almost no differential expression. This may be explained by the fact that the discovery cohort from Chapter 3 (244) lacked a noncancer group consisting of BPH and normal individuals. Furthermore, the patients in Cohort A were not stratified on the basis of BPH versus normal designations, and hence candidates that were selected for verification were not specifically aimed at verifying their differential expression between BPH and normal EPS-urines.

Currently, the analyses of the data presented in this chapter are ongoing. In collaboration with Dr. Paul Boutros (OICR), a multi-feature diagnostic and prognostic signature is currently being generated using machine learning approaches. The features that are being taken into
consideration include each of the 34 peptides that were measured in the 207 heterogeneous verification sample cohort, as well as the clinical parameters that are available. For instance, serum PSA levels that were taken at the time of sample draw may be combined with a protein signature to provide enhanced prognostic performance of the signature. Although not obtained for the current study, additional parameters such as tumor microenvironment parameters may also provide information about patient outcome. Indeed, in a recent study by Lalonde et al., intra-prostatic hypoxia was combined with DNA indices (copy number alterations) to predict 5-year biochemical recurrence (BCR) (144). Such multi-feature signatures may provide a comprehensive, powerful predictor of patient outcome. Patients enrolled in Cohort B of this study have clinical information regarding BCR. In the pT2 group, 11 individuals out of 61 developed BCR at least 2 years post-RP; in the pT3 group, 10 individuals out of 29 developed BCR (Table 4.3). Although not yet explored in this study, one approach to the analysis of the data may be to comparatively analyze the protein expression changes between individuals who develop recurrence and those that did not.

4.5 Materials and Methods

4.5.1 Sample collection and properties

Samples were obtained from patients following informed consent and use of Institutional Review Board approved protocols at Urology of Virginia and Eastern Virginia Medical School and the Research Ethics Review Board at the University Health Network. Twenty ml of EPS-urine was centrifuged at 1100 g for 15 minutes at 4 °C to pellet debris. The resulting EPS-urine was aliquoted in volumes of 3.5 ml and diluted with 2.5 ml of PBS (pH 7.4). The mixture was vortexed, combined, and centrifuged at 1100 g for 15 minutes. The resulting supernatants were stored at -40 °C. The clinical information for patients enrolled in the study are detailed in Appendix Tables 4.4 and 4.5. Prostate cancer patients were selected on the basis of pathological stage pT2 or pT3. Noncancer individuals had biopsy confirmed BPH or were considered as individuals with no indication of prostatic disease based on biopsy results.
4.5.2 Sample preparation for mass spectrometry

Ultrapure-grade tetrafluoroethylene (TFE), trifluoroacetic acid (TFA), iodoacetamide (IAA), and dithiotreitol (DTT) were from Sigma-Aldrich. HPLC-grade solvents (methanol, acetonitrile, and water) and formic acid were from Fisher Scientific. Mass spectrometry-grade trypsin/lys-c was from Promega (Madison, WI). Amicon spin filter columns, 0.5 ml, 3 kDa MWCO, were from Millipore. Solid phase extraction C18 tips were from Agilent.

Four ml of urine and EPS-urine were concentrated to approximately 500 μl by using a spin column with a molecular weight cutoff of 3 kDa, and proteins were precipitated overnight by the addition of ice-cold 100% methanol. Protein pellets were washed twice with 100% methanol and air dried. Protein resolubilization and extraction was performed by the addition of 50% TFE at 60 °C for 2 hours. Following reduction with DTT and alkylation with IAA, proteins were digested overnight at 37 °C using mass trypsin/lys-c. The reaction was quenched by the addition of TFA. Desalting was performed by solid phase extraction using C18 tips. Solvents were removed by vacuum centrifugation and peptides were resolubilized in 5% acetonitrile, 0.1 % formic acid. Peptide concentrations were determined by the micro BCA assay kit (Thermo Fisher Scientific).

4.5.3 Target acquisition by selected reaction monitoring mass spectrometry

Samples were analyzed on a TSQ Vantage triple quadrupole mass spectrometer (Thermo Fisher Scientific) equipped with an EASY-Spray (Thermo Fisher Scientific) electrospray ion source. Separations were performed on EASY-Spray columns (15 cm x 75 μm ID packed with 3μm C18 particles, Thermo Fisher Scientific) heated to 50 °C. Peptides were kept at 4° C and loaded onto the column from an EASY-nLC (Thermo Fisher Scientific) autosampler. Chromatographic conditions were as follows: 40 minute gradient at a flow rate of 300 nl/min starting with 100% A (water), stepping up to 5% B (ACN) in 5 minutes, followed by 25% B at 35 minutes, followed by a steep increase to 50% B at 38 minutes and 100% B at 40 minutes.

Targeted acquisition of eluting ions was performed by the mass spectrometer operated in SRM-MS mode with Q1 and Q3 set to 0.7 m/z fwhm resolution and a cycle time of 1 second. For all
SRM-MS runs with the exception of the measurement of Phase 3 peptides EPS-urines from Cohort B, unscheduled methods were used, each targeting approximately 200 transitions over a 1 second cycle time. For Cohort B, a single scheduled method was utilized (2 minute elution window), monitoring approximately 200 transitions (3 transitions per peptide) over a 1 second cycle time.

4.5.4 Phase 2 peptide selection

In Chapter 3, 133 differentially expressed proteins were identified when comparing the proteome profiles of 16 direct-EPS samples from individuals with EC and OC prostatic tumors. A spectral library was built from the resulting LTQ-Orbitrap XL data obtained using the Skyline software tool (version 2.1.0) (350). All spectra had scores that passed a stringent peptide score as determined by an X!Tandem target decoy search criteria of 0.5% decoy spectral matches. Protein sequences were converted to FASTA format and uploaded into Skyline for the prediction of proteotypic peptides. Peptides were chosen based on previously reported specifications (241). A total of 232 proteotypic peptides (Phase 1 peptides) were selected and purchased as bulk heavy-isotope labeled peptide standards (JPT Peptide Technologies). In order to assess the suitability of the Phase 1 peptides for SRM-MS, 250 fmol of each heavy peptide standard was spiked into 1 ug of EPS-urine-derived peptides, with 4-6 transitions monitored over a 40-minute chromatographic gradient. Of the 232 Phase 1 peptides, 147 (Phase 2 peptides) were reproducibly detectable with a minimum of three transitions in the complex EPS-urine background.

4.5.5 Phase 2 peptide quantification and Phase 3 peptide selection

A cohort of individual EPS-urines (n = 74) from a heterogeneous population of patients with EC, OC, and control (BPH, normal) (Cohort A) was used to analyze the Phase 2 peptides. A total of 1 µg of peptide from each sample was spiked with 200 fmol of the heavy peptide standards. Visualization and inspection of peaks was performed on Skyline. Each peptide was quantified in a sample by integrating the quantifier ion (most intense ion) of the light peptide with its co-eluting heavy peptide ion, in order to derive a light-to-heavy peptide ratio. The student’s T-test was used to compare the ratios between cancer and controls, as well as EC and OC. A K-fold cross-validation was performed to investigate the diagnostic and prognostic power of the peptides at different p-value cutoffs. For the diagnostic and prognostic peptide candidates, p-
value cutoffs of 0.05 and 0.1 were used, respectively. Further refinements to the lists were made by including peptides that did not meet the p-value cutoffs but were potentially promising. For instance, peptides SSEDPNEDIVER from protein IGJ and TPAQFDABLELR from protein ANXA1 were added to the list of putative prognostic candidates because they were the only candidates that were elevated in the EC tumor group (p-value = 0.25). Two KLK3 peptides (HSQPWQVLVASR and LSEPAELTDAVK) were also added in order to monitor PSA levels in EPS-urine. A total of 34 peptides (24 diagnostic and 14 prognostic peptides, of which 4 overlapped) were taken forward for verification. These peptides are henceforth referred to as Phase 3 peptides.

4.5.6 Generation of response curves and determination of figures of merit

Each of the Phase 3 peptides were obtained as highly purified, soluble light and heavy peptides at defined concentrations (AQUA peptides, Thermo Fisher). A reversed response curve was generated for each peptide by spiking various concentrations (0-250 fmol, 7 points) of the heavy peptide into a background of 1 μg urine peptides and 50 fmol of the corresponding light peptide. Each concentration point was measured in 3-5 technical replicates. The LOD and LLOQ for each peptide were derived using the following equation: \( \text{LOD or LOQ} = \frac{F \times SD_{blank}}{b} \), where \( F = 3.3 \) for LOD and 10 for LOQ and \( b = \text{slope} \).

4.5.7 Multiplexed, scheduled selected reaction monitoring mass spectrometry method

In order to increase throughput, a multiplexed SRM-MS assay was developed by scheduling all 34 candidates in a single 40-minute chromatographic gradient. A total of 3 transitions were monitored for the light and heavy versions of each peptide for a total of 204 transitions per analysis. A 2-minute acquisition time window was scheduled around the expected peptide elution time.

4.5.8 Verification of Phase 3 peptides in EPS-urines

The verification a heterogeneous population of patients with pathological stage pT3 and pT2 prostate tumors, BPH, and normal individuals (n = 207) was enrolled (Cohort B). One μg of total
peptide from each sample was spiked with 100 fmol of heavy peptide and 10 fmol of corresponding light peptide for all candidates with the exception of the KLK3 peptide, HSQPWQVLVASR, which was spiked in at 500 fmol of heavy peptide. Visualization and inspection of peaks was performed on Skyline. Each sample was analyzed in two technical replicates using the same instrument parameters as described above.

4.5.9 Quantitative and statistical analyses

Each of the light and heavy peptides were checked for quality of data by observing co-elution of all three transitions, alignment of light and heavy peptide elution times, and reproducibility between technical replicates. Relative ratios of the area under the curve of the most predominant ion (quantifier) of the light peptide versus the corresponding heavy quantifier were calculated. For statistical analyses, the SRMstats software tool (version 1.6) (358) was utilized. Peptides were comparatively tested for abundance differences between Cancer versus Normal; Cancer versus BPH; Cancer versus Noncancer; BPH versus Normal; and pathological stages pT3 versus pathological stage pT2. Peptides with a p-value \( \leq 0.05 \) and a fold change \( \geq 1.5 \) were considered significant.
Chapter 5
Summary and future direction

5 Summary and future direction

5.1 Summary of key findings

In the current thesis, I used a proteomics platform encompassing the systematic discovery and verification of putative protein biomarkers in prostate-proximal fluids. Overall, this body of work provides a high-quality archive of proteins in expressed prostatic secretions (EPS). In addition, proteins that are differentially expressed among subtypes of prostate cancer – those that are extracapsular (EC) and those that are organ-confined (OC) – were identified. By systematically following up on these proteins, a number of peptides were verified to be differentially abundant among these prostate cancer subtypes and a quantitative assay was developed. Ongoing bioinformatics and statistical analyses are being used to generate a multi-feature signature. These findings may facilitate the clinical management of prostate cancer by identifying patients who may be spared from unnecessary medical procedures and have favourable prognoses, circumvent the need for biopsies, and may be used for longitudinal tracking of disease progression.

Furthermore, proteomic analyses of additional EPS-urine-derived exosomes (126) were performed during my candidacy, but were not covered in depth in this thesis. Taken together, over 2000 proteins from prostate-proximal fluids have been comprehensively catalogued, with an emphasis on the identification of prostate cancer biomarkers (Figure 5.1). A full database of these datasets is available and can be further interrogated to answer various clinical questions.

5.1.1 In-depth proteomic analyses characterizing the direct and urinary expressed prostatic secretion proteomes

In the first investigation into the EPS proteome in our lab, myself and colleagues analyzed direct-EPS from nine prostate cancer patients (260). Using multidimensional protein identification technology (MudPIT) and spectral counting, 916 unique proteins were identified. This was integrated with publicly available datasets, such as mRNA expression levels across various human tissues, antibody availability, other proteomics datasets from related fluids and prostate cancer cell secretomes, in order to characterize this proteome.
To describe the EPS-urine proteome, myself and colleagues identified 1022 unique proteins from EPS-urines from prostate cancer and benign prostatic hyperplasia (BPH) patients (127). Additionally, in order to identify prostate-enriched proteins, five patients with prostate cancer and five noncancer individuals were screened pre- and post-digital rectal examination to obtain urine and EPS-urine, respectively. This valuable, internally controlled sample set provided a total of 444 unique proteins of which 49 proteins were significantly enriched in EPS-urine compared to urine using a semi-quantitative method.

5.1.2 Discovery phase: identification of candidate biomarkers of extracapsular prostate cancer

In the discovery phase, a shotgun approach was taken to catalogue the proteomes of a prostate-proximal fluid, direct expressed prostatic secretions (direct-EPS), from EC and OC prostate cancers. In total, over 600 proteins were identified, of which 133 were differentially abundant in the two groups by spectral counting. In an initial verification step, a systematic data mining strategy was used to select a small subset of differentially expressed proteins. These 14 candidates were then analyzed by Western blotting in a cohort of direct-EPS linked with clinical outcome information related to biochemical recurrence (BCR) within a two-year period post-prostatectomy. Five candidates – SFN, MME, PARK7, TIMP1, TGM4 – showed a similar pattern of differential expression between recurrent and non-recurrent patients as they did for EC and OC, respectively. As one of the goals of my thesis was to demonstrate the detection of potential candidates in a clinically relevant biofluid for non-invasive detection, EPS-urine was obtained from EC and OC prostate cancers. As a proof-of-concept, selected reaction monitoring mass spectrometry (SRM-MS) was employed for quantification of these candidates. By targeting at least one proteotypic peptide, all candidates were quantified in the pooled EPS-urines by SRM-MS. In conclusion, I established EPS as a promising resource of prostate cancer biomarkers. Furthermore, I demonstrated SRM-MS as a valid method of rapid verification of putative candidates in biological fluids.

5.1.3 Verification phase: targeted quantification of selected prostate cancer biomarker candidates

The successful application of SRM-MS to EPS-urines in the discovery phase prompted me to utilize this high-throughput method for the accurate quantification of a larger number of
biomarker candidates. By revisiting the 133 differentially abundant proteins determined by spectral counting, 147 proteotypic peptides corresponding to 82 proteins, were selected and relatively quantified by SRM-MS in a cohort of individual EPS-urines from EC and OC patients, as well as noncancer control individuals. From this analysis, 34 candidates were demonstrated to have diagnostic and prognostic potential. Next, to verify these candidates, absolute quantification was performed utilizing highly purified peptide standards and a single multiplexed SRM-MS assay in a new cohort of EPS-urines. This high-quality information is being utilized to generate a multi-feature diagnostic and prognostic signature with a high potential for facilitating the clinical management of prostate cancer patients.

5.1.4 Additional proteomics analysis of prostate-proximal fluids

Although not covered in detail in this thesis, I carried out an additional proteomics investigation into EPS-urine derived exosomes (126). In this study, myself and colleagues analyzed EPS-urine-derived exosomes from prostate cancer patients and noncancer individuals using mass spectrometry (126). A total of 877 unique proteins were identified. Characterization of this proteome revealed a fraction of proteins that were not previously detected in the EPS (direct-EPS and EPS-urine) proteome or other exosome proteomes.

Taken together, I comprehensively characterized the proteomes of various prostate-proximal fluids. Through this, I demonstrated that these biofluids are rich sources of prostate-related proteins and may be a valuable resource of potential disease biomarkers (Figure 5.1).
Figure 5.1 Overview of all proteomics datasets generated for various prostate-proximal fluids. A total of approximately 2000 proteins are available in an in-house database. PCa, prostate cancer; NC, noncancer; EC, extracapsular; OC, organ-confined; BPH, benign prostatic hyperplasia.

5.2 Future directions

The data presented in this thesis provides verified proteomic information to generate a potential non-invasive biomarker panel for the prognosis of prostate cancer. The studies fall into Phases 1 (preclinical exploratory studies) and 2 (clinical assay development) out of the 5 phases of the biomarker development pipeline as set out by the Early Detection Research Network (112). The results fulfill the primary aims of Phase 1 by identifying and prioritizing potential biomarkers in a systematic manner, and providing quantitative measures of protein abundance changes between subtypes of prostate cancers. In Phase 2, an assay is developed for a non-invasive specimen. Its primary aim is to assess the clinical characteristics of the assay (such as its sensitivity and specificity) and its ability to distinguish patient groups. In partial fulfillment of this, a single multiplexed SRM-MS assay was systematically developed and utilized to generate a biomarker signature.
Following the work presented in the current thesis, a number of critical steps need to be undertaken in order to further validate the biomarkers for clinical use. First, the verified differentially expressed peptides are currently being statistically evaluated to generate a multi-feature signature for the non-invasive diagnosis and prognosis of prostate cancer. Following this, a multi-institutional evaluation of the signature should be performed. For this, standardized procedures for the harmonization of sample collection, processing, and storage need to be instituted. The SRM-MS assay should be optimized for a clinical laboratory setting and standard operating procedures should be drafted. Once implemented, the intra- and inter-laboratory reproducibility of the standardized assay, and its analytical robustness must be validated. Furthermore, the biomarkers presented in the thesis are continuous variables; thus clinically relevant thresholds and time points that can accurately stratify patients need to be established and validated (361). One of the major advantages of utilizing EPS-urine is the ease with which the specimens can be drawn from patients. Therefore, the longitudinal collection of EPS-urine at regular intervals should be performed in order to monitor changes in the levels of biomarkers and their relationship to disease progression. Prospective and cancer-control studies should be performed in large patient cohorts in order to determine the clinical utility of the biomarkers. The primary aim of a prospective study is to determine the detection rate and the false referral rate of the biomarker in a relevant population (112). In cancer-control studies, the aim is to estimate the reduction in cancer burden in the population as a result of testing. Factors that may inhibit a biomarker from being taken into the clinic include morbidity associated with the test, poor compliance, and over-diagnosis of cancers that, if left undetected, would pose no threat to patients (112).

One area that was not investigated in this thesis is the biological functions of putative candidates and the molecular mechanisms that they are involved in. In Chapter 3, it was shown that a number of different pathways are enriched in my dataset, including those that regulate the cell cycle and metabolism (Figures 3.9 and 3.10). These and other candidates can be followed up by studying the effects of silencing or over-expressing them in prostate cancer cell lines. Characteristics such as proliferation, apoptosis, migration, and invasiveness can be evaluated using methods such as proliferation and colony formation assays, annexin V/propidium iodide apoptosis assay, wound healing, and Matrigel invasion assays. Furthermore, by integrating
genomic, transcriptomic, and epigenetic information into the analysis, a systems level view of prostate cancer biology might be uncovered.

Our laboratory has been cataloguing the proteome of EPS for over 6 years (126, 127, 244, 260), having identified over 2000 proteins in total (Figure 5.1). Advancing techniques are expected to provide even deeper insight into this immensely complex, yet highly valuable class of biofluids. Subfractionation, by extracting exosomes or glycoproteins, for instance, is an active area of investigation for the identification of novel secreted biomarkers and molecular players in disease progression. Strategies that enrich for exosomes (e.g., differential ultracentrifugation) and glycoprotein capture (e.g., solid phase extraction of glycopeptides) greatly enhance the dynamic range and limit of detection of proteins in highly complex patient specimens.

Extracellular vesicles are small (40-5000 nm diameter) membrane-bound vesicles that carry biomolecular cargo (proteins, nucleic acids, and lipids) from their cell of origin (178, 362), and are involved in a myriad of biological processes (179, 363-367). Cancer cells are reported to increase the release of extracellular vesicles into the extracellular environment compared to noncancer cells (368-370), and thus these factors may represent an important source of fluid-derived cancer biomarkers. Indeed, studies have demonstrated the presence of “prostasomes” – prostate-derived extracellular vesicles – in body fluids, cancer cell lines, and tissue (126, 371-373). In our lab, myself and colleagues have characterized the proteomes of EPS-urine-derived exosomes using shotgun proteomics (126). Almost 900 proteins were identified, of which approximately 200 were not previously detected in direct-EPS and EPS-urines analyzed in our lab.

Protein glycosylation, the post-translational modification of proteins by the covalent addition of sugar moieties, plays a fundamental role in a wide range of biological processes (374, 375). Indeed, alterations in protein glycosylation patterns is a recognized hallmark of cancer progression (376, 377). Profiling the changes that occur in this subproteome as a result of disease can provide novel biomarkers for diagnosis and prognosis of cancer. The majority of membrane and secreted proteins are glycosylated (378); in fact, many of the clinical biomarkers in use today, such as PSA in prostate cancer, CA125 in ovarian cancer, and Her2/neu in breast cancer, are glycoproteins.
As a continuation of the work presented in this thesis, an in-depth proteomic profiling of such EPS subproteomes will provide additional putative biomarker candidates that were not previously identified. As mentioned above, by enriching for these proteins, low abundance species may be more readily resolved.

Proteomics technologies are continuously evolving to provide even deeper proteome coverage, better quantitation, and higher throughput. The data independent acquisition (DIA) mode of mass spectrometry analysis has recently emerged as a powerful tool for simultaneous peptide identification and quantification (379). In DIA mode, all peptides in a mass-to-charge (m/z) window of a defined width and a retention time range are fragmented, and MS/MS data is acquired from all detectable precursor ions in the isolation window. This process is repeated for the next m/z window in an iterative fashion across the full m/z range, producing a complete and permanent time-resolved record of all fragment ions of detectable peptide precursors in a sample (380). With recent improvements in DIA data analysis platforms, the highly complex, composite fragment ion spectra can be linked back to the precursors from which they originate using a priori information contained in MS/MS spectral libraries (381). This method increases coverage and reproducibility of the data by acquiring MS/MS data from all detectable precursor ions in a sample, while making retrospective data analysis a possibility.

Some of the limitations of SRM-MS assays can be circumvented by a high-resolution, accurate-mass mass spectrometer. For instance, the Thermo Scientific Q Exactive is a hybrid quadrupole-Orbitrap mass spectrometer that can be utilized for sensitive targeted quantification of proteins in complex biological material. In parallel reaction monitoring (PRM) mode, a target precursor ion is isolated in a quadrupole, fragmented by high energy collisional dissociation, and all resulting product ions are detected in an Orbitrap mass analyzer, yielding a MS/MS spectrum (382, 383). Quantification is achieved post-acquisition, by extracting one or more of the fragment ions. Peterson and colleagues demonstrated similar analytical performance of PRM to that of SRM-MS in terms of dynamic range, linearity, and precision (382). The major advantage of PRM over SRM-MS is the improved selectivity by recording all fragment ions in the MS/MS scan, over the few transitions recorded in SRM-MS. In this way, the assay development time is also reduced since there is no need to preselect target transitions.
5.3 Concluding remarks

Personalized medicine, a field of health care that utilizes each individual’s unique clinical, genetic, genomic, and environmental information to optimize and guide clinical decision making (384), is emerging as a feasible and promising modality of cancer management. This is largely due to tremendous progress made in genomic medicine that encompasses information from genomes and their cognate factors such as RNA, proteins, and metabolites. This information is helping researchers navigate the complex landscape of tumor biology and providing a comprehensive view, enabling more precise risk assessment, diagnosis and prognosis, and response to therapy (385). For example, protein expression patterns, such as those that were presented in this thesis, can definitively classify patients into different prognostic subgroups, facilitating strategic health planning. Continued scientific efforts are expected to lead the way in personalized medicine and shift the paradigm of healthcare from disease treatment to disease prevention (384).
References


microdissected prostate carcinomas reveals a high frequency of allelic loss on chromosome 8p12-21. *Cancer research* 56, 2411-2416


insignificant disease using free/total prostate specific antigen levels and needle biopsy findings. *J Urol* 160, 2407-2411


proteomic and bioinformatic analyses to identify putative biomarkers. *J Proteome Res* 7, 339-351


