PROTEOMIC ANALYSIS OF PROSTATE CANCER CELL LINE CONDITIONED MEDIA FOR THE DISCOVERY OF CANDIDATE BIOMARKERS FOR PROSTATE CANCER

by

Girish Sardana

A thesis submitted in conformity with the requirements for the Degree of Doctor of Philosophy
Graduate Department of Laboratory Medicine and Pathobiology
University of Toronto

© Copyright by Girish Sardana 2008
PROTEOMIC ANALYSIS OF PROSTATE CANCER CELL LINE CONDITIONED MEDIA FOR THE DISCOVERY OF CANDIDATE BIOMARKERS FOR PROSTATE CANCER

Girish Sardana

Doctor of Philosophy 2008

Department of Laboratory Medicine and Pathobiology

University of Toronto

ABSTRACT

Early detection of prostate cancer is problematic due to the lack of a marker that has high diagnostic sensitivity and specificity. The prostate specific antigen test, in combination with digital rectal examination, is the gold standard for prostate cancer diagnosis. However, this modality suffers from low specificity. Therefore, specific markers for clinically relevant prostate cancer are needed. Our objective was to proteomically characterize the conditioned media from human prostate cancer cell lines to identify secreted proteins that could serve as novel prostate cancer biomarkers.

An initial proof of principle study of the PC3 prostate cancer cell line was conducted. From this study over 200 proteins were identified in the conditioned media. Through gene ontology analysis and literature searches Mac-2 binding protein was selected as a candidate biomarker for validation in the serum of prostate cancer patients. A preliminarily validation showed that Mac-2 binding protein has
discriminatory ability in prostate cancer diagnosis. However, an extended validation did not confirm this.

Based on our proof of principle study we optimized our workflow and extended our analysis by culturing three different prostate cell lines [PC3 (bone metastasis), LNCaP (lymph node metastasis), and 22Rv1 (localized to prostate)]. We conducted a bottom-up analysis of each cell line by 2-dimensional liquid chromatography and tandem mass spectrometry. Of the 2124 proteins identified, 12% (329) were classified as extracellular and 18% (504) as membrane-bound. Among the identified proteins were known prostate cancer biomarkers such as PSA and KLK2. To select the most promising candidates for further investigation, tissue specificity, biological function, disease association based on literature searches, and comparison of protein overlap with the proteome of seminal plasma and serum were examined. Based on these results, several candidates were selected for validation in serum of patients with and without prostate cancer. Of these four novel candidates: follistatin, chemokine (C-X-C motif) ligand 16, pentraxin 3 and spondin 2 showed discriminatory ability.

Of the four candidates, follistatin was further studied in an extended validation in serum of patients with biopsy confirmed prostate cancer and tissues of prostate cancer patients of low and high grade tumours by immunohistochemistry. In addition, follistatin was also investigated in the tissue of colon and lung cancer where intense staining was observed in one specimen of lung squamous carcinoma.
ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisor Dr. Eleftherios P. Diamandis for his guidance and vision for this project. It has been a maturing and character building experience and I am truly grateful for the opportunity to have been a part of the ACDC Lab.

I would like to also thank my advisory committee members Drs. Alex Romaschin and Pui-Yuen Wong for their continued review of my work throughout my project. As well as Drs. Robert Nam and Oliver John Semmes for their perspectives and constructive criticism on my thesis. I would like to give special thanks to Drs. John Marshall, Theodorus van der Kwast and Carsten Stephan for their collaboration on this project. I am also grateful to the following for providing the necessary resources to conduct my research: The Department of Laboratory Medicine and Pathobiology, The University of Toronto, The Samuel Lunenfeld Research Institute, Mt. Sinai Hospital and Proteomic Methods Inc.

There is a saying - “It’s the people that make the lab”. I would not have been able to experience the true meaning of graduate studies without the friends and colleagues that I met along the way. Special thanks to all the members of the ACDC Lab past and present.

Finally, I would like to thank my family for their support during my graduate studies.
# TABLE OF CONTENTS

| ABSTRACT | ii |
| ACKNOWLEDGEMENTS | iv |
| TABLE of CONTENTS | v |
| LIST of TABLES | viii |
| LIST of FIGURES | ix |
| LIST of ABBREVIATIONS | xi |

## CHAPTER 1: INTRODUCTION

1.1 Prostate Cancer

1.1.1 Epidemiology and Statistics 2
1.1.2 Anatomy and Histology of the Prostate 5
1.1.3 Pathobiology of Prostate Cancer 7
1.1.4 Current Methods for Diagnosis 10
1.1.5 Current Treatments 13

1.2 Cancer Biomarkers 19

1.2.1 Introduction to Cancer Biomarkers 19
1.2.2 Methods for Identifying Novel Cancer Biomarkers 20
    1.2.2.1 Genomic Approaches 21
    1.2.2.2 Proteomic Approaches 24
1.2.3 Clinical and Analytical Properties of a Cancer Biomarker 27
1.2.4 Applications of Cancer Biomarkers 29
1.2.5 Current and Emerging Prostate Cancer Biomarkers 31

1.3 Prostate Cancer Model Systems 51

1.3.1 Mouse Models 51
1.3.2 Cell Culture Models 52

1.4 Continued Need for Novel Prostate Cancer Biomarkers 54

1.5 Rationale and Objectives of the Present Study 55

    1.5.1 Rationale 55
    1.5.2 Hypothesis 57
    1.5.3 Objectives 57
CHAPTER 2: PROOF OF PRINCIPLE: PROTEOMIC ANALYSIS OF THE PC3 CELL LINE CONDITIONED MEDIA

2.1 Introduction 60

2.2 Materials and Methods 61
   2.2.1 Roller Bottle Cell Culture 61
   2.2.2 Dialysis 63
   2.2.3 Fast Performance Liquid Chromatography of CM 63
   2.2.4 Lyophilization and Digestion of Fractions 63
   2.2.5 Liquid Chromatography–MS Analysis 64
   2.2.6 Database, Genome Ontology, and Literature Search 65
   2.2.7 ELISAs for Kallikrein 5, 6, and 11 66
   2.2.8 Protein Recovery 66
   2.2.9 Mac-2BP ELISA 66

2.3 Results 67
   2.3.1 Total Protein, KLK5 and KLK6 Concentrations in Culture Over Time 67
   2.3.2 Recovery of KLK5 and KLK6 During Sample Preparation 69
   2.3.3 Proteins Identified by Mass Spectrometry 69
   2.3.4 Overlap of Proteins Identified in Batches 1 and 2 73
   2.3.5 Mac-2BP Concentrations in Patients with Prostate Cancer vs. Healthy Men and in the Conditioned Media of the PC3(AR)$_6$ cell line 74

2.4 Discussion 79

CHAPTER 3: OPTIMIZATION OF CELL CULTURE AND PROTEOMIC WORKFLOW

3.1 Introduction 83

3.2 Materials and Methods 84
   3.2.1 Cell Culture 84
   3.2.2 Measurement of Total Protein, LDH and PSA, KLK5 and KLK6 84
   3.2.3 Sample Preparation 85

3.3 Results 86
   3.3.1 Cell Culture Optimization 86
   3.3.2 Sample Preparation Optimization 90

3.4 Discussion 92
CHAPTER 4:
COMPARATIVE PROTEOMIC ANALYSIS OF THE CONDITIONED MEDIA OF THREE PROSTATE CANCER CELL LINES

4.1 Introduction

4.2 Materials and Methods

4.2.1 Cell Culture
4.2.2 Measurement of Total Protein, Lactate Dehydrogenase and PSA, KLK5, KLK6
4.2.3 Conditioned Media Sample Preparation and Trypsin Digestion
4.2.4 Strong Cation Exchange High Performance Liquid Chromatography
4.2.5 Online Reversed Phase Liquid Chromatography – Tandem Mass Spectrometry
4.2.6 Database Searching and Bioinformatics
4.2.7 Validation of Candidates

4.3 Results

4.3.1 Proteins Identified by Mass Spectrometry
4.3.2 Identification of Internal Control Proteins
4.3.3 Reproducibility between Replicates
4.3.4 Differences in Proteins Identified Between Cell Lines
4.3.5 Genome Ontology Distributions of Proteins
4.3.6 Secreted and Membrane Proteins
4.3.7 Overlap with Previous Data
4.3.8 Overlap with Seminal Plasma Proteins
4.3.9 Biological Network Analysis
4.3.10 Overlap with Breast Cancer Secretome
4.3.11 Validation of Follistatin, Chemokine (C-X-C motif) ligand 16, Pentraxin 3 and Spondin 2

4.4 Discussion

CHAPTER 5:
VALIDATION OF CANDIDATE BIOMARKERS IN PROSTATE CANCER PATIENT SAMPLES

5.1 Introduction

5.2 Materials and Methods

5.2.1 Conditioned Media, Serum Samples and Tissue Specimens
5.2.2 Quantification of candidates in biological fluids
5.2.3 Immunohistochemistry of Follistatin
5.2.4 Statistical Data Analysis
5.3 Results

5.3.1 Chemokine (C-X-C motif) ligand 5 136
5.3.2 Chemokine (C-X-C motif) ligand 1 138
5.3.3 Lipocalin-2 140
5.3.4 Tumour necrosis factor receptor superfamily, member 6b, decoy 142
5.3.5 Extended Validation of Follistatin 144

5.4 Discussion 151

CHAPTER 6: SUMMARY AND FUTURE DIRECTIONS
6.1 Summary 157

6.2 Key Findings 157

6.3 Future Directions 159

REFERENCES 161

APPENDIX: Assay Reproducibility and Precision 201
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>List of candidate biomarkers for prostate cancer and their possible</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>clinical utility.</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>Extracellular candidate tumor markers identified in culture medium</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>of the PC3 (AR)₆ roller bottle culture</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>Known prostate biomarkers identified in the conditioned media of</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>PC3, LNCaP, and 22Rv1 cell lines</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>List of candidate biomarkers selected based on selection criteria.</td>
<td>127</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Schematic representation of the workflow for proteomic analysis of</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>roller bottle CM</td>
<td></td>
</tr>
<tr>
<td>2.2</td>
<td>Measurement of KLK5, KLK6 and total protein</td>
<td>68</td>
</tr>
<tr>
<td>2.3</td>
<td>Cellular location of proteins from PC3(AR)6 CM</td>
<td>70</td>
</tr>
<tr>
<td>2.4</td>
<td>Mac-2BP concentrations and the correlation between serum Mac-2BP</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>concentrations and PSA</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>Concentrations of KLK5, KLK6, and KLK11 in serum of 26 CaP patients</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>and 17 healthy men</td>
<td></td>
</tr>
<tr>
<td>2.6</td>
<td>Serum concentrations of Mac-2BP in CaP, BPH and normal patients</td>
<td>78</td>
</tr>
<tr>
<td>3.1</td>
<td>Measurements of PSA and LDH in the CM of the 22Rv1 cell line</td>
<td>87</td>
</tr>
<tr>
<td>3.2</td>
<td>Measurements of PSA and LDH in the CM of the LNCaP cell line</td>
<td>88</td>
</tr>
<tr>
<td>3.3</td>
<td>Measurements of KLK5, KLK6 and LDH in the CM of the PC3 cell line</td>
<td>89</td>
</tr>
<tr>
<td>3.4</td>
<td>Overview of the optimized sample preparation workflow</td>
<td>91</td>
</tr>
<tr>
<td>4.1</td>
<td>Workflow of proteomic method employed</td>
<td>100</td>
</tr>
<tr>
<td>4.2</td>
<td>Overlap of the 3 replicates from PC3, LNCaP and 22Rv1 conditioned</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>media</td>
<td></td>
</tr>
<tr>
<td>4.3</td>
<td>Overlap of proteins identified between each cell line</td>
<td>110</td>
</tr>
<tr>
<td>4.4</td>
<td>Overlap of the extracellular and membrane proteins identified in</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>each cell line</td>
<td></td>
</tr>
<tr>
<td>4.5</td>
<td>Classification of proteins by cellular location</td>
<td>112</td>
</tr>
<tr>
<td>4.6</td>
<td>Classification of proteins by cellular location for each cell line</td>
<td>113</td>
</tr>
<tr>
<td>4.7</td>
<td>Molecular functions related to diseases associated with Follistatin</td>
<td>116</td>
</tr>
<tr>
<td>4.8</td>
<td>Molecular functions related to diseases associated with Chemokine</td>
<td>117</td>
</tr>
<tr>
<td></td>
<td>(C-X-C motif) ligand 16, Pentraxin 3 and Spondin 2</td>
<td></td>
</tr>
<tr>
<td>4.9</td>
<td>Concentrations of each candidate in the conditioned media of each</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>cell line and the control flask</td>
<td></td>
</tr>
<tr>
<td>4.10</td>
<td>Validation of Follistatin, Chemokine (C-X-C motif) ligand 16,</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>Pentraxin 3 and Spondin 2 in serum</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Measurement of CXCL5 in CM of CaP cell lines and serum of CaP</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>patients and normals</td>
<td></td>
</tr>
<tr>
<td>5.2</td>
<td>Measurement of CXCL1 in CM of CaP cell lines and serum of CaP</td>
<td>139</td>
</tr>
<tr>
<td></td>
<td>patients and normals</td>
<td></td>
</tr>
<tr>
<td>5.3</td>
<td>Measurement of LCN2 in CM of CaP cell lines and serum of CaP</td>
<td>141</td>
</tr>
<tr>
<td></td>
<td>patients and normals</td>
<td></td>
</tr>
<tr>
<td>5.4</td>
<td>Measurement of TNFRSF6B in CM of CaP cell lines and serum of CaP</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>patients and normals</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>Follistatin serum levels and correlations with PSA, Gleason score</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>and age</td>
<td></td>
</tr>
<tr>
<td>Figure</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.6</td>
<td>PSA, %fPSA, KLK2 and KLK11 serum concentrations in patient serum</td>
<td>147</td>
</tr>
<tr>
<td>5.7</td>
<td>Prostate tissue specimens stained for follistatin</td>
<td>148</td>
</tr>
<tr>
<td>5.8</td>
<td>Lung tissue specimens stained for follistatin</td>
<td>149</td>
</tr>
<tr>
<td>5.9</td>
<td>Colon tissue specimens stained for follistatin</td>
<td>150</td>
</tr>
</tbody>
</table>
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-DE</td>
<td>two-dimensional gel electrophoresis</td>
</tr>
<tr>
<td>A2M</td>
<td>α-2-macroglobulin</td>
</tr>
<tr>
<td>ACN</td>
<td>acetonitrile</td>
</tr>
<tr>
<td>ACT</td>
<td>α-1-antichymotrypsin</td>
</tr>
<tr>
<td>AIPC</td>
<td>androgen independent prostate cancer</td>
</tr>
<tr>
<td>AMACR</td>
<td>α-methylacyl CoA racemase</td>
</tr>
<tr>
<td>ANN</td>
<td>artificial neural networks</td>
</tr>
<tr>
<td>ANXA3</td>
<td>annexin A3</td>
</tr>
<tr>
<td>API</td>
<td>α1-protease inhibitor</td>
</tr>
<tr>
<td>AR</td>
<td>androgen receptor</td>
</tr>
<tr>
<td>AUC</td>
<td>area under curve</td>
</tr>
<tr>
<td>BPH</td>
<td>benign prostate hyperplasia</td>
</tr>
<tr>
<td>BPSA</td>
<td>BPH associated PSA</td>
</tr>
<tr>
<td>CaP</td>
<td>prostate cancer</td>
</tr>
<tr>
<td>CDCHO</td>
<td>chemically-defined Chinese Hamster Ovary</td>
</tr>
<tr>
<td>CGH</td>
<td>comparative genomic hybridization</td>
</tr>
<tr>
<td>cPSA</td>
<td>complexed PSA</td>
</tr>
<tr>
<td>CM</td>
<td>conditioned medium</td>
</tr>
<tr>
<td>CXCL16</td>
<td>chemokine (C-X-C motif) ligand 16</td>
</tr>
<tr>
<td>DiEtOH</td>
<td>diethanolamine</td>
</tr>
<tr>
<td>DRE</td>
<td>digital rectal exam</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>ELISA</td>
<td>enzyme linked immunosorbent assay</td>
</tr>
<tr>
<td>EPCA</td>
<td>early prostate cancer antigen</td>
</tr>
<tr>
<td>ESI</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>EZH2</td>
<td>enhancer of zeste homolog gene 2</td>
</tr>
<tr>
<td>FCS</td>
<td>fetal calf serum</td>
</tr>
<tr>
<td>fPSA</td>
<td>free PSA</td>
</tr>
<tr>
<td>FISH</td>
<td>fluorescence in situ hybridization</td>
</tr>
<tr>
<td>FPLC</td>
<td>fast-performance liquid chromatography system</td>
</tr>
<tr>
<td>FPR</td>
<td>false positive rate</td>
</tr>
<tr>
<td>FSH</td>
<td>follicle-stimulating hormone</td>
</tr>
<tr>
<td>GO</td>
<td>gene ontology</td>
</tr>
<tr>
<td>GSTP1</td>
<td>glutathione S-transferase pi</td>
</tr>
<tr>
<td>IGF</td>
<td>insulin-like growth factors</td>
</tr>
<tr>
<td>IGFBP</td>
<td>insulin-like growth factor binding proteins</td>
</tr>
<tr>
<td>IL-6</td>
<td>interleukin 6</td>
</tr>
<tr>
<td>iPSA</td>
<td>intact PSA</td>
</tr>
<tr>
<td>KLK1</td>
<td>pancreatic/renal kallikrein 1</td>
</tr>
<tr>
<td>KLK2</td>
<td>human kallikrein-related peptidase 2</td>
</tr>
<tr>
<td>LCM</td>
<td>laser-capture microdissection</td>
</tr>
<tr>
<td>LC-MS/MS</td>
<td>liquid chromatography-tandem mass spectrometry</td>
</tr>
<tr>
<td>LDH</td>
<td>lactate dehydrogenase</td>
</tr>
<tr>
<td>LH</td>
<td>luteinizing hormone</td>
</tr>
<tr>
<td>LHRH</td>
<td>LH releasing hormone</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Mac-2BP</td>
<td>mac 2 binding protein</td>
</tr>
<tr>
<td>MALDI</td>
<td>matrix assisted laser desorption ionization</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>MS/MS</td>
<td>tandem MS</td>
</tr>
<tr>
<td>NCBI</td>
<td>national center of biotechnology information</td>
</tr>
<tr>
<td>NGEP</td>
<td>new gene expressed in the prostate</td>
</tr>
<tr>
<td>NPV</td>
<td>negative predictive value</td>
</tr>
<tr>
<td>%fPSA</td>
<td>percent free PSA</td>
</tr>
<tr>
<td>PAP</td>
<td>prostatic acid phosphatase</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate-buffered saline</td>
</tr>
<tr>
<td>PCA3</td>
<td>prostate cancer antigen 3</td>
</tr>
<tr>
<td>PCPT</td>
<td>the prostate cancer prevention trial</td>
</tr>
<tr>
<td>PIN</td>
<td>prostatic intraepithelial neoplasia</td>
</tr>
<tr>
<td>PPV</td>
<td>positive predictive value</td>
</tr>
<tr>
<td>PSA</td>
<td>prostate specific antigen</td>
</tr>
<tr>
<td>PSAD</td>
<td>PSA density</td>
</tr>
<tr>
<td>PSAD-TZ</td>
<td>transitional zone PSA density</td>
</tr>
<tr>
<td>PSMA</td>
<td>prostate specific membrane antigen</td>
</tr>
<tr>
<td>PSAV</td>
<td>PSA velocity</td>
</tr>
<tr>
<td>proGRP</td>
<td>progastrin-releasing peptide</td>
</tr>
<tr>
<td>PSCA</td>
<td>prostate stem cell antigen</td>
</tr>
<tr>
<td>PSP94</td>
<td>prostate secretory protein 94</td>
</tr>
<tr>
<td>PSPBP</td>
<td>PSP94-binding protein</td>
</tr>
</tbody>
</table>
PTEN phosphatase and tensin homologue deleted from chromosome 10
PTX3 pentraxin 3
RB retinoblastoma
RT-PCR real-time polymerase chain reaction
ROC receiver operating characteristic curve
SAX strong anion exchange
SCX strong cation exchange
SELDI-TOF surface enhanced laser desorption ionization time of flight
SFM serum-free media
SILAC stable isotopic labelling with amino acids in cell culture
SPON2 spondin 2
TBST tris-buffered saline with tween 20
TGF-β1 transforming growth factor β1
TFA trifluoroacetic acid
tPSA total PSA
TMN tumour node metastasis
TRAMP transgenic adenocarcinoma of the mouse prostate
TRUS trans-rectal ultrasound
TZ transitional zone
uPA urokinase-plasminogen activation
CHAPTER 1:
INTRODUCTION
1.1  Prostate Cancer

The prostate is recognized for having a relatively high occurrence of tumours in men, especially older men. As a result, prostate cancer (CaP) is the most prevalent malignancy in men and is the second leading cause of cancer deaths in North America with one in six men having a lifetime risk of being diagnosed and a 3.4% chance of death due to CaP(1). Most patients currently are being diagnosed with early stages of the disease with no symptoms(2). As a result of this increase in early stage tumour diagnoses or stage migration, the classical ways of prognosis, such as the Partin Tables and Kattan Nomograms are no longer as effective. The focus has now moved from early detection to the determination of the clinical significance of these early stage tumours. One objective is to find ways of distinguishing clinically relevant tumours that have the ability to metastasize. Currently, 30% of tumours removed by radical prostatectomy are deemed clinically insignificant and would not have required such invasive treatment. Most patients diagnosed have a latent, non-aggressive form of CaP(3), thus it is important that these patients are not over-treated. Currently, little is known about the molecular pathogenesis of CaP(4). Control of CaP could be achieved through early detection and appropriate selection of treatment; however we have yet to reach this level of sophistication.

1.1.1  Epidemiology and Statistics

It is estimated that there will be over 500,000 new cases worldwide of CaP making it the sixth cancer in the world in terms of incidence. It is the most common
cancer in North America, Europe and parts of Africa. Current statistics from the Canadian Cancer Society show that 24,700 estimated new cases with 4,300 deaths in 2008(5). Before the age of 40, CaP is rarely diagnosed however its clinical incidence increases with age. The mean age of diagnosis of CaP is 72-74 with the majority of patients diagnosed after the age of 65. Prostate cancer incidence rates are expected to rise as life expectancies continue to increase. The rise in incidence rates since the mid 1990’s has been attributed to the implementation of the prostate specific antigen (PSA) test for population screening in Western countries. While incidence rates have been rising, mortality rates in Canada have been shown to decrease significantly by 2.9% from 1995 to 2004(5). While this has been presumed to be a result of screening and improved treatment there are two prospective clinical trials underway that are assessing this: the European Randomized study of Screening for Prostate Cancer and Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial.

While CaP claims the highest incidence rates among all cancers in men, this does not take into account the prevalence of small adenocarcinomas that are found incidentally upon autopsy or during other surgical procedures. The prevalence of these ‘latent’ tumours is thought to be over 50% in men over 50(6). This highlights that men are more likely to die with CaP rather than from it. This attribute of CaP makes it very difficult in its diagnosis where the high prevalence of clinically insignificant tumours could result in an increase in unnecessary treatment.

Geographically the highest incidence in the world of CaP is centered mainly in the United States and Scandinavian countries and the lowest in Mediterranean
countries and Japan(7). Within countries, CaP differs in ethnic groups. With respect to the US, African-American males harbour an increased incidence versus Caucasian, native Chinese and native Japanese(8). An increase in the incidence of CaP in Japanese males who move to Western countries suggests that life-style factors may play a role in the development of detectable CaP(9). Other factors that may account for these differences include access to health care and how CaP is reported.

Differences in genetic susceptibility are also attributed as being a predisposing factor for CaP. The chance for developing CaP increases by two to three times when a first degree relative is afflicted with the disease and increases when there are more affected individuals in a family, between 10-15% of CaP patients have a relative who is affected(10). Twin studies have also assessed the link between genetics and environmental factors showing a possible autosomal dominant inheritance of a high risk gene(11). Strong candidate genes for genetic association of CaP are those involved with androgen regulation(12) and metabolism(13).

Androgen hormones are required for the development of the normal prostate and have been linked to CaP progression. As a result anti-androgen therapies have proven to be of great use in treating CaP in 75-80% of patients with metastatic CaP. While testosterone and dihydrotestosterone have shown to induce CaP tumours there is very little epidemiological data to support androgen levels as a risk factor for CaP(14).
In addition to genetics and hormonal influences, nutritional factors have been shown to play a role in CaP incidence. Studies have shown that CaP is associated with the Western lifestyle with respect to a diet that contains high fat, red meat and dairy(15). Studies continue to show conflicting data but there are some consistently reoccurring components such as α-linolenic acid(16) and calcium(17). Conversely, protective effects of some diets such as those found in Asia have been attributed to high intakes of phyto-oestrogens from soybeans which have a preventative effect on CaP(18). Intake of lycopene from tomato based products has also been shown to reduce risk of CaP by acting as an antioxidant(19). However, the most promising data of nutritional prevention of CaP comes from selenium and vitamin E. Both are considered antioxidants that also enhance the immune system and induce apoptosis in cancerous cells. The value of these two nutrients in prevention of CaP is currently being studied in the Selenium and Vitamin E Cancer Prevention Trial (SELECT) and final results will be available by 2013(20).

1.1.2 Anatomy and Histology of the Prostate

The prostate is an accessory reproductive organ that consists of exocrine glands that are organized in a fibromuscular architecture surrounding the proximal urethra. It is located in the abdominal cavity inferior to the bladder where the urethra from the bladder passes through the prostate before entering the penis. The portion of the urethra in the prostate is termed the prostatic urethra. Connected to the prostate are two seminal vesicles which connect via two deferential ducts that originate from the epididymis. The seminal vesicles and deferential ducts connect to
the prostatic urethra via two ejaculatory ducts within the prostatic gland. Within the pelvis the prostate is surrounded by striated musculature as well as fatty tissue and neurovascular bundles.

The glandular tissue within the prostate can be divided into different regions anatomically. These are the five lobes of the prostate: the anterior, middle, posterior and two lateral lobes. The prostate can also be subdivided functionally into two zones. These are the centrally located transition zone, the posteriorly and laterally located peripheral zone. The transition zone is the main site for benign prostate hyperplasia (BPH) development while the peripheral zone is where malignant tumours are found(21).

The prostate produces substances that add to those produced by the seminal vesicles and the spermatozoa. Approximately 60% of semen consists of seminogelin – a protein produced by the seminal vesicles. Thirty percent is produced by the prostate which adds proteins and enzymes that affect the liquidity of the semen, and the remaining 10% are secretions from the bulbourethral glands that lie ‘downstream’ from the prostate. These secretions contribute to the protective liquid environment of the semen and for nourishment of the sperm.

The histology of the prostate is similar in all regions consisting of ducts and acini which are lined by a double epithelial layer. The luminal epithelial cells are cuboidal or columnar in shape and the basal epithelial cells are flattened. Below these two layers is the basal membrane. Throughout the gland are scattered neuroendocrine cells that lie above the basal cell layer. The glands are supported by a dense stroma of smooth muscle fibres and fibrous connective tissue which
exerts its main function during ejaculation when prostatic secretions are ejected together with semen.

The luminal cells in the ducts and acini produce proteins that become constituents of semen. These include enzymes such as prostatic acid phosphatase (PAP) and PSA. Prostate specific antigen is responsible for liquefying the seminal clot by digesting seminogelin and thus increasing the motility of sperm. Basal cell function is still not yet known but have been hypothesized to act as either stem-like cells to replenish the luminal epithelium or act as functional cell barrier between luminal cells and the stroma and are involved in growth and differentiation of the epithelium(22). Neuroendocrine cells in the prostate produce peptides such as chromogranin A, serotonin and progastrin-releasing peptide. The exact function of this is currently unknown, however rare tumours are derived from neuroendocrine cells and these peptides have been shown to be useful as tumour markers in their diagnosis(23,24).

1.1.3 Pathobiology of Prostate Cancer

Cancers including CaP have been shown to develop from accumulating genetic alterations that translate into increased cell growth and propensity to invade surrounding tissues. Morphologic alterations in prostate tissue histology are evident that indicate it is a multistep process that involve genetic mutations including deletions, amplifications, rearrangements and epigenetic modifications(25). Phenotypic changes during progression of CaP have been characterized to follow transition states from benign glands to the malignant phenotype. Initially, epithelial
cells are seen to progress to prostatic intraepithelial neoplasia (PIN) within benign glands. Here we begin to see dysplastic epithelial cells that can be categorized as high and low grade PIN(26). High grade PIN is considered a precursor to CaP however not all PIN lesions detected progress to CaP(27). It is hypothesized that tumour cells are themselves genetically instable and are thus prone to genetic mutations(28). Transformed cells that develop a survival advantage then proliferate over the surrounding cells and form a tumour.

Results from epidemiologic studies have shown a genetic susceptibility risk for men with a family history of CaP as well as through twin studies(29,30). From linkage analysis there have been several CaP susceptibility loci that have been reported(31). Candidate genes have emerged that may play a role in CaP progression. These include RNASEL which encodes a ribonuclease that can induce apoptosis upon viral infection(32) and MSR1 which is a lipopolysaccharide receptor found at sites of inflammation in the prostate(33). While the prevalence of familial CaP is low, most tumours arise from somatic mutations that accumulate over several decades. An alteration that occurs frequently in the CaP genome is a gain or loss of a region of DNA. Common gains are at 7p, 7q, 8q, Xq and common losses are at 8p, 10q, 13q and 16q(34-36). The frequency of these alterations varies with reports; however, by identifying functions of disrupted genes a molecular pathway may be developed to explain the process of the growth, metastasis and progression to androgen independent prostate cancer (AIPC). However, due the heterogeneity of the tumours this process is complicated and requires careful dissection(37).
Deletion, inactivation by mutation or suppression of genes by promoter methylation that function in regulating apoptosis and cell proliferation may act to permit the progression of CaP. These tumour suppressor genes have been shown to be important in several aspects of CaP cell growth. A gene lost early at the stage of PIN due to promoter methylation is the glutathione S-transferase pi-1 (GSTP1) gene which helps to neutralize oxidative stress in basal and epithelial cells(38). Additional loss of function of the retinoblastoma (RB) and TP53 gene that regulate cell division have been shown in CaP cell lines and metastatic disease(39). A prominent tumour suppressor gene lost in many CaP tumours is the phosphatase and tensin homologue deleted from chromosome 10 (PTEN). This gene encodes a phosphatase that inactivates the PI3 kinase and Akt pathways that result in regulating apoptotic activity(40).

In addition to inhibition of tumour suppressor genes, activation of proto-oncogenes in CaP are observed. Gain of function mutations are evident in these genes that result in their over-expression or constitutive activation resulting in either increased proliferation or decreased cell death. Growth factors and their receptors that have been shown to be overexpressed include epidermal growth factor, transforming growth factor-α and β, platelet-derived growth factor and insulin-like growth factor-1 and 2(41-44). The Her-2/neu gene encodes a transmembrane tyrosine kinase where overexpression in breast cancer has shown prognostic and therapeutic significance. Her-2/neu amplification is uncommon in CaP, however overexpression has been observed in metastatic CaP and AIPC(45,46). Another potent oncogene is c-myc amplification which is associated with AIPC(47). The
cyclin-dependent kinases, a family of protein kinases that regulate the cell cycle have also been shown to have altered expression in AIPC lesions(48). Mutations in genes involved in cell survival pathways such as telomerase, BCL-2 and the ras family of oncogenes have also been implicated in CaP progression(47,49,50).

The androgen receptor (AR) plays a pivotal role in the development of the prostate and the progression of CaP(51). Androgen resistant tumours are thought to arise through selection pressure for cells containing mutations in the AR gene allowing it to grow independently of testosterone. One mechanism is the increased expression of the AR in AIPC that would permit the growth of the tumour in a low testosterone environment(52). Mutations in the ligand binding domain of the AR result in gain of function of the receptor by allowing it to be able to bind and be activated by a variety of ligands(53).

1.1.4 Current Methods for Diagnosis and Prognosis

The diagnosis of CaP is usually based on symptoms, an abnormal digital rectal exam (DRE) or PSA test. Confirmation of diagnosis is then performed by a trans rectal ultrasound (TRUS)-guided biopsy, and histological examination of the prostate tissue is performed to assess the presence of cancer. It is common practice in North America that men over the age of 50 have a routine DRE and PSA test performed. Men that are at high risk may have the exams performed earlier. This form of screening for CaP has resulted in the stage migration to organ confined tumours, however, the benefit in terms of increasing overall survival is hotly debated.
A CaP patient may present with symptoms stemming from urinary obstruction. These can be hesitancy, nocturia, incomplete emptying, hematuria, diminished urinary stream or impotence. These symptoms are also commonly associated with BPH and warrant further examination by DRE and PSA test. Currently, patients rarely present with symptoms of metastatic disease such as anemia, bone pain or neurologic related symptoms associated with cord compression or metastases to the brain.

A DRE is usually performed to assess the physical features of the prostate and to determine areas of the prostate that have hardened and to determine if there is extension to the area surrounding the prostate. This exam is subjective and the results do not correlate well with tumour volume and extent of disease progression. However, it is useful in determining if the tumour has spread to outside the capsule of the prostate or whether it is confined.

Accompanying a DRE is usually the PSA test which is a serum blood test for the presence of elevated levels of PSA which is a highly prostate specific protein that is produced in large amounts by prostate epithelial cells. It has been shown that normal and BPH tissue produce more PSA than CaP tissue with poorly differentiated CaP producing less than well-differentiated CaP (54). Elevated serum levels of PSA due to CaP growth result from disruption of the basement membrane surrounding the epithelial cells and leakage of into capillaries and lymphatics. The normal range of PSA is less than 4 ng/mL, however, it has been realised that PSA can not be used as a dichotomous marker and levels are considered as a continuum. As a result,
there is considerable overlap between levels of PSA in benign conditions and CaP, thus reducing the specificity of PSA.

Based on the results of the PSA and DRE a TRUS is performed of the prostate to determine staging, prostate volume and to determine sites for needle biopsy. Cancerous tissue is usually hypoechoic compared to the surrounding prostate tissue and thus can be visualized with TRUS, however, this method can not predict CaP with certainty(55). Based on the TRUS results, a needle biopsy is performed transrectally using a TRUS-guided spring-loaded needle gun. Areas that are felt or visualized to be abnormal are sampled with at least 10 to 12 cores. Each core is then assessed by a pathologist to determine the spread of the tumour within the prostate. Biopsies are usually taken from the peripheral or transition zone of the prostate.

To assess the presence and differentiation of CaP the Gleason grading system is used(56). Gleason grading is based on the pattern of the glands and combines in the first and second most prominent patterns evident. Grading is on a scale of 1 to 5 increasing from well differentiated tissue architecture to poorly differentiated structures. The grading from each pattern is then summed to attain the Gleason score from 2 to 10. Based on the Gleason score further imaging of the surrounding tissues may be warranted to determine staging and spread of the tumour outside the prostate. Imaging modalities that are currently used include: ultrasound, computed tomography, magnetic resonance imaging in combination with contrast agents are used to determine spread to the lymph nodes and bone. Based on these results the tumour may be staged based on the Tumour Node Metastasis
(TNM) staging system. From this a prognosis may be determined and appropriate treatment decisions can be made.

1.1.5 Current Treatments

Clinical treatment decisions for CaP, especially in CaP that is localized and detected early, are problematic when determining whether to treat or not to treat the patient. Several factors contribute to determining this decision namely the risk benefit ratio of the adverse events of treatment versus the life expectancy of the patient. Due to the slow growing nature of CaP it is important to determine the prognosis of the patient. Recommendations for treatment based on risk stratification has recently been suggested by the National Institute for Health and Clinical Excellence(57). Ultimately it is up to the healthcare professional to ensure the patient receives adequate information so as to make an informed decision.

Initially, a conservative management approach is taken that is referred to as “watchful waiting”. Here patients enter a clinical follow-up program in which the tumour size and progression is measured by serum PSA testing. If there is evidence of CaP progression then curative or palliative treatments may be considered. For a substantial number of CaP patients, especially older patients with localised and low grade tumours, a watchful waiting approach might be the better option. Data from a recent study from the Surveillance, Epidemiology, and End Results Program showed that patients who undertook conservative management of clinically localized CaP had a 10-year survival of 94% for Gleason 2 to 4 tumours and 45% for Gleason 8 to 10(58).
Radical Prostatectomy

Due to the use of PSA testing there has been an increase in the number of organ confined tumours, the majority of which are deemed to be clinically relevant (59). Thus, these tumours are at the stage where they may be removed to cure a patient’s CaP. Advances in surgical techniques over the decades have revolutionized the ability to cure locally confined CaP. Radical prostatectomy is a surgical procedure where all or part of the prostate gland is removed either as a result from a progressing tumour or if there is urinary obstruction due to enlargement of the prostate from BPH or CaP that restricts the flow of urine in the urethra. The ideal candidate for radical prostatectomy would have a life expectancy of 10 to 20 years post surgery, should be free of co-morbidities and have a clinically relevant tumour that is confined to the prostate and thus curable by surgery. Current surgical techniques allow for nerve sparring where erectile function is preserved and incontinence can be prevented. However, these are still possible long term side effects from the procedure. Radical prostatectomy is not recommended for patients with metastatic disease.

Radiation Therapy

Similar to surgery, the objective of radiation therapy is to cure patients with localized CaP by eliminating the tumour from the prostate. In this case, external-beam radiation or implanted radioactive ‘seeds’ termed brachytherapy are delivered to the prostate with the least possible harm to surrounding tissues. The ionizing radiation given during radiation therapy damages the DNA in cells. Cancer cells
have a reduced ability to repair this damage and thus are unable to proliferate, while
normal cells that receive this radiation are able to repair any DNA damage.

Since the advent of external-beam radiation for treatment of localized CaP
there have been several advances that have allowed improvements in treatment
planning and delivery. As a result, the curative ability of external-beam radiation for
localized CaP has improved over the past decades(60). The type of radiation used is
usually X-rays or gamma rays and is given daily over the course of 6 to 10 weeks.
The implementation of three-dimensional conformal radiation therapy has been able
to incorporate more precise definition of tumour volume and thus calculation of dose
delivered. The ability to exclude normal tissues from receiving high-dose radiation
allows for higher tumour doses without increased surrounding tissue toxicity(61). To
obtain three-dimensional images, computed tomography images are taken to
produce high-resolution three-dimensional images of the prostate and surrounding
tissues. From these decisions on delivery and dose of radiation are determined(62).
A further advancement has been the implementation of intensity-modulated radiation
therapy which allows dose distributions that conform to the tumour area with
enhanced precision(63).

The use of brachytherapy relies on the principle that high doses of radiation
can be given directly within tumours without the radiation leaking to surrounding
tissues. Currently, the mode of delivery is through TRUS-guided implantation of the
radioactive isotopes $^{125}$I or $^{103}$Pd with or without external-beam radiation(64). Use of
permanent metal seed implants in combination with external-beam radiation therapy
has shown improvements in survival(65). Some side effects of both external-beam
and brachytherapy include incontinence, impotence, infertility, bowel problems and fatigue.

*Androgen Ablation Therapy*

Steroid hormones such as testosterone and other androgens have been shown to play an essential role in the growth and development of the prostate. Circulating testosterone is converted to its more potent form in the prostate by 5α-reductase to dihydrotestosterone. This molecule then enters prostate epithelial cells and binds the androgen-binding region of the AR which then binds to androgen response elements on DNA promoting transcription of androgen-responsive genes and prostate cell growth. The major source of testosterone production is from the testes, with a small proportion coming from the adrenal glands. Testosterone secretion is regulated by secretion of luteinizing hormone (LH) and follicle-stimulating hormone (FSH) from the anterior pituitary which stimulates the Leydig cells in the testes and causes them to produce testosterone. Secretion of LH is regulated by LH releasing hormone (LHRH) also known as gonadotropin-releasing hormone which is secreted from the hypothalamus. The secretion of LH and LHRH is under the control of a negative feedback loop where the concentration of testosterone in the circulation acts to inhibit the secretion of LH and LHRH.

The relationship with prostate and CaP growth and androgens was discovered over sixty years ago(66) and was marked with the awarding of the 1966 Nobel Prize. Since then we know that CaP cells deprived of androgens slow in proliferation and under apoptosis while normal prostate tissue undergoes atrophy and a reduction in prostate gland size. Thus, methods to block the action of
androgens on the prostate were developed and have since been used in the
treatment of metastatic CaP.

The gold standard for androgen ablation therapy and treatment for metastatic
CaP has been bilateral orchiectomy (removal of the testes). By this method serum
levels of testosterone are reduced by 95% and due to the reduction in testosterone
there is permanent rise in LH and FSH. Administration of estrogens also reduces
testosterone levels by down-regulating the secretion of LH and FSH by the pituitary.
Diethylstilbestrol is used during this therapy with castrate levels of testosterone
being achieved between 3 to 9 weeks(67). Side effects from estrogen therapy
include nausea, vomiting, gynecomastia, fluid retention as well as cardiovascular
effects. Anti-androgen therapy has also been used to reduce the effects of
androgens to treat CaP. There are two classes of anti-androgens: steroidal and
non-steroidal. Steroidal anti-androgens include cyproterone acetate and megestrol
acetate which act by inhibiting C21-9 decarboxylase – an enzyme that synthesises
androgens, and by inhibiting LHRH release. Non-steroidal anti-androgens act by
competitively blocking the binding of dihydrotestosterone to the androgen receptor.
Three agents that are commonly used are flutamide, bicalutamide and nilutamide.
The agents will decrease the intracellular levels of testosterone and
dihydrotestosterone but in contrast to steroidal anti-androgens there will be
increased serum levels of testosterone due to increased LHRH. The chemical
androgen ablation therapy of choice has become LHRH agonists. Two agonists that
are used are leuprolide acetate and goserelin acetate which are taken
subcutaneously. They function by negatively inhibiting LHRH release in the pituitary
through a negative feedback loop and by downregulating the receptors for LH in the pituitary. An initial rise in LH and FSH is observed initially and this is compensated for by the administration of anti-androgens. A meta-analysis of monotherapies used in androgen ablation showed that LHRH agonists were equivalent to patients treated with orchiectomy(68).

Approximately 70% of CaP patients respond well initially to hormonal treatment(69). However, most tumours eventually acquire androgen independence and can grow without the presence of androgens. The prognosis of these patients is poor with median survival time of less than 10 months(69). New chemotherapeutic approaches have been developed that combat CaP at this stage.

*Cytotoxic Chemotherapy for Androgen Independent CaP*

Owing to the slow growing nature of early CaP, chemotherapeutic agents are not effective in targeting these cells preferentially. As a result, uses of chemotherapies are restricted to treatment of patients with AIPC. Current chemotherapeutic agents used for treatment of AIPC have not been able to show a survival benefit, however, they have been found useful for palliative care and to improve overall quality of life. Current chemotherapies for AIPC include: mitoxantrone and prednisone for the palliation of pain of bone metastases(70). Mitoxantrone acts by inhibiting type II topoisomerase thus disrupting DNA synthesis and repair. Doxorubicin and etoposide act in similar ways by intercalating with DNA and disrupting topoisomerase activity. Estramustine, which is a derivative of estrogen and acts as an alkylating agent that binds and disrupts DNA, has been shown with etoposide to decrease PSA and show a non-statistical survival
benefit(71). Vinblastine is a microtubule inhibitor and functions by binding tubulin and halting cell cycle progression at M phase. Paclitaxel and its semi-synthetic derivative docetaxel are both from the drug class of taxanes and both bind tubulin in microtubules with high affinity thus stabilizing the structure of the cytoskeleton and preventing depolymerisation and as a result, freezing the cell from undergoing cell division. Docetaxel has shown reductions in PSA and non-statistical increases in progression-free survival(72).

1.2 Cancer Biomarkers

1.2.1 Introduction to Cancer Biomarkers

Cancer biomarkers, or tumour markers, can be defined as measurable analytes in biological tissues or fluids that act as a surrogate indication for the presence of a tumour. They may be produced by the tumour itself or by the host in response to the presence of the tumour. The analyte may be measured qualitatively or quantitatively by several methods such as proteomic, genomic or chemical to determine its presence. The ideal tumour marker should be both specific and sensitive for the detection of a certain type of tumour early in its progression. Currently, most tumour markers are not sensitive or specific enough to be used in routine population screening and are used in monitoring patients after therapy.

Tumour biomarkers have been shown to span many different forms. These include small molecule, gene or protein based analytes. Small molecule metabolites have been identified resulting from the metabolic pathways of cancer cells. In addition, some cancers have shown higher concentrations of certain micronutrients.
Nucleic acid based analysis of cancers has resulted in a plethora of differing forms of cancer biomarkers. These include DNA based mutations in cancer cells, single nucleotide polymorphisms, altered expression of certain genes measurable by their transcript levels and epigenetic modifications that alter gene expression. Changes in peptide and protein levels have also been assayed as cancer biomarkers. These include altered expression of peptides and proteins, post-translational modifications, variants of protein isoforms, mutations in proteins, as well as autoantibodies. Both genomics and proteomics have been utilized in multi-parametric approaches such as gene and protein microarrays and mass spectrometric proteomic patterns. In addition, computational approaches to the formation of multi-panel biomarkers through the use of artificial neural networks have resulted in increased performance of several biomarkers. With evolving technologies such as high throughput genomics and proteomics, the search for novel biomarkers that are of high quality is ongoing.

1.2.2 Methods for Identifying Novel Cancer Biomarkers

The opportunity to identify candidate cancer biomarkers has materialized in the past decade with the completion of the human genome sequence and the advent of high throughput sequencing technologies, microarrays and mass spectrometric approaches for the identification of proteins. As a result, there have been large efforts in developing methods to identify novel tumour markers that show clinical utility for diagnosis, prognosis or therapeutic monitoring. Here we discuss two main approaches, genomics and proteomics that have been applied to the search for cancer biomarkers.


1.2.2.1 Genomic Approaches

Of recent there has been an emergence of genetic tools that have made available the comparison of genetic profiles of normal and cancerous tissue and cell cultures. The gold standard to assess gene transcript levels has been the use of real-time polymerase chain reaction (RT-PCR) which is able to quantitatively determine levels of mRNA transcripts in several samples simultaneously. However, the most common high throughput tool in use is the DNA microarray which can assess either DNA gene copy number alterations or transcript levels of a gene’s mRNA on a genome wide scale. Microarrays have thus provided several insights into the molecular mechanisms of cancer biology, as well as led to the discovery of new markers for disease and therapeutic response. The application of microarrays to CaP was thought to help elucidate the genetic underpinnings of the differences between indolent tumours and tumours with a clinically aggressive phenotype. Due to the several challenges posed by the nature of CaP this has resulted in limited success.

There are unfortunately limited models for CaP such as cell lines which are mainly attained from metastatic sites and patients with AIPC(73). Several animal models established have been studied but their relevance to human CaP progression is controversial, albeit recent transgenic models have shown promise for study(74). When using fresh frozen tissue samples for genomic analysis the tissue samples are invariably contaminated with normal glandular architecture and stromal cells, thus confounding results obtained due to high background from normal cells. Issues also occur when selecting normal tissues for control since a field effect may
be present that affects normal epithelium or the presence of PIN lesions(75). Therefore, tissue samples should be carefully examined histologically before being used for RNA extraction in a genomic study. Improvements in tissue dissection including laser-capture microdissection (LCM) have shown to improve the purity of tissue samples obtained(76). However, this procedure has been subject to contamination with artefacts.

In the quest for the identification of differentially expressed biomarkers, several studies have compared gene expression of cancerous to normal tissues or BPH tissue. A number of novel genes have been discovered that have shown differential expression. One gene, AMACR, was found to be expressed in almost all CaP and confirmed by immunohistochemistry and has since become a routine diagnostic tool for pathologists(77). Due to the increased number of datasets emerging from microarray studies of CaP, discrepancies between datasets were seen. This may in part have been due to selection bias of samples used for each dataset, or differences in sample preparation or the genes represented on microarrays used. To overcome these differences a meta-analysis of independent data sets was conducted to validate and establish a model for CaP gene expression(78). From this 50 genes were identified to be overexpressed and 103 genes were underexpressed in CaP with 11 novel genes being found to score high that were not noted in their original datasets. Value has also been gained in the reanalysis of a public data set of Lapointe et al.(79). Tomlins et al.(80) used cancer profile outlier analysis to identify translocations of gene fusions. This involved searching multiple microarray data sets for gene expression differences that were
consistently different between normal and CaP tissues. Through this they discovered
the novel gene fusion of the TMPRSS2 serine protease gene to two ETS family
transcription factor genes which were found to be overexpressed in CaP; ERG and
ETV1, which is under current investigation as being a prognostic marker for CaP.

Correlating clinical outcome with transcript levels has also been an
approached used for stratification of tumour subtypes. An example is hepsin which is
seen to be elevated in CaP relative to normal, but tumours with relatively lower
levels of hepsin have a worse prognosis(81). A comparison of normal, localized and
metastatic tumours resulted in identification of EZH2 which was correlated by
immunohistochemistry with recurrent CaP after surgery(82). In contrast to identifying
single genes that discriminate CaP from normal, another approach has been to use
gene expression patterns to identify subtypes of tumours and predict prognosis. In
this instance unsupervised hierarchical clustering is used to separate tumours based
on clinical outcome(79). From the data used by Lapointe et al. they were able to
distinguish 4 genes positively and 19 genes negatively associated with recurrence of
CaP tumours. Of these MUC1 And zinc-α-2-glycoprotein were also confirmed by
immunostaining to have poor prognosis and tumour recurrence. The comparison of
gene expression data from cell lines and animal models with CaP tissue samples
has also proven useful. A tumour stem cell gene signature was identified by
Glinskey et al. that predicted failure of therapy(83).

In addition to gene expression changes, alterations in gene copy number due
to chromosomal aberrations are associated with malignant transformation(4). Several platforms are available that are used to assess DNA copy number
alterations by comparative genomic hybridization (CGH) including BAC arrays, SNP chips, and cDNA microarrays. Numerous studies have been performed to analyze copy number alterations including those done using cell lines\(^{84,85}\) and with tumour samples\(^{86,87}\). The significance of these changes will facilitate the identification of genes correlating with clinical outcome and identify the critical regions of genetic alteration in CaP progression.

1.2.2.2 Proteomic Approaches

High-throughput technologies that are used to proteomically profile biological specimens have evolved tremendously over the last decade. A large focus has been on the identification of novel biomarkers for disease. The rationale used for a proteomic approach versus a genomic approach to biomarker discovery is that proteins are the effector molecules of the cell, and differences caused during the progression to a malignant phenotype will be evident in differences in the proteome. These proteomic alterations are then candidates for identification as surrogate biomarkers either in tissues or biological fluids.

Proteomic studies first involved the use of two dimensional gel electrophoresis (2-DE), however, novel technologies that have emerged include protein microarrays, ‘soft’ ionization techniques such as electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI), improvements in mass spectrometer design such as the LTQ-Orbitrap (ThermoFisher Scientific), and quantitative isotopic labelling approaches such as iTRAQ\(^{88}\) and stable isotopic
labelling with amino acids in cell culture (SILAC)(89). All of these have combined to contribute to the search for protein biomarkers.

Several approaches can be taken to biomarker discovery utilizing proteomics. The most common, as in genomic studies, is to utilize cancer and normal tissue or biological fluids from patients and perform a proteomic analysis on each to determine differences in the abundance of proteins or protein modifications. The most commonly used sources for searching for biomarkers for prostate cancer has been serum, prostate tissue and seminal fluid. However, one of the main issues with proteomic analysis of biological fluids is the large dynamic range of proteins present. Usually, the proteins of interest are low abundance tumour proteins and the presence of high abundance serum proteins is seen to suppress their ionization thus precluding them from detection. In an effort to simplify this complex mixture debulking or chromatographic methods are used to remove the interference of high abundance proteins.

A common approach that was initially employed was the use of surface enhanced laser desorption ionization time of flight (SELDI-TOF) mass spectrometry (MS) analysis of serum. It was thought to have low sample preparation, was able to simplify the sample mixture prior to MS analysis and have a high throughput. The technology relies on the use of chips with affinity-based chromatographic surfaces where serum samples would be deposited and then washed off. The bound proteins and peptides to the chip surface would then be subjected to MALDI-TOF analysis. This would generate a collection of unknown peptide mass spectra from each sample of which discriminatory algorithms would be used to determine spectral
peaks that corresponded to normal and cancerous states. This method was used as an approach in the discovery of CaP biomarkers(90-92), and has since been debunked as a method for biomarker discovery due to its poor reproducibility and artifactual results(93).

The proteomic analysis of tissues also presents with similar hurdles as with serum but to a lesser extent. Current approaches have been to use LCM to capture pure populations of cancerous and normal cells and perform a comparative proteomic analysis(94). In addition, the use of archived formalin fixed prostate tissue biopsies have also been used for proteomic analysis(95). A simpler approach, albeit in vitro, has been to use CaP cell lines and study their response to stimuli(96) or their secreted proteins in culture(97,98).

To identify protein biomarkers that may be secreted by the tumour and end up in the circulation, it is relevant to study the proximal fluids that are produced by the prostate gland. Since, proteins present in this fluid would contain secreted proteins from tumour glands as well as normal secretions; it is highly likely that these secreted proteins would be detectable in the circulation through tissue leakage. Two recent studies highlight this approach, each using prostatic fluid(99) or urine(100) containing prostatic fluid from patients with and without CaP. Bioinformatics analysis of the proteins identified resulted in selection of novel candidates for further validation.

It has also been speculated that antibodies are generated to tumour antigens and are then amplified and present in the circulation. A novel approach to identify autoantibodies to tumour antigens utilizes protein microarrays that are spotted with
antibodies from the serum of patients with CaP. This method measures the humoral response to tumour proteins. This approach was used to detect a response against AMACR a well known CaP biomarker(101). Wang et al. demonstrated that they were able to generate phage-display library of proteins derived from CaP tissue. Using this they identified a 22-phage peptide detector for autoantibodies in CaP patients(102). Thus, the clinical significance of autoantibodies in CaP has been shown to be prominent.

1.2.3 Clinical and Analytical Properties of a Cancer Biomarker Assay

Clinical efficacy and analytical robustness are two of the main criteria used in selecting tumour makers for use. A tumour marker should be specific for its application be it screening, prediction or therapeutic response monitoring. Overall it should exhibit high sensitivity and specificity with a high area under curve (AUC) in a receiver operating characteristic curve (ROC). It should also be well defined biochemically with the availability of detection reagents such as antibodies. The analytical method for measuring the marker should be accurate and precise with no cross-reactivity and should be automated for efficient execution.

When developing a test for a tumour marker in a biological fluid for example it is important that the reproducibility of the assay first be optimized. Next the dynamic range of the assay should be established such that analytical sensitivity and limit of detection be determined along with the maximal value measurable that is accurate. After a working assay is developed the clinical validation of the marker may be assessed. It is important to establish reference values of the tumour marker in
healthy individuals to determine upper limits of the marker in the healthy population. This will in turn help in deciding cut-offs for the tumour marker in determining clinical diagnosis.

The clinical sensitivity and specificity of a tumour marker assay can be defined as the percentage of people tested with the disease that the test correctly finds positive and percentage of people with the disease that the test correctly finds negative, respectively. These values are in essence indications of the true positive and false positive rates of the test. Calculations for each are shown below:

Sensitivity = True Positives / (True Positives + False Negatives)

Specificity = True Negatives / (False Positives + True Negatives)

Based on these values a ROC curve may be drawn depicting the range of sensitivity and specificity across a range of cut-off values. The sensitivity values are plotted on the y-axis with 1-specificity on the x-axis. This aids in determining the trade off between sensitivity and specificity and helps in choosing a cut-off value. It is important to note the consequences of a false positive or false negative when choosing the appropriate cut-off. The added value of a ROC curve is that multiple tests can be compared on the same graph to determine which has the highest AUC which translates the ability of the test to discriminate those with and without the disease. These values range from 0.5 (a test that has a 50% probability of being correct) and 1 (a test that has a 100% of being correct).
A measure of depicting how clinically relevant the tumour marker assay is of detecting disease in the patient is shown in its predictive value. The positive predictive value (PPV) of a test is the percentage of people testing positive for the test that indeed have the disease. The negative predictive value (NPV) of a test is the percentage of people testing negative for the test who indeed don’t have the disease. The calculations of which are shown below:

\[ PPV = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \]

\[ NPV = \frac{\text{True Negatives}}{\text{False Negatives} + \text{True Negatives}} \]

Taken together in combination with the prevalence of the disease will determine if the test is suitable for population screening or in just high risk populations.

1.2.4 Applications of Cancer Biomarkers

There are several uses for tumour markers in aiding clinical management of patients with cancer. Based on the properties of the marker and the test used to measure it dictates which application the marker is suitable for. These include population screening, diagnosis, prognosis, prediction of therapeutic response and monitoring effectiveness of therapy, determining tumour recurrence or remission, or being used as an imaging marker in localizing the tumour. Below we discuss the requirements of the marker and assay for each application.
Screening for cancer in the population requires that a marker is elevated when the disease is in its early stages and curable. With the exception of PSA, most markers are elevated in late stage disease and have a low diagnostic sensitivity. It is also beneficial if the marker is specific to a particular tissue since elevations in other tissues due to other diseases or inflammatory conditions will reduce its specificity and increase its false positives. It is also required that the test for the marker be non-invasive and low in cost to increase compliance, e.g. a blood test. Screening must also demonstrate that it delivers a clear benefit to the patient in reducing morbidity and mortality associated with the disease.

A tumour marker used in diagnosis of cancer involves the same considerations as a marker for screening. Most markers used have either low diagnostic sensitivity or specificity. However, for patients in high-risk groups where prevalence is relatively high, using a diagnostic marker will aid in deciding whether more invasive tests are required.

Determining the prognosis of a tumour is essential in deciding the course of therapy to be taken. Most tumour markers are correlated with cancer prognosis and other prognostic indications such as tumour grade and staging. However, the clinical decision to determine a therapeutic course of action can not be taken based on the levels of the tumour marker alone. This is evident with PSA and its derivatives.

Tumour markers may also be used to predict and monitor therapeutic response. With a marker such as PSA, its high tissue specificity allows it to be used to monitor a patient after radical prostatectomy and other therapies. This determines
if the therapy was effective and also determines if there is tumour recurrence as the PSA level will be seen to increase again. Tumour markers have also found use as targets for therapy. The most prominent marker being Her-2/neu in breast cancer as a target for the antibody based therapy Herceptin. Tumours overexpressing Her-2/neu are responsive and tumours lacking will not respond, thus this tailors the type of therapy to the patient.

1.2.5 Current and Emerging Prostate Cancer Biomarkers

The current biomarker used for CaP diagnosis is PSA. It is considered both the best tumour marker available for any cancer and as a marker with many shortcomings. Originally, it was used for monitoring patients with CaP and was subsequently implemented for screening.

Prostate Specific Antigen-Derived Forms:

It has become clear that the operating characteristics of PSA need to be improved. One approach has been to measure PSA derivates that include PSA rate change over time (PSA velocity, PSAV), ratio of PSA to prostate volume (PSA density, PSAD), and age specific PSA ranges. In addition, novel improvements in measuring PSA and PSA related proteins have allowed the measuring of percent free PSA (%fPSA) which is a ratio of free to total PSA (tPSA); complexed PSA which measures how much PSA in serum is bound to either α-2-macroglobulin (A2M), α₁-protease inhibitor (API) or α-1-antichymotrypsin (ACT); as well as different cleavage isoforms of PSA.
**Age-Specific PSA:**

Age-specific PSA ranges adjust PSA thresholds based on age with the assumption that older men have higher normal PSA levels due to prostate enlargement since PSA correlates linearly with prostate size. Thus, a lower cut-off for younger men should be employed to improve sensitivity, and a higher cut-off in older men would reduce false positives, increasing specificity. Recently, a retrospective study showed that men ≤ 50 years with a PSA of 0.51-1ug/L compared to ≤ 0.5ug/L had a 2.5 fold greater risk of CaP up to 25 years later. In addition, men with PSA of 2-3ug/L had a 19 times greater chance of CaP. Thus, this test can serve to stratify patients based on risk and to monitor and tailor therapy. This test still remains controversial as studies have shown reduction in sensitivity and lower diagnostic accuracy compared to the 4.0ug/L cut-off.

**PSA Density:**

PSA density takes into account the volume of the peripheral and transitional zone (TZ) of the prostate gland and is calculated by dividing the serum PSA level by prostate size determined by TRUS. A higher PSAD has been shown in men with CaP versus BPH, while others have not been able to reproduce these results. In BPH the TZ is mostly enlarged and not a frequent site of carcinoma. Taking into account TZ size specifically (PSAD-TZ) has shown improvements in specificity but has not been reproducible. A PSAD ratio of more than 0.15 is used to indicate an increased risk of CaP. This test would increase specificity and sensitivity in large prostates in men with small prostates respectively. While the use of PSAD is not in clinical use for screening due
the cumbersome need for TRUS, utility may be derived in selecting men for repeat biopsy.

**PSA Velocity:**

PSA velocity is calculated by measuring serial PSA measurements over time and determining the rate increase. It has been seen that BPH patients show a linear increase in PSA, while CaP patients display exponential increases in PSA(117). However, variations in PSA levels due to biological or analytical variability can result in imprecision of this test(118,119). In addition, infection and chronic inflammation are confounding factors that can increase PSA levels. A consensus on the appropriate parameters to use has still yet to be determined. Carter *et al.*(120,121) showed that at least three measurements over a period of at least two years are required and a PSAV >0.75ug/L per year was significantly associated with CaP with men with 4-10ug/L showing specificities of >90% PSA. In addition, a 2.0ug/L yearly PSAV has been associated with a shorter time to PSA relapse and mortality following surgery for localized CaP(122-124). Thus, this could identify patients who may not benefit from localized therapy. Initial results from the European Randomized Study of Screening for Prostate Cancer confirmed PSAV values differ significantly between men with and without CaP and was limited in value in predicting biopsy outcome(125).

**Free PSA:**

Measuring the ratio of fPSA to PSA (%fPSA) has shown to be beneficial by improving the operating characteristics of PSA in the ranges of 4-10ug/L and <4ug/L in several retrospective and prospective studies(126-129). Higher %fPSA correlates
with decreased risk of CaP and increased risk of BPH and this ratio is now being used routinely since the mid 1990’s and has been approved by the USA Food and Drug Administration in 1998. Sensitivity for %fPSA testing has been shown at 90-95% resulting in reduction of unnecessary biopsies by 15-20%(126,130). However, confounding variables such as prostate volume(131), stage and grade(132), total PSA(133), PIN lesions(134), race, age and drug treatment history should be considered(135). A recent study utilizing 10-12 core biopsy versus the sextant biopsy method depicted %fPSA as having reduced diagnostic efficacy(136). Conflicting studies(137-139) have resulted in a debate as to what the optimal %fPSA threshold should be.

*Complexed PSA:*

Prostate specific antigen circulating in serum has been shown to be mostly bound to protease inhibitors, with the majority (~75%) consisting of the PSA-ACT complex(140-142). Measuring these forms of complexed PSA (cPSA) have shown to improve specificity marginally in the PSA range of 2-10ug/L in a prospective trial as well as in meta-analyses of CaP patients(143,144). Currently there are two assays available: one is for PSA bound to ACT (PSA-ACT) and one is for cPSA which measures PSA-ACT and PSA bound to API (PSA-API) but not bound to A2M(144,145). This assay is of limited use since PSA-ACT increases(146) where as PSA-API decreases in CaP patients(147). Currently, it is very difficult to measure PSA-A2M since the levels are low and A2M sterically hinders access to PSA epitopes(148). Complexed PSA however, in combination with tPSA (cPSA/tPSA) results in comparable sensitivity and specificity as %fPSA(149-152). In addition,
there have been conflicting multi-center studies depicting either benefits of cPSA compared to tPSA(143,144) or showing no advantage when used independently(147,153). Thus, the context of the usefulness of cPSA is still debated. Measurements of PSA-A2M(154) and PSA-API(155) complexes in serum have been reported(156) and have both shown elevation in BPH patients vs. CaP, while PSA-API has shown to be increased and an improvement over tPSA but not fPSA.

**Cleavage Isoforms of PSA:**

Prostate specific antigen is secreted from prostate epithelial cells as a pro-enzyme with a seven amino acid activation peptide that is cleaved after secretion. However, there are several different isoforms and cleavage products of PSA that result and that have been assayed(157,158). The non-clipped 'intact' PSA (iPSA) has been assayed and shown to be informative in discriminating CaP from BPH(159,160). This free form of PSA is enzymatically inactive and the ratio with fPSA has been shown to be higher in CaP patients. proPSA contains 244 amino acids (-7)(161) compared to fPSA (237 amino acids) and can be differentially cleaved to produce different forms termed (-7), (-5), (-4), (-2), (-1)(162). proPSA in combination with fPSA has been shown to improve detection of CaP in ranges less than 4ug/L(163) and in the 4-10ug/L range and associated with aggressive CaP’s more so than tPSA, fPSA and cPSA alone(157,164,165). (−2)proPSA has shown to be increased in serum of CaP patients compared to BPH and can discriminate CaP from BPH in patients with elevated fPSA(166). The efficacy of the (−5, −7) proPSA forms has been evaluated and has shown inadequate ability compared with %fPSA to enhance detection of CaP(167,168). However, in combination with fPSA it was
correlated with increased Gleason grade and metastatic disease and was able to improve specificity in PSA levels of 2-4ug/L(169).

Benign or BPH associated PSA (BPSA) is a cleaved sub-form of fPSA that has been correlated with presence of benign prostate tissue in the TZ of the prostate(170). Benign PSA is unable to discriminate between BPH and CaP in serum independently(171), however it could serve as a marker for BPH when used in conjunction with %fPSA and therapeutic monitoring of BPH patients(172). The proPSA/BPSA ratio was measured in CaP patients with fPSA less than 15% and attained a sensitivity and specificity of 90% and 46% respectively(173). BPSA is elevated in patients with symptomatic BPH versus asymptomatic patients and correlates with prostate enlargement more closely than tPSA and %PSA(174). In healthy men BPSA is almost undetectable and is not affected by age.

Molecular derivatives of PSA still need to be validated in larger sample sets to determine their clinical utility for diagnosing BPH or CaP. Currently, there is still debate on which combination of derivatives produces the optimal diagnostic and prognostic characteristics. Thus, there is a great need for novel identification of biomarkers to aid in increasing the specificity of CaP detection and prognostication.

**Human Kallikrein-Related Peptidase 2:**

Human kallikrein-related peptidase 2 (KLK2) is a secreted serine protease from the same gene family as PSA and shares an 80% sequence homology. Like PSA it's highest expression is in the prostate, albeit it is expressed in other tissues,(175) it is also androgen regulated in the prostate however its concentrations in the circulation are 100 times less than PSA, albeit its mRNA levels are half that of
PSA. Unlike PSA, KLK2 differs in enzymatic activity and is mostly found in the free unbound form. Prostate cancer tissue data has shown that KLK2 is relatively elevated during CaP progression and thus may be a useful CaP biomarker(176,177). Serum studies have displayed improvements in diagnosis of CaP(178) in combination with tPSA(179) and fPSA(180-185) specifically with respect to extracapsular extension and tumour volume(186,187). KLK2 also showed improved independent prognostic information versus PSA for risk of biochemical recurrence in men with PSA ≤10ug/L(188). Like PSA, KLK2 is also found in a bound and free form(189), where the bound form has been shown to diagnose CaP with higher specificity in complex with protease inhibitor 6(190). Additional validation studies are required to elucidate the full prognostic potential of KLK2.

Other Tissue Kallikreins:

Up until recently PSA, KLK2 and pancreatic/renal kallikrein 1 (KLK1) were the only genes identified in the kallikrein locus on chromosome 19. Now we know that the locus spans 300kb and consists of 15 genes that share significant homology and sequence similarity at the DNA and amino acid sequence(175,191). In addition to KLK2, other kallikreins have shown utility as biomarkers for CaP and other diseases(192-196). There are 8 kallikreins that are expressed relatively highly in prostate tissue – KLK2-4, 10-13, and 15(175). Of these KLK11 shows promise as a serum biomarker for CaP(197). In combination with tPSA and %fPSA showed improvement in prediction of CaP(198,199).
Prostate Specific Membrane Antigen:

Prostate specific membrane antigen (PSMA), formally known as folate hydrolase 1, is a membrane glycoprotein highly expressed in epithelial cells of normal and CaP. It was first discovered from a monoclonal antibody derived from mice immunized with the membrane fraction of the CaP cell line LNCaP(200). It was found that there was increased relative expression of PSMA in epithelial cells of CaP tissue; as well PSMA was shown to have folate hydrolase and carboxypeptidase activity. The function of PSMA in the prostate is currently unknown; however, in the brain it has shown to release glutamate from N-acetylaspartylglutamate. A commercial imaging test for PSMA was developed (Prostascint, Cytogen Corp.) by using the 7E11 antibody to PSMA conjugated to indium-111 and used for radioimmunoscinography. This approach was validated against computed tomography, magnetic resonance imaging, and biopsy results and was shown to be more sensitive and specific in detecting metastatic spread to the lymph nodes(201). Currently, there has been conflicting data showing reproducible use of PSMA as a serum marker for CaP(202-205). Detection of PSMA versus PSA via RT-PCR for the identification of circulating tumour cells did not show a strong correlation with CaP(206). Finally, PSMA has been studied as a target for therapy by using antibodies conjugated to radioisotopes or toxins, or the activation of dendritic cells against PSMA(207-211). The use of PSMA has not been adopted into clinical practice as yet and its role as a diagnostic and therapeutic tool is still evolving.
**Prostate Cancer Antigen 3:**

Also known as DD3, prostate cancer antigen 3 (PCA3) is a non-coding RNA that is expressed almost exclusively in the prostate and shown to be highly overexpressed in CaP tissue, including metastasis, versus BPH. Several assays can measure PCA3 mRNA in urine sediment. The only commercially available test is APTIMA® (Gen-Probe Inc., San Diego, CA) which utilizes transcription-mediated amplification. A PCA3 score is derived from PCA3 mRNA levels normalized to PSA. A recent large multi-institutional study of patients undergoing biopsy included PCA3 analysis in voided urine after prostatic massage; an AUC of 0.66 versus 0.57 for PSA using a PCA3 score cut-off of 58 was obtained. In a study of 233 patients undergoing repeat biopsy after a negative biopsy, PCA3 had an AUC of 0.68 and a sensitivity of 58% and specificity of 72%. This test could potentially be useful in improving the specificity of PSA. A combination of PCA3 and another three urinary biomarkers, GOLPH2, SPINK1 and TMPRSS2:ERG fusion, improved the sensitivity and specificity of PCA3 alone.

**Early Prostate Cancer Antigens:**

Changes in nuclear matrix proteins have been shown to be associated with carcinogenesis. Early prostate cancer antigen (EPCA) is a nuclear matrix protein that was initially detected by proteomic profiling of rat prostate tissue. It has since then been shown promise as a diagnostic marker for CaP. Immunohistochemical studies by auto-antibodies to EPCA in biopsies of CaP tissue showed increased staining relative to non-cancerous specimens. A field effect was also seen in non-cancerous adjacent areas to tumour tissue, in 86% of CaP...
tissue and aids in identifying patients at risk who have a negative biopsy. A blood-based assay for EPCA was developed showing a sensitivity of 92% with 94% specificity in a small cohort of 12 CaP and 34 healthy patients(219). Another study measuring the EPCA-2 protein in serum showed a 92% specificity and 94% sensitivity for identifying CaP and was able to differentiate localized from metastatic CaP with an AUC of 0.89(220). Larger and other independent studies are awaited to confirm this promising data. However, methodological deficiencies were identified with such markers, casting doubt on their actual validity(221).

**Alpha-Methylacyl-CoA Racemase:**

Alpha-methylacyl CoA racemase (AMACR) is an enzyme involved in the oxidative metabolism and synthesis of branched chain fatty acids found in dairy products and red meat(222). As well as being strongly expressed in CaP tissue, it is located in a gene region (5p13.3) that contains polymorphisms that are associated with CaP(223). A meta-analysis of microarray data showed AMACR to be up-regulated in CaP with high confidence(78). A multi-institutional study of immunohistochemical staining of AMACR helped in distinguishing benign from cancerous prostate tissue with a 97% sensitivity and 92% specificity(224). Recently, decreased AMACR expression was also shown to have prognostic value in predicting biochemical recurrence and death due to CaP(225). Circulating levels of AMACR mRNA have been measured by RT-PCR in serum and urine(226). Protein levels are low in serum, however, AMACR protein has been detected by Western blot in urine(227). Increased levels of autoantibodies to AMACR were able to discriminate CaP patients from healthy subjects, in the PSA range 4-10ug/L, and
showed a sensitivity of 62% and specificity of 72% (101). Additional studies are underway to fully elucidate the potential of AMACR as a biomarker for CaP.

_Urokinase Plasminogen Activator and Receptor:_

The degradation of the extracellular matrix has been associated with cancer progression and the urokinase-plasminogen activation (uPA) cascade has been shown to participate in this. Plasminogen is converted to its active form plasmin through the activation of the serine protease uPA, binding to its receptor uPAR. One study demonstrated increased levels in BPH and CaP compared to normal, albeit there was no statistically significant association with CaP (228). Detection of isoforms of uPAR, in combination with PSA isoforms and KLK2 showed improved prediction of biopsy outcome in patients with elevated PSA in both univariate and multivariate models (229). Increased tissue levels in CaP were associated with osteoblastic metastases (230), as well as advanced CaP progression (231). Recently a study has shown increased serum levels of uPA and uPAR in patients with metastatic bone CaP (232). These studies depicted preoperative plasma uPA as a predictor of biochemical recurrence and metastatic disease indicating the presence of distant disease at time of localized therapy. Large prospective studies are needed to elucidate the full prognostic potential of uPA and uPAR for preoperative models of disease progression and metastases.

_Insulin-Like Growth Factors and Binding Proteins:_

Serum levels of insulin-like growth factors (IGF) and their binding proteins (IGFBP) have been found to be associated with CaP. The IGF family consists of two ligands (IGF-1, IGF-2), two receptors (IGFR-1, IGFR-2) and six binding proteins
(IGFBP-1 to 6). Increased levels of IGF-1 and decreased levels of IGFBP-3 have been correlated with increased risk of developing CaP(233). Another prospective study found that IGF-1 levels were increased slightly with CaP risk but did not outperform PSA(234). However, others have failed to reproduce these results and found no association with CaP progression. In addition, the main IGFBP produced by the prostate is IGFBP-2 and was shown to be increased in CaP. Albeit, levels in localized tumours were correlated inversely to tumour size and indicators of CaP progression(235). Serum levels of IGFBP-3 were shown to be inversely correlated with presence of metastases to the bone, however did not show any difference to localized CaP and normals(235).

**TMPRSS2:ERG/ETV Gene Fusion:**

Gene rearrangements have been implicated in cancers, in particular haematological malignancies. One such rearrangement between the transcription factor genes ERG (21q22.2) and ETV1 (7p21.1) with membrane-anchored serine protease TMPRSS2 (21q22.3) was shown to occur in 80% of CaP by cancer outlier profile analysis(80). The fusion product of these genes tested in urine has been shown in 42% of CaP patients and in 20% of patients with PIN and rarely in BPH(236). A prospective study which followed 252 men with stage T1a/b CaP for 9 years showed that the TMPRSS2:ERG fusion was associated more than the TMPRSS2:ETV fusion in Gleason sums>7, metastatic disease and death due to CaP(237). An isoform of the TMPRSS2:ERG fusion has been shown by fluorescence in situ hybridization (FISH) analysis to be present in 80-95% of CaP tissue, and could be a potential target for therapy.
Transforming Growth Factor Beta:

Transforming growth factor \( \beta_1 \) (TGF-\( \beta_1 \)) is a widely acting growth factor involved in a variety of molecular processes such as cellular differentiation, immune response, angiogenesis and proliferation. Studies with model systems of CaP have shown a role of TGF-\( \beta_1 \) in CaP progression(238). Increased levels of TGF-\( \beta_1 \) in CaP tissue have been correlated with tumour grade and stage and lymph node metastasis(239). An ELISA for TGF-\( \beta_1 \) was used to measure pre-operative plasma levels, which have shown to be increased in CaP patients(240) and correlate with extracapsular extension, seminal vesicle invasion, metastasis(241) and biochemical recurrence(242). Thus, TGF-\( \beta_1 \) could prove useful as a prognostic marker for CaP.

Enhancer of Zeste Homolog Gene 2:

Enhancer of zeste homolog gene 2 (EZH2) is part of the polycomb family of proteins involved with regulation of gene expression. It has been shown that EZH2 is expressed more in metastatic CaP compared to localized CaP and BPH through gene expression profiling of CaP tissue from autopsies of men who died from metastatic CaP(82). In addition, it was found to outperform preoperative PSA and Gleason score for determination of CaP progression. It was also shown in combination with E-cadherin to predict CaP recurrence after localized therapy(243). Development of a serum assay would aid in the validation of this candidate biomarker to identify patients at risk of developing metastatic disease.

Glutathione S-Transferase Pi hypermethylation:

Hypermethylation of tumour suppressor genes at their promoter regions at cytosine/guanine (CpG) nucleotide islands have been implicated in CaP. Glutathione
S-transferase pi (GSTP1) is an enzyme that protects DNA from free radical damage. Reduced expression of the GSTP1 gene due to hypermethylation of the promoter has been consistently shown in CaP and has been studied in urine sediment to determine the need for biopsy(244). This assay has been improved through prostatic massage before collecting urine(245). Panels of genes, including GSTP1, have been studied in a similar manner(246).

_Prostate Secretory Protein 94 and Binding Protein:_

Prostate secretory protein 94 (PSP94), also known as h-microseminoprotein, is a highly abundant protein in semen and plays a role in regulation of cell proliferation and apoptosis. Bound forms of PSP 94 exist that are complexed with PSP94-binding protein (PSPBP). Serum levels of PSP94 / free PSP94 and PSPBP in CaP patients after local surgery were associated with Gleason sum, biochemical recurrence and surgical margin status(247). Large prospective studies are still required to validate this candidate biomarker.

_Markers for Neuroendocrine Differentiation:_

Chromogranin A is a peptide produced by the neuroendocrine cells in the prostate and is currently used for diagnosis and prognosis of CaP tumours that show neuroendocrine differentiation. Increased chromogranin A levels in serum have been correlated with androgen independent CaP progression and poor prognosis(23), and have been shown to precede PSA elevation(248), and improve specificity in combination with fPSA(249).

Progastrin-releasing peptide (proGRP) is a growth factor found to be released in the neuroendocrine type of CaP. Increased levels were detected in metastatic
CaP and was associated with its progression (24). It was also shown to be predictive of the androgen-independent phenotype (250). Thus, both chromogranin A and proGRP may be used to monitor patients with late stage disease for hormone-refractory CaP in a subset of patients that display neuroendocrine differentiation.

**E-cadherin:**

Cell-cell adhesion plays an important role in normal tissue architecture and carcinogenesis. E-cadherin is a cell adhesion molecule expressed in epithelial cells and whose expression has been shown to predict CaP prognosis. An immunohistochemical study showed reduced expression of E-cadherin in 50% of CaP tumours, where normal prostate tissue showed uniform expression (251). This was then further studied and correlated with grade, tumour stage, and survival. A lower immunohistochemical expression of E-cadherin was associated with CaP patients with a shorter survival (252).

**Annexin A3:**

Annexin A3 (ANXA3) is a calcium-binding protein that is part of the annexin family of proteins. ANXA3 has been shown to be involved with the activation of the immune response as well as membrane trafficking and lymphocyte migration (253, 254). Recently, ANXA3 has been studied as a promising prognostic tissue marker for CaP and was shown to have lower expression in CaP versus BPH and PIN and normal tissues by immunohistochemistry (255, 256). ANXA3 was able to stratify a large group of intermediate-risk patients in to high and low risk subgroups.
Prostate Stem Cell Antigen:

Prostate stem cell antigen (PSCA) is a membrane glycoprotein that shows fairly specific expression to the prostate. PSCA was detected by immunohistochemistry in CaP tissues and its RNA was found in blood samples(257). Increased expression of PSCA was correlated with increased risk of CaP, Gleason score, stage and presence of metastasis(258). PSCA was also investigated as a target for therapy(259). However, larger validation studies are required to confirm its clinical utility.

Hepsin:

Hepsin is a membrane serine protease that is highly expressed in prostate tissue and was first identified in the human liver from cDNA libraries(260). Overexpression of the hepsin gene has been shown in 90% of CaP tumours through mRNA expression profiling studies(261). Immunohistochemical staining of hepsin in one study showed it to be highly expressed in PIN lesions of the prostate, and preferentially expressed in CaP versus BPH(262). Further studies in serum and urine are required to fully elucidate its diagnostic potential.

Interleukin-6 Ligand and Receptor:

Interleukin 6 (IL-6) is a cytokine secreted by a variety of cell types and involved in the immune and acute-phase responses. Elevated levels of IL-6 and its receptor IL-6R have been shown in metastatic and AIPC(263), and has thus far have been suggested as a candidate markers for CaP progression(264). Studies with IL-6 in combination with TGF-β1 have also shown promise in CaP diagnosis(265).
Circulating Tumour-associated DNA:

Dissemination of tumour cells is a prerequisite to metastasis; thus early detection of these cells in the circulation can be useful in prognosis of CaP patients. This has been shown by RT-PCR which has proven to be sensitive and used to increase the accuracy of staging and prediction of disease recurrence by using markers specific to the prostate(257,266).

Autoantibodies:

It is known that the immune system elicits an autoantibody response to some antigens overexpressed by tumours. Humoral responses to huntingtin-interaction protein 1, prostasomes and AMACR have been reported(101,267,268). By using phage display and protein microarrays, Wang et al.(102) were able to identify autoantibodies to peptides derived from CaP tissue in a new approach termed “cancer immunomics”. They were able to generate a 22 phage peptide array that was able to discriminate 68 CaP serum samples from 60 controls with 88.2% specificity and 81.6% sensitivity with an AUC of 0.93, which was superior to PSA (AUC 0.8). Studies are underway to further validate this detection tool in a larger cohort. A recent study by the same group applied a similar approach, followed by biological network analysis to determine deregulated pathways in CaP progression(269). One concern is that the needle biopsies themselves may be eliciting an autoimmune response.

Nomograms:

Nomograms are multi-parametric tools that combine clinical features such as tumour grade/stage and biomarkers to provide physicians with standardized patient
care. They utilize evidence-based approaches for decisions regarding treatment at each stage of disease management. The value of a nomogram is derived from its performance characteristics and user-friendliness. There are numerous nomograms that have been developed for CaP including a TGF-β₁ and IL-6 standard nomogram for biochemical recurrence(270), as well as nomograms predicting outcome of biopsy(271). A review by Karakiewicz and Hutterer summarizes nomograms that have been developed for CaP(272).

**Multi-Parametric Tests / Artificial Neural Networks:**

The use of combined biomarkers for improvement of disease prediction has been used widely for many different disease states. There is heterogeneity among individuals and as a result, disease states within these individuals differ in their biology. Thus, the use of multi-parametric tests will most likely be more applicable for population screening versus one marker. Recently, Parekh *et al.* (273) used a biomarker panel consisting of 54 proteins that included adipokines, metalloproteinases, adhesion molecules and growth factors. They used age-matched controls and measured pre-diagnostic serum concentrations of patients later diagnosed with CaP. Their results did not prove that the marker panel was able to outperform the risk factors from the The Prostate Cancer Prevention Trial (PCPT) calculator. Artificial neural networks (ANN) have been used to model complex relationships between variables and find patterns in data. Stephan *et al.* have used the ANN approach to assess various combinations of kallikrein biomarkers to determine their clinical utility for CaP diagnosis(274).
Proteomic Patterns:

High throughput proteomic analysis of biological fluids, tissues and cell lines has recently become a popular approach for the identification of novel biomarkers. In particular, the application of SELDI-TOF has been used frequently to profile biological samples. With respect to CaP, Adam et al. (91) used a decision tree algorithm to ascertain a peak fingerprint that could discriminate CaP from normal individuals with a sensitivity and specificity of 83% and 97%, respectively. Petricoin et al. (275) used 266 serum samples from CaP patients and controls to achieve 95% sensitivity and 78% specificity. Qu et al. (92) used a boosted decision tree algorithm for analysis of their SELDI-TOF data and were able to achieve 97% sensitivity and specificity. Other studies have also used proteomic profiling for CaP diagnosis showing usefulness of the approach (276). However, the use of proteomic pattern fingerprinting has come under scrutiny and as a result the NCI/EDRN has conducted a multi-institutional study to objectively validate this approach and the results were recently published (93). Even though stage 1 of the validation confirmed the analytical reproducibility of the approach, stage 2 was unable to determine if the approach could predict CaP in a case-control series across institutions. The cause of this failure has been attributed to pre-analytical, analytical and bioinformatics biases, as described in previous literature (277, 278).
Table 1.1: List of candidate biomarkers for prostate cancer and their possible clinical utility.

<table>
<thead>
<tr>
<th>Candidate CaP Biomarker</th>
<th>Assessed Clinical Utility</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>KLK2</td>
<td>Diagnostic and prognostic predictor of extracapsular extension, tumour volume and biochemical recurrence</td>
<td>(186,188)</td>
</tr>
<tr>
<td>KLK11</td>
<td>Early predictor of CaP in serum</td>
<td>(199)</td>
</tr>
<tr>
<td>PSMA</td>
<td>Imaging marker and target for therapy</td>
<td>(201,209,210)</td>
</tr>
<tr>
<td>PSP94</td>
<td>Predictor of Gleason sum, surgical margin status and biochemical recurrence after local surgery</td>
<td>(247)</td>
</tr>
<tr>
<td>PSCA</td>
<td>Immunohistochemical marker associated with Gleason sum and stage. Target for therapy.</td>
<td>(258,259)</td>
</tr>
<tr>
<td>PCA3</td>
<td>Urinary biomarker for detection of CaP</td>
<td>(214,215)</td>
</tr>
<tr>
<td>EPCA/EPCA2</td>
<td>Immunohistochemical detection of CaP, serum marker to differentiate local from metastatic CaP</td>
<td>(218-220)</td>
</tr>
<tr>
<td>Hepsin</td>
<td>Immunohistochemical detection in PIN, CaP vs. BPH</td>
<td>(262)</td>
</tr>
<tr>
<td>AMACR</td>
<td>Increased detection of autoantibodies in CaP, immunohistochemical expression as a prognostic factor for biochemical recurrence and death</td>
<td>(101,224,225)</td>
</tr>
<tr>
<td>EZH2</td>
<td>CaP tissue gene expression predicts progression</td>
<td>(82)</td>
</tr>
<tr>
<td>uPA/uPAR</td>
<td>Increased tissue and serum levels predicts biochemical recurrence and metastasis</td>
<td>(230,232)</td>
</tr>
<tr>
<td>IGF/IGFBP</td>
<td>IGF-1 slightly increased in CaP serum, IGFBP’s inversely correlated to CaP progression</td>
<td>(234,235)</td>
</tr>
<tr>
<td>TGF-β1</td>
<td>Increased immunohistochemical and serum levels with CaP progression and biochemical recurrence</td>
<td>(239-241)</td>
</tr>
<tr>
<td>TMPRSS2:ERG/ETV</td>
<td>Increased detection in urine of CaP, PIN patients vs. BPH, gene fusion present in CaP tissue by FISH</td>
<td>(236,237)</td>
</tr>
<tr>
<td>GSTP1</td>
<td>Hypermethylation of promoter detected in urine to assess for biopsy</td>
<td>(244)</td>
</tr>
<tr>
<td>IL-6</td>
<td>Elevated serum levels in late stage CaP</td>
<td>(263,264)</td>
</tr>
<tr>
<td>Chromogranin A</td>
<td>Monitoring of patients with androgen independent late stage CaP with neuroendocrine differentiation</td>
<td>(23,248,249)</td>
</tr>
<tr>
<td>Annexin A3</td>
<td>Decreased expression in tissues of CaP by immunohistochemistry, prognostic risk marker</td>
<td>(255,256)</td>
</tr>
<tr>
<td>Progastrin-releasing peptide</td>
<td>Monitoring of patients with metastatic CaP with neuroendocrine and androgen-independent phenotype</td>
<td>(24,250)</td>
</tr>
<tr>
<td>E-cadherin</td>
<td>Reduced immunohistochemical expression in CaP correlated with stage and reduced survival</td>
<td>(251,252)</td>
</tr>
</tbody>
</table>
1.3 Prostate Cancer Model Systems

The utilization of mouse and cell culture based model systems for CaP has allowed the portrayal of this disease phenotype and has enabled the understanding of the cancer biology and aided in developing novel therapies. Here we discuss both mouse model and cell culture based approaches to CaP model systems.

1.3.1 Mouse Models

Mouse models have been shown to be immensely valuable in comprehending CaP cancer biology. Knockouts of genes in a prostate specific fashion have elucidated the associations with CaP progression. One of the most commonly used mouse models for CaP is the transgenic adenocarcinoma of the mouse prostate (TRAMP) mouse in which the Simian virus SV40 large and small T antigens are expressed under the regulation of a rat probasin promoter(279). In the TRAMP mouse, CaP pathogenesis develops within 12 weeks of birth with initial PIN lesions being seen and metastatic spread occurring at 30 months. Another model also utilizing the SV40 large t-antigen is the LADY mouse model. This is an early model of CaP progression with PIN lesions developing at 20 months and no metastatic spread observed(280). Of use have also been models using c-myc(281) and a heterozygous PTEN mouse model(282). The PTEN model shows that loss of PTEN recapitulates features of human CaP progression and mice also lacking p27 present with more aggressive tumours with earlier onset, albeit without metastasis. A rat model of CaP, known as the Dunning rat model, was one of the first to be developed
and used widely\cite{283}. It has shown utility for the study of AIPC growth of CaP cells and the molecular determinants of metastases.

Due to the \textit{in vivo} nature of the mouse model the interplay between host responses and the tissue microenvironment can result in a representative phenotype that recapitulates the morphology in humans. However, due the complex nature of these interactions, elucidating the mechanisms involved become more of a challenge, added to this is the maintenance of a colony of mice. A canine model for PIN development has also been used, albeit their long life span makes them unsuitable for many studies\cite{284}. With respect to the identification of circulating protein biomarkers for CaP in animal models, similar issues are present as with analyzing human specimens. In this case a simplified cell culture based approach may be warranted to study CaP cells in a homogenous environment.

\subsection{Cell Culture Models}

The use of cell lines derived from prostate tumours allows the analysis of prostate tumour cells in a simplified and controlled environment. There have been numerous cell lines derived from prostate tumours, albeit most have been from metastatic sites\cite{73}. Prominent cell lines that are commonly used include the PC3, LNCaP, DU-145, 22Rv1 and MDA PCa 2b. All are obtained from metastatic sites except for 22Rv1 which was obtained from a localized tumour to the prostate. Normal epithelial phenotype cells have been developed such as the RWPE-1 cell line series, albeit their sustained growth in culture draws question to the extent of the normal phenotype.
There are several advantages and disadvantages to using cancer cell lines over animal models. These then dictate the nature of the experiment that can be conducted. Firstly, the cost involved with maintaining them is significantly less than maintaining mice. They are readily available and studies can be performed relatively quickly. Large quantities and volumes of cells may be propagated to create high-throughput studies. Cell lines are extremely versatile in the types of studies they may be used in. Not only can they be grown in vitro but also can be injected into mice to form xenograft models of CaP progression. They can be modified and studied over time to determine sequential events that occur as a result of specific stimulus. As well as the products produced from the cells such as their ‘secretome’ can be analyzed readily. Disadvantages that are associated with cell lines are that they do not represent the heterogeneity of the tumour microenvironment as well as the necessarily heterogeneous nature of tumours with a patient and between patients. As a result multiple cell lines may be required to address the full heterogeneity seen in a tumour phenotype. Cell lines are also subject to genetic alterations in culture that may alter their phenotype over the course of a long experiment. The path to the progression of the tumour is lost and does not provide insight in the pathogenic process necessarily.

Developing CaP cell lines has been regarded as being notoriously difficult. As a result the CaP cell lines that are utilized today have been around for quite some time and have been characterized very well. Each cell has shown to have its own unique phenotype which has become useful in comparing the proteins secreted by each in the search for CaP biomarkers(285).
1.4 Continued Need for Novel Prostate Cancer Biomarkers

The discovery of PSA and its introduction in the clinic in the early 1990’s has had a profound impact on the early diagnosis of CaP and resulted in an increase in CaP incidence(286). PSA is currently used as a dichotomous marker for diagnosis but it is now being realized that its values represent a relative degree of risk for CaP(287). The upper limit of normal set at 4ug/L fails to detect a significant number of cancers and the PCPT determined that there is no level of PSA where cancer can be ruled out(288). Measurement of total PSA has been shown to be useful as a prognostic tool, with high preoperative values associated with advanced disease and poor clinical outcome(104). The controversy surrounding the use of this marker is being currently debated since it is not clear if PSA screening has led to a decline in mortality due to CaP(289). In 2008 and 2009 two major randomized prospective clinical trials will report if PSA screening reduces mortality: the European Randomized study of Screening for Prostate Cancer and Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. The relationship of PSA with tumour grade is also not clear. It has been shown that tissue PSA decreases with increasing Gleason sum(54), albeit serum levels increase due to disruption of the basement membrane surrounding the prostate epithelial cells and the overall prostate tissue architecture. PSA is not specific for prostate cancer and can serve as a marker for benign prostate hyperplasia and prostate volume growth(290). Key statistics for the test have been shown to be inadequate, with positive predictive values of 37% and patients in the grey zone of 4-10ug/L having 25% chance of displaying CaP(291) and 15% of men with levels of <4ug/L displaying CaP(287). Due to the inadequacies
of PSA there is a need for novel markers of CaP to prevent over treatment of indolent tumours. In addition to diagnosis, prognostic, predictive and therapeutic markers are needed to act as surrogate endpoints in forecasting disease severity, choosing treatments and monitoring response to therapies, respectively.

1.5 Rationale and Objectives of the Present Study

1.5.1 Rationale

Advancements in sequencing technologies for genomics have led to the completion of many genomes including the human(292). Coupled with advances in mass spectrometry, particularly soft ionization techniques for MS such as ESI and MALDI, has of recent made a large impact in the field of biomarker discovery(293,294). These new advances have enabled the generation of large databases housing theoretical protein sequences of all the genes in an organism. With these advancements, the identification of complex mixtures of proteins by MS has become routine. In our global analysis of proteins present in the CM of CaP cell lines we will utilize this technology in our search for novel prostate tumour markers.

We have chosen a cell line based model for identifying secreted proteins for our discovery approach. By collecting and concentrating CM produced from cell lines, the proteins secreted from the cells will accumulate in the CM and be present at higher levels than those found in tissues, thereby making their identification through MS more facile. In addition, given that the cell lines to be used are specific to epithelial metastatic CaP cells, the proteins present in the CM will originate from the cancer cell specifically and not from the stroma environment, thereby
avoiding unnecessary complications in our analysis. This will eliminate any extraneous proteins identified that are not specific to the tumour. To select our candidates, our focus will be on secreted or shed membrane proteins, since in the past they have proven to be the most useful as biomarkers\(^{(295)}\). In addition, using prostate cell lines derived from clinically relevant tumours will aid in our discovery of tumour markers that are specific to clinically relevant tumours.

Through qualitative analysis of the PC3, LNCaP and 22Rv1 malignant prostate cell lines, proteins identified will be investigated for their potential as candidate tumour markers. We assume that these molecules have not yet been identified because their concentration in serum is too low and therefore cannot be measured or purified, unless specific immunological reagents and highly sensitive enzyme linked immunosorbent assay (ELISA) methods are available.

To facilitate our analysis of the CM, we will use protein and peptide free chemically defined serum-free media (SFM) in our cell culture system. This will provide us with a distinct advantage over fetal calf serum (FCS) containing media, since all the proteins secreted into the media will be from the cell line and can thus be easily identified without worry of contaminating proteins from the FCS. Conditioned cell culture media is also significantly easier to characterize compared to human serum or plasma which contains many substances that interfere with its proteomic analysis such as lipids. In addition, there are many high abundance proteins in human serum that will mask the detection by MS of lower abundant proteins. There are also concerns regarding sample acquisition and handling that
makes blood samples an imperfect source when performing biomarker discovery studies (296).

1.5.2 Hypothesis

Proteins produced by tumour cells are secreted or shed into the circulation and may act as surrogate tumour markers. These tumour markers can be used to aid in the diagnosis and prognosis of CaP patients. We hypothesize that candidate protein tumour markers for the early detection of clinically significant CaP are secreted in vitro by CaP cell lines into their tissue culture media. These proteins can subsequently be identified through qualitative proteomic analysis of the CM by MS based proteomics. Candidate tumour markers identified will then be selected to validate their clinical utility in serum and tissues of patients with and without CaP.

1.5.3 Objectives

1) Perform a proof of principle study to demonstrate the proteomic analysis of the secretome of the PC3(AR)6 cell line by mass spectrometry

   a) Optimize the culture of PC3(AR)6 in large volumes of SFM for an extended period of time by measuring control proteins KLK5, KLK6 and total protein

   b) Collect and prepare the CM for fractionation by fast performance liquid chromatography

   c) Process the fractions for liquid chromatography tandem mass spectrometry
d) Analyze the identified proteins and select candidates for further validation through bioinformatics and literature searches

e) Validate candidates in serum of patients with and without CaP

2) Conduct and extended proteomic analysis of the CM of three CaP cell lines and a control flask
   
   a) Optimize culture of each cell line in SFM to optimize secreted protein production and reduce release of intracellular proteins

   b) Collect and prepare CM for peptide fractionation through high performance liquid chromatography

   c) Process samples for analysis by liquid chromatography tandem mass spectrometry

   d) Analyze the identified proteins and determine protein identification probabilities and false positive rates

   e) Subject the list of proteins to bioinformatics and literature searches and select novel candidates for validation

3) Perform a pre-clinical validation on selected candidates in serum of patients with and without CaP

4) Conduct and extended validation on the most promising candidate in the serum and tissues of patients with and without CaP
CHAPTER 2:
PROOF OF PRINCIPLE: PROTEOMIC ANALYSIS OF THE PC3 CELL LINE CONDITIONED MEDIA

The work presented in this chapter was published in Clinical Chemistry:

Sardana, G., Marshall J. and Diamandis, E.P.

*Discovery of Candidate Tumor Markers for Prostate Cancer via Proteomic Analysis of Cell Culture–Conditioned Medium*


Copyright permission has been granted.
Chapter 2: Proof of Principle: Proteomic Analysis of the PC3 Cell Line Conditioned Media

2.1 Introduction

Prostate cancer is the most common malignancy in men and the 2nd leading cause of cancer-related deaths (297). Early diagnosis of cancer improves clinical outcomes, but detection methods for clinically relevant preclinical CaP are limited (298). Tumour markers can be used to detect cancer, determine prognosis, or monitor treatment (299). The established CaP tumour marker PSA, has low diagnostic specificity (300); increased concentrations are also seen in BPH and prostatitis (301). The sensitivity and specificity of the PSA test for CaP have been improved by modifications such as measuring PSAV and measuring the ratio of fPSA to tPSA or KLK2 in addition to PSA (302-304). Other methods for detecting CaP are not highly specific for CaP and are uncomfortable to patients (305). Serum proteomic profiling has emerged as a new method for detecting CaP (306,307) but has not been adequately validated (308,309).

Determining the clinical significance of a prostate tumour is a major concern of CaP testing (310). An autopsy study of men who died of other causes revealed CaP or precursor lesions in 29% of men 30 to 40 years old and 64% of those >60 years old (311). Because treatments for CaP (androgen ablation, radical prostatectomy, radiation, and chemotherapy) have serious side effects, there is a need to differentiate patients who require treatment from those who do not. Mass spectrometry for biomarker discovery (312-314) has generated large databases of protein sequences. We used a cell culture–based proteomic approach to search for novel candidate prostate tumour markers in the proteins secreted into the CM of the CaP cell line PC3(AR)6.
Chapter 2: Proof of Principle: Proteomic Analysis of the PC3 Cell Line
Conditioned Media

2.2 Materials and Methods

2.2.1 Roller Bottle Cell Culture

We grew the CaP epithelial cell line PCR(AR)₆, kindly provided by Dr. Theodore Brown (Toronto, Ontario, Canada), in a humidified incubator at 37°C and 5% CO₂ in RPMI 1640 (Gibco) with 80 mL/L fetal calf serum (HyClone) to confluence (20 × 10⁶ cells/flask) in two 175-cm² tissue culture flasks (Nunc). The cells were trypsinized and transferred to an 850-cm² roller bottle flask with a vented cap (Corning), placed on a roller culture apparatus (Wheaton Science Products), and incubated for 2 days in 150 mL RPMI with 8% FCS to allow the cells to adhere (Figure 2.1). Afterwards, the medium was discarded and the interior was rinsed twice with 150 mL phosphate-buffered saline (PBS) (137 mmol/L NaCl, 10 mmol/L phosphate, 2.7 mmol/L KCl, pH 7.4). Next, 400 mL of chemically defined Chinese hamster ovary medium (CDCHO) (Gibco), supplemented with glutamine (8 mmol/L) (Gibco), was added, and the roller bottle was incubated for 14 days. During the culture period, we measured total protein by the Coomassie (Bradford) assay (Pierce Biotechnology) and KLK5 and KLK6 by ELISA (192,315). The CM was collected, spun down (3,000g) to remove cellular debris, and frozen at −20°C for later use. We processed 2 replicates for MS analysis; these replicate cultures are referred to as batch 1 and batch 2.
Figure 2.1: Schematic representation of the workflow for proteomic analysis of roller bottle CM. CM from roller bottles was collected and dialyzed overnight. The dialyzed medium was directly loaded onto a SAX column and eluted by FPLC. Ten fractions were collected, lyophilized, and trypsin-digested. The resulting peptides were ZipTip desalted and separated by reversed-phase C-18 chromatography coupled online to an ion-trap mass spectrometer. The acquired MS/MS data were searched by Mascot, and identified proteins were manually categorized by Genome Ontology and literature searches through NCBI.
2.2.2 Dialysis

The thawed CM was dialyzed in tubing with a molecular weight cut-off of 3.5 kDa (Spectra/Por) in 10 L of 20 mmol/L diethanolamine (DiEtOH) (Sigma-Aldrich), pH 8.9, overnight at 4°C. A sample aliquot was taken after dialysis.

2.2.3 Fast Performance Liquid Chromatography of CM

The dialyzed CM was loaded onto an HR10/10 column (GE Amersham) containing SOURCE15 strong anion exchange (SAX) beads (GE Amersham). An AKTA fast-performance liquid chromatography system (FPLC) was used running Unicorn v4.12 software equipped with a P-960 sample pump and Frac-900 fraction collector (GE Amersham), at a flow rate of 1 mL/min followed by a 2-stage elution gradient at a flow rate of 3 mL/min (0% to 60% elution buffer within 40 min followed by a ramp from 60% to 100% within 10 min) using 20 mmol/L DiEtOH, pH 8.9 running buffer, and 1 mol/L NaCl elution buffer. Absorbance was monitored at 214 nm. The first 10 fractions of 10mL each were collected, taking sample aliquots from each, and the flow-through. A SAX protein standard (Bio-Rad) was run each time to evaluate the quality of the column before each sample loading.

2.2.4 Lyophilization and Digestion of Fractions

The collected fractions were lyophilized overnight to dryness, resuspended in 1 mL dH₂O, and trypsin-digested using a 10X digest buffer [5% acetonitrile (ACN), 200 mmol/L urea, and 50 mmol/L tricine, pH 8.8] to digest ~100μg protein from each lyophilized fraction; 1μg trypsin (Promega) was used per digest. The digests were
incubated overnight at 37°C and reduced the following day with dithiothreitol (DTT) (1 mmol/L) for 1 h at 25°C. We added a final 1μg trypsin, and the samples were incubated at 37°C for approximately 3 h.

2.2.5 Liquid Chromatography–MS Analysis

Digested samples were collected on a C-18 ZipTip (Millipore) to purify and desalt the peptides. The ZipTip was primed with 50% ACN in 0.1% acetic acid and washed with 0.1% acetic acid before the digested samples were passed through the ZipTip. The peptides were eluted from the ZipTip with 2μL of 0.1% acetic acid in 65% ACN, and dH₂O was added to give a final volume of 20μL. The desalted peptides were injected at 2μL/min onto a C-18 reversed-phase (RP) chromatography column (Vydak 300μm X 15 cm) via an Agilent 1100 series HPLC system coupled to a Bruker HCT ion-trap ESI mass spectrometer (Bruker Daltronics) via a metal electrospray needle. The sample was injected in 5% acetonitrile in 0.1% acetic acid and, after loading for 5 min, a 1-min gradient to 12.5% acetonitrile was followed by a 90-min gradient to 65% acetonitrile in 0.1% acetic acid. We analyzed the eluted peptides by tandem MS (MS/MS), and data were acquired and mass spectral peaks were deconvoluted to consolidate different charge states observed for each peptide with the software supplied by Bruker. The instrument was standardized with a tryptic digest of alcohol dehydrogenase, cytochrome C, and glycogen phosphorylase to assess mass accuracy and sensitivity of the instrument before and after each set of runs.
2.2.6 Database, Genome Ontology, and Literature Search

The resulting MS/MS spectra was searched in MGF format from both batch 1 and 2 using the Mascot algorithm search engine (version 2.1) with default variables and trypsin specified. Specifically, one missed cleavage was allowed, a variable oxidation of methionine residues and a fixed modification of carbamidomethylation of cysteines was set with a fragment tolerance of 0.4 Da and a parent tolerance of 3.0 Da. The database used was a custom-built non-redundant compilation of human, mouse, and rat sequences from GenBank, Ensembl, and SwissProt, compiled January 2005. A bioinformatics program was used through Protana Inc. (Mississauga) to identify peptides from the MS/MS spectra present from each fraction, giving each peptide a score. The identified peptides from all the fractions within each respective batch were clustered with other peptides that were common to a particular protein, and each group of peptides was then given a cluster score. Any peptide with a Mascot score <20 and any protein with a score <40 was removed from the data. The identified proteins were manually analyzed and classified by their genome ontology cellular component classification and PubMed literature searches were conducted on each protein.

The false-positive rate of protein identification was measured by searching a random database, in which every sequence entry from the "normal" database was randomly shuffled. The number of hits from each search was categorized based on score, and for each scoring interval, the false-positive rate was calculated as number of random hits/(number of random hits + number of normal hits).
2.2.7 ELISAs for Kallikrein 5, 6, and 11

Kallikreins 5, 6, and 11 were measured by sandwich-type ELISA, 96-well plates, as described earlier (192,198,315) with use of biotinylated detection antibodies, alkaline phosphatase-conjugated streptavidin and diflunisal phosphate substrate. Plates were read by time-resolved fluorescence(316).

2.2.8 Protein Recovery

To assay for sample recovery of proteins during dialysis and fractionation, we analyzed sample aliquots that had been taken during this procedure for KLK5 and KLK6 by the aforementioned ELISA assays.

2.2.9 Mac-2BP ELISA

We obtained the s90K/Mac-2BP ELISA reagent set from Bender Medsystems. Briefly, serum samples were diluted 1:500 and CM samples 1:10 in sample diluent buffer (provided by the manufacturer) and loaded onto 96-well strips pre-coated with anti–Mac-2BP antibody. Samples were incubated at 37°C for 45 min with shaking at 100 rpm. The wells were washed 3 times with wash buffer (as provided), a detection antibody was added, and the plate was incubated for 45 min. The plate was washed and substrate solution (as provided) was added to each well, and the plate was incubated with shaking at room temperature for 10 min, following which a stop solution (as provided) was added to each well. Absorbance was measured at 490 nm by a Wallac–Victor² plate reader (Perkin-Elmer). An extended validation of Mac-2BP was performed with 210 CaP, 53 BPH and 17 normal serum
samples donated by Carsten Stephan from the Department of Urology, Charité -
Universitätsmedizin Berlin, Germany.

2.3 Results

2.3.1 Total Protein, KLK5 and KLK6 Concentrations in Culture Over Time

To demonstrate the accumulation of secreted proteins over time in the roller
bottle culture, we measured 2 secreted proteins that are known to be produced by
the PC3(AR)6 cell line, namely KLK5 and KLK6, over the 14-day culture period
(Figure 2.2A, B). The concentrations of KLK5 and KLK6 increased with time in
culture and began to plateau after ~10 days. We measured the total protein in the
CM, which increased steadily throughout the culture period (Figure 2.2C).
Figure 2.2

**Figure 2.2**: Measurement of KLK5, KLK6 and total protein. KLK5 (A), KLK6 (B), and total protein (C) concentrations over time in CM of the PC3(AR)$_6$ roller bottle culture. The concentrations of KLK5 and KLK6 were monitored by ELISA. Two replicates are shown.
2.3.2 Recovery of KLK5 and KLK6 During Sample Preparation

The recoveries of KLK5 and KLK6 were 25% and 17%, respectively. The major protein losses were seen after lyophilization. In addition, for KLK5, a significant amount went in the flow-through after column loading.

2.3.3 Proteins Identified by Mass Spectrometry

After LC-MS/MS and searching by Mascot from both batches, we identified 262 proteins from the SAX FPLC fractions that had a Mascot score of at least 40. Each protein identified was tabulated and cross-referenced with the genome ontology database for cellular components (317) (Figure 2.3). A large percentage (39%) of all proteins identified are classified as extracellular (23%) or membrane (16%) proteins. Many identified proteins are classified as intracellular (50%), whereas 11% were unclassified. From the list of the 262 proteins, we selected candidate biomarkers (Table 2.1) based on the following criteria:

1. Proteins were searched manually against the Genome Ontology database (317) for their cellular localization. Proteins that were classified as secreted and membrane-bound were selected.

2. Literature searches through the National Center of Biotechnology Information (NCBI) PubMed database were then performed to determine:
   a. if these proteins are novel molecules that have yet to be explored as potential biomarkers;
   b. if these proteins are known to participate in critical pathways implicated in cancer initiation and progression.
Chapter 2: Proof of Principle: Proteomic Analysis of the PC3 Cell Line
Conditioned Media

**Figure 2.3**

Classification of proteins by cellular location from PC3(AR)$_6$ CM, batches 1 and 2. Each protein identified after Mascot searching was classified by its cellular location using Genome Ontology classifiers (www.geneontology.org).

**Figure 2.3: Cellular location of proteins from PC3(AR)$_6$ CM.** Classification of proteins by cellular location from PC3(AR)$_6$ CM, batches 1 and 2. Each protein identified after Mascot searching was classified by its cellular location using Genome Ontology classifiers (www.geneontology.org).
Table 2.1: Extracellular candidate tumor markers identified in culture medium of the PC3 (AR)6 roller bottle culture.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein Description</th>
<th>Mascot Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDF15</td>
<td>Prostate differentiation factor; PTGF-β</td>
<td>40.81</td>
</tr>
<tr>
<td>COL2A1</td>
<td>Alpha 1 type II collagen, isoform 1, preproprotein; collagen II</td>
<td>44.96</td>
</tr>
<tr>
<td>PAM*</td>
<td>Peptidylglycine α-amidating monooxygenase, isoform c, preproprotein</td>
<td>45.64</td>
</tr>
<tr>
<td>IGFBP2</td>
<td>Insulin-like growth factor binding protein 2 precursor (IGFBP-2) (IBP-2)</td>
<td>45.7</td>
</tr>
<tr>
<td>S100A8*</td>
<td>S100 calcium-binding protein A8; cystic fibrosis antigen; calgranulin A</td>
<td>47.9</td>
</tr>
<tr>
<td>TLR9*</td>
<td>Toll-like receptor 9, isoform A precursor</td>
<td>49.44</td>
</tr>
<tr>
<td>TPT1*</td>
<td>Tumor protein, translationally controlled 1</td>
<td>50.73</td>
</tr>
<tr>
<td>CTSL*</td>
<td>Cathepsin L</td>
<td>52.09</td>
</tr>
<tr>
<td>PLG</td>
<td>Plasminogen</td>
<td>54.92</td>
</tr>
<tr>
<td>LGALS3</td>
<td>Lectin, galactoside-binding, soluble, 3 (galectin 3); Lectin, galactose-binding</td>
<td>55.53</td>
</tr>
<tr>
<td>LTBP2*</td>
<td>Latent transforming growth factor β binding protein 2</td>
<td>65.36</td>
</tr>
<tr>
<td>MUC5B*</td>
<td>Mucin 5, subtype B, tracheobronchial</td>
<td>66.33</td>
</tr>
<tr>
<td>GALNT6*</td>
<td>Polypeptide N-acetylgalactosaminyltransferase 6</td>
<td>66.35</td>
</tr>
<tr>
<td>IGFBP6</td>
<td>Insulin-like growth factor binding protein 6</td>
<td>72.26</td>
</tr>
<tr>
<td>INHBB</td>
<td>Inhibit β B subunit precursor; Inhibit, β-2</td>
<td>74.86</td>
</tr>
<tr>
<td>TIMP2</td>
<td>Tissue inhibitor of metalloproteinase 2 precursor</td>
<td>75.21</td>
</tr>
<tr>
<td>LAMC2</td>
<td>Laminin, γ 2, isoform a precursor; nicein (100 kDa);</td>
<td>75.96</td>
</tr>
<tr>
<td>CXCL3*</td>
<td>Macrophage inflammatory protein-2-β precursor (MIP2-β)</td>
<td>80.8</td>
</tr>
<tr>
<td>AZGP1</td>
<td>α-2-glycoprotein 1, zinc; α-2-glycoprotein</td>
<td>82.29</td>
</tr>
<tr>
<td>NPC2*</td>
<td>Niemann-Pick disease, type C2; Niemann-Pick disease, type C2 gene</td>
<td>82.8</td>
</tr>
<tr>
<td>MDK*</td>
<td>Midkine (neurite growth-promoting factor 2)</td>
<td>90.34</td>
</tr>
<tr>
<td>SERPINA1</td>
<td>Serine (or cysteine) proteinase inhibitor</td>
<td>93.92</td>
</tr>
<tr>
<td>KLK6*</td>
<td>Kallikrein 6 (neurosin, zyme); protease M; protease, serine, 9 (neurosin)</td>
<td>95.71</td>
</tr>
<tr>
<td>SEMA3F*</td>
<td>Semaphorin 3F precursor (Semaphorin IV) (Sema IV) (Sema III/F)</td>
<td>100.2</td>
</tr>
<tr>
<td>C19orf10*</td>
<td>Chromosome 19 open reading frame 10; interferleukin 25; interferleukin 27</td>
<td>103.18</td>
</tr>
<tr>
<td>LRG1</td>
<td>Leucine-rich α-2-glycoprotein</td>
<td>109.66</td>
</tr>
<tr>
<td>COL6A2</td>
<td>α 2 type VI collagen, isoform 2C2</td>
<td>109.73</td>
</tr>
<tr>
<td>GALNT2</td>
<td>Polypeptide N-acetylgalactosaminyltransferase 2; UDP-GalNAc transferase 2</td>
<td>114.5</td>
</tr>
<tr>
<td>IGFBP5</td>
<td>Insulin-like growth factor binding protein 5</td>
<td>120.38</td>
</tr>
<tr>
<td>RNASET2*</td>
<td>Ribonuclease 6 (EC 3.1.27.—) precursor</td>
<td>124.53</td>
</tr>
<tr>
<td>FBLN1</td>
<td>Fibulin 1 isoform C precursor</td>
<td>137.3</td>
</tr>
<tr>
<td>DAG1*</td>
<td>Dystroglycan 1 precursor; α-dystroglycan; Dystrophin-associated glycoprotein-1</td>
<td>145.92</td>
</tr>
<tr>
<td>S100A9</td>
<td>S100 calcium-binding protein A9; calgranulin B</td>
<td>152.1</td>
</tr>
<tr>
<td>TGFB2</td>
<td>Transforming growth factor, β 2</td>
<td>152.59</td>
</tr>
<tr>
<td>SPINT1*</td>
<td>Hepatocyte growth factor activator inhibitor precursor</td>
<td>154.24</td>
</tr>
<tr>
<td>B2M</td>
<td>β-2-microglobulin</td>
<td>177.47</td>
</tr>
<tr>
<td>LCN2*</td>
<td>Lipocalin 2 (oncogene 24p3)</td>
<td>181.58</td>
</tr>
<tr>
<td>KLK5*</td>
<td>Kallikrein 5; stratum corneum tryptic enzyme</td>
<td>181.94</td>
</tr>
<tr>
<td>LCN7*</td>
<td>P3ECSL; glucocorticoid-inducible protein; oxidized-LDL responsive gene 2</td>
<td>184.67</td>
</tr>
<tr>
<td>DF</td>
<td>Complement factor D preproprotein; adipsin; properdin factor D</td>
<td>187.96</td>
</tr>
<tr>
<td>SFN*</td>
<td>Stratifin</td>
<td>193.8</td>
</tr>
<tr>
<td>APP*</td>
<td>Amyloid β (A4) precursor protein (protease nexin-II, Alzheimer disease)</td>
<td>215.47</td>
</tr>
<tr>
<td>Protein</td>
<td>Description</td>
<td>MW (Da)</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>---------</td>
</tr>
<tr>
<td>TIMP1</td>
<td>Tissue inhibitor of metalloprotease 1 precursor; Erythroid-potentiating activity</td>
<td>227.55</td>
</tr>
<tr>
<td>CTS</td>
<td>Complement component 1, s subcomponent</td>
<td>254.89</td>
</tr>
<tr>
<td>GRN</td>
<td>Granulin</td>
<td>258.53</td>
</tr>
<tr>
<td>PLTP*</td>
<td>Phospholipid transfer protein</td>
<td>282.02</td>
</tr>
<tr>
<td>STC2*</td>
<td>Stanniocalcin 2; stanniocalcin 2; stanniocalcin-related protein</td>
<td>303</td>
</tr>
<tr>
<td>PTX3*</td>
<td>Pentaxin-related gene, rapidly induced by IL-1β; Pentraxin-3</td>
<td>308.44</td>
</tr>
<tr>
<td>CYR61*</td>
<td>Cysteine-rich, angiogenic inducer, 61; cysteine-rich heparin-binding protein 61</td>
<td>327.5</td>
</tr>
<tr>
<td>IL6</td>
<td>Interleukin 6 (interferon, β 2)</td>
<td>336.16</td>
</tr>
<tr>
<td>PSAP*</td>
<td>Saposin precursor</td>
<td>359.65</td>
</tr>
<tr>
<td>PLAU</td>
<td>Plasminogen activator, urokinase</td>
<td>361.2</td>
</tr>
<tr>
<td>FSTL1*</td>
<td>Follistatin-like 1 precursor; follistatin-related protein</td>
<td>367.11</td>
</tr>
<tr>
<td>P4HB*</td>
<td>Procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase)</td>
<td>375.33</td>
</tr>
<tr>
<td>AGRN*</td>
<td>Agrin</td>
<td>486.69</td>
</tr>
<tr>
<td>HSPG2*</td>
<td>Basement membrane-specific heparan sulfate proteoglycan core protein precursor</td>
<td>506.82</td>
</tr>
<tr>
<td>INHBA*</td>
<td>Inhibin β A subunit precursor; Inhibin, β-1; EDF</td>
<td>507.77</td>
</tr>
<tr>
<td>NUCB1*</td>
<td>Nucleobindin 1 precursor (CALNUC)</td>
<td>631.69</td>
</tr>
<tr>
<td>LGALS3BP*</td>
<td>Galectin 3 binding protein; L3 antigen; Mac-2-binding protein; serum protein 90K</td>
<td>632.55</td>
</tr>
<tr>
<td>COL6A1</td>
<td>Collagen, type VI, α 1 precursor; collagen VI, α-1 polypeptide</td>
<td>766.46</td>
</tr>
<tr>
<td>CTSD</td>
<td>Cathepsin D (lysosomal aspartyl protease)</td>
<td>771.08</td>
</tr>
<tr>
<td>THBS1</td>
<td>Thrombospondin 1 precursor</td>
<td>865.57</td>
</tr>
<tr>
<td>BF</td>
<td>Complement factor B preproprotein; B-factor, properdin; C3 proactivator</td>
<td>1071.18</td>
</tr>
<tr>
<td>FN1</td>
<td>Fibronectin 1 isoform 2 preproprotein; cold-insoluble globulin</td>
<td>1978.92</td>
</tr>
<tr>
<td>C3</td>
<td>Complement component 3</td>
<td>3339.99</td>
</tr>
</tbody>
</table>

*Proteins linked to cancer in previous literature reports.
We determined the overlap of the proteins identified between the 2 batches through visual inspection of the Mascot data.

Proteins in this list (Table 2.1) that are marked with an asterisk have not been previously evaluated as serum biomarkers for CaP as determined through PubMed literature searches specific for the protein.

We calculated false-positive protein identification rates based on a random database search. False-positive rates for specific scoring intervals were as follows: 40–50, 44%; 50–60, 35%; 60–70, 7%; 70–80, 8%; 90–150, 7%; and >150, 0%. The presence of a higher false-positive rate in the 70–80 scoring interval is the result of a statistical fluctuation attributable to additional protein identification in the random database search.

### 2.3.4 Overlap of Proteins Identified in Batches 1 and 2

To determine the reproducibility of the method, the proteins identified from batches 1 and 2 were manually compared for overlap. We identified 145 proteins in both batches (55% overlap). Additionally, we identified 78 proteins only in batch 1 and 39 proteins only in batch 2. Combined, the total number of identified proteins was 262. As expected, the more abundant proteins were preferentially identified in both batches.
2.3.5 Mac-2BP Concentrations in Patients with Prostate Cancer vs. Healthy Men and in the Conditioned Media of the PC3(AR)$_6$ Cell Line

From the proteins identified, Mac-2BP was chosen as one novel biomarker candidate for further validation. The concentrations of Mac-2BP increased in the CM over time (Figure 2.4A), as expected and in a similar fashion to KLK5 and KLK6 (Figure 2.2). We measured serum concentrations of Mac-2BP from 26 men with CaP and 17 healthy men using a Mac-2BP ELISA. The median Mac-2BP concentrations in CaP patients were almost twice as high as those in healthy men (Figure 2.4B), with 50% of the CaP patients having increased Mac-2BP concentrations compared with the healthy men (at the 100th percentile of healthy men as a cutoff). The difference in medians of the 2 populations by Mann-Whitney test was highly significant ($P = 0.003$). The negative correlation between Mac-2BP and PSA in these 26 patients was also significant, with a Spearman correlation coefficient ($r_s$) of $-0.63$ ($P < 0.001$) (Figure 2.4C).

We further measured KLK5, KLK6, and KLK11 (a previously identified prostate and ovarian cancer biomarker) (318) in the same set of patients and controls, as above. The data (Figure 2.5) showed decreased concentrations of KLK5 in CaP ($P < 0.0001$ by Mann-Whitney test), decreased concentrations of KLK6 in CaP ($P = 0.03$ by Mann-Whitney test), and increased concentrations of KLK11 in CaP ($P < 0.0001$ by Mann-Whitney test).

Spearman correlations for all pairs of measured concentrations (Mac-2BP, KLK5, KLK6, KLK11, and PSA) included only one statistically significant negative correlation between Mac-2BP and PSA (Figure 2.4C).
An extended validation of Mac-2BP was performed using serum samples that were confirmed cases of CaP, BPH and healthy males. As can be seen from Figure 2.6 the trend initially observed in the preliminary validation is reversed with the median levels of CaP patients showing lower Mac-2BP serum concentrations than normals. Levels are also lower in BPH patients. A significant difference is observed by the Kruskal-Wallis non-parametric one-way ANOVA of $p = 0.0028$. Post-hoc analysis by the Dunn’s multiple comparisons test showed there was a significant difference between the CaP and normal group, $p < 0.01$. 
Figure 2.4: Mac-2BP concentrations and the correlation between serum Mac-2BP concentrations and PSA. (A), Mac-2BP concentrations in CM of PC3(AR)$_6$ cell line over time. (B), Mac-2BP concentrations in serum of 26 CaP patients and 17 healthy men. Horizontal lines indicate medians. $P$ value was calculated with the Mann–Whitney test. (C), Correlation between serum Mac-2BP concentrations and PSA in the 26 cancer patients. $r_s = $ Spearman correlation coefficient.
Figure 2.5: Concentrations of KLK5, KLK6, and KLK11 in serum of 26 CaP patients and 17 healthy men. Horizontal lines indicate medians. P value was calculated with the Mann–Whitney test.
Figure 2.6: Serum concentrations of Mac-2BP in CaP, BPH and normal patients. Horizontal lines indicate median values. Significance was determined using the Kruskal-Wallis non-parametric one-way ANOVA test.
2.4 Discussion

We used a proteomic method for identification of secreted proteins from the CM of the metastatic CaP cell line PC3(AR)$_6$ as a model to discover novel markers for CaP. Because large amounts of cells are needed for confident MS detection of low-abundance cellular proteins, we determined if a cell line could be grown in a large volume of SFM for an extended period. The use of protein- and peptide-free chemically defined Chinese hamster ovary serum-free medium simplified analysis, providing a distinct advantage over fetal calf serum–containing medium, which would contaminate the CM.

The loss of KLK5 in the flow-through during SAX FPLC was attributable to incomplete capture of KLK5 by the SAX column. The incomplete capture is consistent with its relatively high pI of ~ 8. Appreciable losses of both KLK5 and KLK6 also occurred after lyophilization of the FPLC fractions, possibly attributable to incomplete solubilization of the freeze-dried protein (319).

Independent MS detection of KLK5 and KLK6, proteins known to be secreted by PC3(AR)$_6$, was confirmed for batch 1 and 2 in the expected fractions. The complex mixture of proteins present at varying concentrations in CM necessitated fractionation before MS to increase the depth of identification. However, not all proteins in a mixture can be ionized and detected in 1 run, with the lower abundance proteins not being identified in both batches (320).

Fifty percent of the proteins identified were intracellular. Their presence in the CM is to be expected because of cell death and their high abundance within cells. Because we were primarily interested in investigating proteins that are secreted or
shed from CaP cells in vivo, we selected the extracellular and membrane proteins for further evaluation. Each protein was examined to establish if it had been previously evaluated as a CaP biomarker or if it had any link to cancer. The selected candidates are listed in Table 2.1.

We performed preliminary validation of Mac-2BP by ELISA. Mac-2BP has been shown to be a serum prognostic marker in lymphoma (321), and lung (322), breast (323), hepatocellular (324), ovarian (325), and colon (326,327) carcinoma. Serum concentrations have not been evaluated in CaP, however, despite the correlation of immunohistochemical staining for Mac-2BP with Gleason grade (328). Serum Mac-2BP was increased in 50% of the CaP patients. (Figure 2.4). The correlation of Mac-2BP concentrations and PSA concentrations in these patients was weak and negative (Figure 2.4C). As can be seen in our extended validation of Mac-2BP, the trend seen in the preliminary validation was reversed (Figure 2.6). Even though Mac-2BP is seen to be elevated at the tissue level in CaP this was not translated into an elevation seen in serum in our larger sample set. Upon further study of Mac-2BP it was realised that it is ubiquitously expressed in almost every tissue in the human body (Unigene database). As a result, the elevations due to CaP most likely are not discernable among the background of endogenous Mac-2BP in serum from other tissues. As a result we did concluded that Mac-2BP would not be a useful serum tumour marker for CaP as it is not specific to the prostate and elevations due to CaP are not detectable in serum. Two kallikreins (KLK5 and KLK6) were also present at lower concentrations in CaP, whereas KLK11 was present at increased concentrations. These data support the theory that secreted
proteins are adjunct biomarkers for CaP, although with less diagnostic accuracy than PSA. No correlation was seen between any pairs of these markers in serum.

The results of this and other similar studies (97,329) suggest that a wealth of knowledge is obtainable by analyzing the CM of cell lines. We observed minimal overlap of our data with those of Martin et al.(97) and Lin et al.(330), who studied the proteins secreted and present in the LNCaP cell line, with 67 proteins from the Martin et al. study and 27 proteins from the Lin et al. study overlapping. This highlights the heterogeneity of cell lines and the data that can be derived from each. Obviously, much work is needed to further evaluate the identified candidate biomarkers, but this study will form the basis of future communications.
CHAPTER 3:

OPTIMIZATION OF CELL CULTURE AND PROTEOMIC WORKFLOW

Sections of this chapter were published in:

Sardana, G., Jung, K., Stephan, C. and Diamandis, E.P.
Proteomic Analysis of Conditioned Media from the PC3, LNCaP and 22Rv1 Prostate Cancer Cell Lines: Discovery and Validation of Candidate Prostate Cancer Biomarkers
Journal of Proteome Research, 2008 Aug 1;7(8):3329-3338

Copyright permission has been granted.
3.1 Introduction

As with most proof of principle studies, based on the results obtained an optimization is needed before further experiments are conducted. From the data shown in the previous chapter there were some key issues that were identified that required improvement.

First, proteins identified from one cell line are not indicative of the heterogeneity of a disease phenotype. Second, the culture volume and amount of total protein produced was not controlled for the capacity of the mass spectrometer used. Third, the amount of cell death was not controlled for during culture with ~50% of the identified proteins identified were from an intracellular source. Fourth, the sample preparation workflow employed was not in agreement with the published literature as an optimal method for sample preparation. Fifth, the amount of protein loss during our sample preparation steps was not quantified. Sixth, the type of mass spectrometer used was considered to have a relatively low sensitivity and accuracy in identifying peptides. Seventh, our bioinformatics approach did not fully take advantage of multiple search engines and assessing prediction probabilities as well as determining false positive rates (FPR).

Taking all these concerns into account, a modified approach was devised based on the best practices used for proteomic analysis from the literature. These changes were then incorporated and applied to modify the proteomic workflow from our proof of principle study.
3.2 Materials and Methods

3.2.1 Cell Culture

The PC3, LNCaP and 22Rv1 cell lines were purchased from the American Type Culture Collection (Rockville, MD). All cell lines were grown in T-175 culture flasks (Nunc) in RPMI 1640 culture medium (Gibco) supplemented with 8% fetal bovine serum (FBS) (Hyclone). Cells were cultured in a humidified incubator at 37°C and 5% CO₂. Cells were seeded at varying densities in duplicate. PC3 and 22Rv1 cells were grown for 2 days in 30mL of RPMI + 8% FBS. Afterwards, the medium was removed, and the flask was gently washed 3 times with 30mL of PBS (137 mM NaCl, 10 mM phosphate, 2.7 mM KCl, pH 7.4). Thirty milliliters of CDCHO (Gibco) medium supplemented with glutamine (8mmol/L) (Gibco) were added to the flasks and incubated for 2 days. The LNCaP cell line was grown as above, except that the cells were incubated for 3 days in RPMI + 8% FBS before the media were changed to CDCHO. The cell lines were grown in duplicate.

After incubation in CDCHO, the CM was collected and spun down (3,000 X g) to remove cellular debris. Measurement of total protein, lactate dehydrogenase (LDH) and PSA, KLK5, KLK6 were taken from each culture.

3.2.2 Measurement of Total Protein, LDH and PSA, KLK5 and KLK6

The total protein of the CM was measured using the Coomassie (Bradford) assay (Pierce Biotechnology) as recommended by the manufacturer.

Lactate dehydrogenase levels in the CM were measured via an enzymatic assay based on conversion of lactate to pyruvate. NADH production from NAD⁺
during this reaction was monitored at 340nm with an automated method and converted to Units per litre (U/L) (Roche Modular Systems).

Kallikrein 3 (PSA), KLK5 and KLK6 were measured with in-house ELISAs as described earlier(192,315,331).

3.2.3 Sample Preparation

We devised a new approach to sample preparation where we monitored the level of PSA by ELISA after each sample preparation step as a marker for overall protein recovery. The sample preparation procedure was devised from the published literature. Approximately 30mL of CM from the LNCaP cell line, which corresponded to 1mg of total protein, were dialyzed overnight using a 3.5kDa cut-off dialysis tubing (Spectra/Por) at 4°C in 5L of 1mM of ammonium bicarbonate solution with one buffer exchange. The dialyzed CM was lyophilized overnight to dryness followed by resolubilization with 322µL of 8M urea, 25µL of 200mM DTT, and 25µL of 1M ammonium bicarbonate. The sample was vortexed thoroughly and incubated at 50°C for 30min. One hundred and twenty five micro litres of 500mM iodoacetamide were added and the sample was incubated in the dark at room temperature for 1h. The sample was then desalted using a NAP-5 column (GE Healthcare) and lyophilized to dryness.
3.3 Results

3.3.1 Cell Culture Optimization

Growth conditions of the three cell lines were optimized in order to reduce cell death and maximize secreted protein levels. Cells were incubated in SFM for 2 days at different seeding densities. Total protein, LDH and the concentration of KLK5 and KLK6 in the CM of PC3, and PSA in the CM of LNCaP and 22Rv1 cells were measured. The ratio of PSA, 5, and 6 concentrations with LDH levels for each culture condition (Figures 3.1, 3.2, 3.3) were compared. The optimal seeding concentrations were $7.5 \times 10^6$, $22 \times 10^6$, and $75 \times 10^6$ cells for PC3, LNCaP and 22Rv1 cell lines, respectively, as these gave the highest ratio of KLK production (indicator of secreted proteins) to LDH (indicator of cell death). The total protein of the CM for the optimized seeding densities was 33 µg/mL, 39 µg/mL and 39 µg/mL for PC3, LNCaP and 22Rv1 cells, respectively. Thus, 30mL of media contained approximately 1mg of total protein.
Figure 3.1: Measurements of PSA and LDH in the CM of the 22Rv1 cell line. Concentrations of PSA (A) and LDH (B) at different seeding densities. (C) Ratio of PSA to LDH at different seeding densities.
Figure 3.2: Measurements of PSA and LDH in the CM of the LNCaP cell line. Concentrations of PSA (A) and LDH (B) at different seeding densities. (C) Ratio of PSA to LDH at different seeding densities.
Figure 3.3: Measurements of KLK5, KLK6 and LDH in the CM of the PC3 cell line. Concentrations of KLK5 and KLK6 (A) and LDH (B) at different seeding densities. (C) Ratio of KLK5 and KLK6 to LDH at different seeding densities.
3.3.2 Sample Preparation Optimization

The consensus reached from the literature search to optimize our sample preparation workflow is shown in Figure 3.4 with the total recovery of PSA after each step also shown. The proteomic workflow of choice was found to be the MuDPiT (multi-dimensional protein identification technology) approach first shown by Washburn et al. (332) in 2001. This involved trypsin digestion of the sample followed by orthogonal chromatography steps. Strong cation exchange (SCX) followed by LC-MS/MS reversed-phase C\textsubscript{18} capillary chromatography. To employ this approach with our requirements for processing the CM, we utilized dialysis and lyophilization as intermediate steps to concentrate the sample prior to denaturing with urea, DTT and iodoacetamide. A desalting step was added to remove any salts in the sample, followed by a lyophilization step prior to resuspension in Buffer A for SCX chromatography.

In order to evaluate sample recovery PSA was measured by ELISA at each step. The results are shown in Figure 3.4. As can be seen the overall recovery is ~20% with the majority of sample loss at the lyophilization steps. Based on the results obtained here, we determined our starting total protein concentration should be 1mg. Thus, ~200ug of total protein would be fractionated prior to LC-MS/MS, and if ~10 fractions were collected, each would contain ~20ug of peptides. Considering the capillary column used during electrospray ionization into the mass spectrometer can only hold 1ug of peptides, the total concentration of peptides in each fraction would be more than enough.
Figure 3.4: Overview of the optimized sample preparation workflow. Levels of PSA are measured by ELISA after each step of the sample preparation procedure to assess sample recovery.
3.4 Discussion

In order to improve our sample coverage and efficiency of identification of proteins the methods employed in the proof of principle study needed to be optimized. This required analyzing our approach at each step of our workflow and determining the most robust procedure for sample preparation and analysis. We addressed each of our concerns raised in the proof of principle study before continuing on with our extended analysis of CaP cell lines.

Initially, our primary concern was the sample preparation procedure used to process the CM prior to LC-MS/MS analysis was not the most commonly used for this type of ‘shotgun’ proteomics application. Instead a multi-dimensional chromatographic approach that fractionated peptides rather than proteins was investigated (332). This MuDPiT approach reduced the sample handling of samples by performing one initial trypsin digest prior to the first dimension of chromatography and in addition employed a high resolution HPLC for the first chromatographic dimension. It was agreed upon that this method provided a robust approach to protein identification.

Our next concern was the culture volume used to grow the cells was too large and we did not optimize the seeding density and duration of culture for the amount of total protein produced to meet our requirements for LC-MS/MS. Based on the amount of total protein required for LC-MS/MS analysis we worked backwards and determined what would be the initial amount of total protein required before cell culture and the sample processing steps. To aid in determining this we first quantified our protein recovery for our new proteomic workflow. By measuring PSA
after each step up to trypsin digestion we were able to determine the overall recovery to be ~20% (Figure 3.4). Using this information and based on the fact that the reversed phase capillary used for LC-MS/MS can only bind 1ug of peptides, we determined that having 1mg of total protein starting material would be sufficient for our purposes.

To encompass more of the heterogeneity of the CaP phenotype we chose to expand the number of cell lines studied. In this respect, we chose to use the PC3, LNCaP and 22Rv1 CaP cell lines since each had been derived from a different metastatic site or localized from the prostate and exhibited a unique morphology and phenotype in culture. As a result each cell line had to be optimized to grow in culture to produce the optimal amount of total protein while minimizing cell death and maximizing secreted protein produced. We grew each cell line at different seeding densities for a short period in SFM. The CM was then collected and the levels of KLK’s for each cell line were measured as well as the level of LDH in the CM. The amount of KLK and LDH in the CM acted as a surrogate markers for the amounts of secreted and intracellular proteins produced by the cell respectively. To determine the optimum seeding density for each cell line the ratio of KLK to LDH was determined and the seeding density with the highest ratio was chosen. The results of the optimization for each cell line are shown in Figures 3.1, 3.2, and 3.3. As can be seen, each cell line had different growth rates and required differing seeding densities. Based on these results the appropriate seeding density for each cell line was chosen.
In addition to improvements in our sample preparation and workflow, we also aimed to improve the mass spectrometric platform used. From the Bruker ion trap used in our proof of principle study we transitioned to using the Thermo LTQ linear ion trap which has enhanced sensitivity. This is due to its ability to trap more ions and analyze them faster due to its shortened duty cycle. This translates into an increase in the number of peptides identified as well as the number of spectra identified per peptide.

To manage the large amounts of data that will be obtained from the analysis of three cell lines and the replicates semi-automated database applications were designed to assist in the annotation and comparison of the data. These included a parsing program to assist with annotation by genome ontology of proteins identified which was designed in collaboration with Adrian Pasulescu at the Samuel Lunenfeld Research Institute; a database application to sort the genome ontology annotated proteins into unique groups, by Peter Bowden at Ryerson University; a protein list comparison database application to compare proteins identified between cell lines an replicates, designed with David Nguyen at Mt. Sinai Hospital. Together these semi-automated applications will increase the efficiency at which data is analyzed.
Chapter 4: Comparative Proteomic Analysis of the Conditioned Media of Three Prostate Cancer Cell Lines

The work presented in this chapter was published in part in the Journal of Proteome Research:

Sardana, G., Jung, K., Stephan, C. and Diamandis, E.P.  
*Proteomic Analysis of Conditioned Media from the PC3, LNCaP and 22Rv1 Prostate Cancer Cell Lines: Discovery and Validation of Candidate Prostate Cancer Biomarkers*  
Journal of Proteome Research, 2008 Aug 1;7(8):3329-3338

Copyright permission has been granted.
Chapter 4: Comparative Proteomic Analysis of the Conditioned Media of Three Prostate Cancer Cell Lines

4.1 Introduction

Currently, serum PSA levels, combined with DRE, are the recommended screening tests for early detection of CaP in asymptomatic men over the age of 50(333). However, there is considerable controversy surrounding the efficacy of the PSA test in reducing the overall mortality of CaP(334,335). These concerns stem from over-diagnosis and the lack of specificity of PSA in discriminating CaP from BPH(335). While PSA has been shown to correlate very well with tumour volume(336), it is unable to predict with certainty the biological aggressiveness of the disease. Several refinements of the PSA test have been shown to increase its sensitivity and specificity(337-341). However, there is still a need to develop non-invasive tests to identify clinically relevant CaP(342).

Mass spectrometry-based proteomic technologies are currently in the forefront of cancer biomarker discovery. Serum or plasma is usually the discovery fluid of choice(343) however, several studies have employed MS for biomarker discovery in various other biological fluids(344,345). In addition, a number of alternative approaches have been used, such as MS spectra profiling(275,346), which refers to the identification of discriminatory mass spectral features without identification of the peptides being identified, peptide profiling(347,348), which refers to the identification of differentially cleaved or modified peptide sequences, and isotopic labelling(349). There are many inherent limitations to using MS for biomarker discovery in complex biological mixtures such as serum(296). The main concern is the suppression of ionization of low abundance proteins by high abundance proteins such as albumin and immunoglobulins. Depletion strategies
have been used to remove high abundance proteins which have aided in improving the detection limit of MS(350,351). One approach to overcome the limitations posed by biological fluids is to study the secretome of cell lines grown in SFM. The proteins identified from the CM are specific to the cell line being cultured as there are no other contaminating proteins, therefore, greatly simplifying MS analysis. Studies analyzing the CM from prostate(97), colon(352), endothelial(353), adipose(354), nasopharyngeal(355) and retinal epithelial cells(356), have already been conducted, thus demonstrating the versatility of this approach.

Previously, we have shown the PC3(AR)6 CaP cell line can be grown in a serum-free environment and the secreted proteins present can be readily identified by MS-based methods(98). In this study, we performed a detailed proteomic analysis of the CM of three CaP cell lines; PC3, LNCaP and 22Rv1. From this analysis we identified 2,124 proteins by using a bottom-up approach, consisting of offline strong cation exchange (SCX) chromatography followed by capillary C-18 reversed-phase liquid chromatography-tandem mass spectrometry (LC-MS/MS). Our extensive lists of proteins, and their cellular and biological classifications, may form the basis for discovering novel CaP biomarkers.

The utility of this approach was examined by validating four candidates by ELISA. Each was found elevated in serum of a subset of CaP patients; a positive correlation was also observed with serum PSA levels. Furthermore, our approach identified many known CaP biomarkers including PSA, KLK2, KLK11, prostatic acid phosphatase and prostate specific membrane antigen, further supporting the view that this unbiased approach may aid in new CaP biomarker discovery efforts.
4.2 Materials and Methods

4.2.1 Cell Culture

The PC3, LNCaP and 22Rv1 cell lines were purchased from the American Type Culture Collection (Rockville, MD). All cell lines were grown in T-175 culture flasks (Nunc) in RPMI 1640 culture medium (Gibco) supplemented with 8% fetal bovine serum (FBS) (Hyclone). Cells were cultured in a humidified incubator at 37°C and 5% CO₂. Cells were seeded at 7.5 X 10⁶, 22 X 10⁶, and 75 X 10⁶ cells for PC3, LNCaP and 22Rv1 cell lines, respectively. PC3 and 22Rv1 cells were grown for 2 days in 30mL of RPMI + 8% FBS. Afterwards, the medium was removed, and the flask was gently washed 3 times with 30mL of PBS (137 mM NaCl, 10 mM phosphate, 2.7 mM KCl, pH 7.4). Thirty milliliters of CDCHO (Gibco) medium supplemented with glutamine (8mmol/L) (Gibco) were added to the flasks and incubated for 2 days. The LNCaP cell line was grown as above, except that the cells were incubated for 3 days in RPMI + 8% FBS before the media were changed to CDCHO. All cell lines were grown in triplicate and independently processed and analyzed. A negative control was also prepared with the same procedures as above, except no cells were seeded.

After incubation in CDCHO, the CM was collected and spun down (3,000 X g) to remove cellular debris. Aliquots were taken for measurement of total protein, lactate dehydrogenase (LDH) (a marker of cell death), and kallikreins 3, 5, 6 (internal control proteins). The remainder was frozen at -80°C until further use.
4.2.2 Measurement of Total Protein, Lactate Dehydrogenase and PSA, KLK5, KLK6

The total protein of the CM was measured using the Coomassie (Bradford) assay (Pierce Biotechnology) as recommended by the manufacturer.

Lactate dehydrogenase levels in the CM were measured via an enzymatic assay based on conversion of lactate to pyruvate. NADH production from NAD$^+$ during this reaction was monitored at 340nm with an automated method and converted to Units per litre (U/L) (Roche Modular Systems).

Kallikrein 3 (PSA), KLK5 and KLK6 were measured with in-house ELISAs as described earlier(192,315,331).

4.2.3 Conditioned Media Sample Preparation and Trypsin Digestion

Our general protocol is shown in Figure 4.1. Approximately 30mL of CM from each cell line, which corresponded to 1mg of total protein, was dialyzed overnight using a 3.5kDa cut-off dialysis tubing (Spectra/Por) at 4°C in 5L of 1mM of NH$_4$HCO$_3$ solution with one buffer exchange. The dialyzed CM was lyophilized overnight to dryness followed by resolubilization with 322µL of 8M urea, 25µL of 200mM DTT, and 25µL of 1M NH$_4$HCO$_3$. The sample was vortexed thoroughly and incubated at 50°C for 30min. 125µL of 500mM iodoacetamide was added, the sample was incubated in the dark at room temperature for 1h. The sample was desalted with a NAP-5 column (GE Healthcare), lyophilized and resuspended in 120µL of 50mM NH$_4$HCO$_3$, 100µL methanol, 150µL H$_2$O and 5µg of trypsin (Promega), vortexed, and incubated for 12h at 37°C.
Chapter 4: Comparative Proteomic Analysis of the Conditioned Media of Three Prostate Cancer Cell Lines

Figure 4.1

Prostate Cancer Cell Lines
PC3, LNCaP, 22Rv1

Prepare Conditioned Media

Trypsin Digestion

SCX - HPLC Chromatography

12 Fractions

LC-MS/MS

Mascot and X!Tandem Database Searching

Identification Probability Scoring by Scaffold

GO Annotation
Replicate Overlap
Cell Line Overlap

Biomarker Candidates

Pre-Clinical Validation

Figure 4.1: Workflow of proteomic method employed. For additional information, refer to Materials and Methods section.
Chapter 4: Comparative Proteomic Analysis of the Conditioned Media of Three Prostate Cancer Cell Lines

4.2.4 Strong Cation Exchange High Performance Liquid Chromatography

The tryptic digests were lyophilized and resuspended in 120µL of 0.26M formic acid in 10% ACN (mobile phase A). The sample was fractionated using an Agilent 1100 HPLC system connected to a PolySULFOETHYL A™ column containing a hydrophilic, anionic polymer (poly-2-sulfoethyl aspartamide) with a 200Å pore size and a diameter of 5µm (The Nest Group Inc.). A one hour linear gradient was used, with 1M ammonium formate and 0.26M formic acid in 10% acetonitrile (mobile phase B) at a flow rate of 200µL/min. Fractions were collected via a fraction collector every 5 min (12 fractions per run), and frozen at -80°C for further use. A protein cation exchange standard, consisting of four peptides, was run at the beginning of each day to assess column performance (American Peptide).

4.2.5 Online Reversed Phase Liquid Chromatography – Tandem Mass Spectrometry

Each 1mL fraction was C18 extracted using a Zip TiP_{C18} pipette tip (Millipore) and eluted in 4µL of 90% ACN, 0.1% formic acid, 10% water, 0.02% trifluoroacetic acid (TFA) (Buffer B). Upon which 80µL of 95% water, 0.1% formic acid, 5% ACN, 0.02% TFA (Buffer A) was added and half of this volume (40µL) was injected via an auto-sampler on an Agilent 1100 HPLC. The peptides were first collected onto a 2 cm C18 trap column (inner diameter 200 µm), then eluted onto a resolving 5 cm analytical C18 column (inner diameter 75 µm) with an 8 micron tip (New Objective). The HPLC was coupled online to an LTQ 2-D Linear Ion Trap (Thermo Inc.). A 120 min gradient was used on the HPLC and peptides were ionized via nano-
electrospray ionization. The peptides were subjected to MS/MS and DTAs were created using the Mascot Daemon v2.16 and extract_msn (Matrix Science). Parameters for DTA creation were: min mass 300, max mass 4000, automatic precursor charge selection, min peaks 10 per MS/MS scan for acquisition and min scans per group of 1.

4.2.6 Database Searching and Bioinformatics

Mascot, v2.1.03 (Matrix Science)(357) and X!Tandem v2.0.0.4 (GPM, Beavis Informatics Ltd.)(358) database search engines were used to search the spectra from the LTQ runs. Each fraction was searched separately against both search engines using the IPI Human database V3.16(359) with trypsin specified as the digestion enzyme. One missed cleavage was allowed, a variable oxidation of methionine residues and a fixed modification of carbamidomethylation of cysteines was set with a fragment tolerance of 0.4 Da and a parent tolerance of 3.0 Da.

The resulting DAT files from Mascot and XML files from X!Tandem were inputed into Scaffold v01_05_19 (Proteome Software)(360) and searched by allocating all DAT files into one biological sample and all XML files into another biological sample. This was repeated three times for each cell line. The cut-offs in Scaffold were set for 95% peptide identification probability and 80% protein identification probability. Identifications not meeting these criteria were not included in the displayed results. The sample reports were exported to Excel, and an in-house developed program was used to extract Gene Ontology (GO) terms for cellular component for each protein and the proportion of each GO term in the
dataset. Proteins that were not able to be classified by GO terms were checked with Swiss-Prot entries and against the Human Protein Reference Database (361) and Bioinformatics Harvester (http://harvester.embl.de) to search for cellular component annotations. Finally, the overlap between proteins identified from each cell line and within each of the replicates per cell line was determined by in-house built software. Each protein was also searched against the Plasma Proteome Database (www.plasmaproteomedatabase.org). The list of proteins were also compared with those found in seminal plasma by Pilch et al. (362) and in breast cancer cell line CM by Kulasingam et al. (363). In addition, we used Ingenuity Pathway Analysis software (Ingenuity Systems) to determine differences in biological networks in the extracellular and membrane proteins of each cell line, as well as overlay molecular functions with respect to disease conditions associated with each of the biomarker candidates.

The same set of spectra produced by the LTQ was searched with the same parameters as above, but against a reversed IPI Human database V3.16 which was created using the Reverse.pl script from The Wild Cat Toolbox (Arizona Proteomics). The DAT and XML files from this “reversed” search were input into Scaffold as before and the identified peptides meeting the pre-set cut-offs were identified. The false positive rate (FPR) was calculated as such: $\text{FPR} = \frac{\# \text{False peptides}}{\# \text{True peptides} + \# \text{False peptides}}$. 

4.2.7 Validation of Candidates

A pre-clinical validation was conducted where the concentrations of four candidate proteins (follistatin, chemokine (C-X-C motif) ligand 16, pentraxin 3 and spondin 2) were measured by ELISA in the serum of patients with or without prostate cancer and in the CM of each cell line. These candidates were selected for validation using similar criteria as Kulasingam et al. Follistatin (R&D Systems), chemokine (C-X-C motif) ligand 16 (R&D Systems) and pentraxin 3 (Alexis Biochemicals) were measured using their respective manufacturer’s protocol. Spondin 2 was measured at diaDexus Corporation, San Francisco, CA using an in-house developed assay. Serum samples from healthy males and CaP patients were collected at the Toronto Medical Laboratories, Toronto, Canada. The patient groups were classified by their PSA levels. Healthy males were patients with PSA levels <1ug/L and CaP patients were those with PSA levels >10ug/L. The study was carried out after Institutional Review Board approval, median age of patients were 75 and 67 for the cancer and non-cancer groups, respectively.
Chapter 4: Comparative Proteomic Analysis of the Conditioned Media of Three Prostate Cancer Cell Lines

4.3 Results

4.3.1 Proteins Identified by Mass Spectrometry

A schematic of the sample preparation and bioinformatics utilized in this study is shown in Figure 4.1. Each cell line was independently cultured in triplicate to determine the reproducibility of our method in identifying proteins between the replicates. After setting a cut-off of 95% peptide probability and 80% protein probability in Scaffold, 2,124 proteins were identified that met the criteria from all three cell lines combined. In total, from the three replicates per cell line, 1,157, 1,285, and 1,116 proteins were found in the PC3, LNCaP and 22Rv1 cell lines, respectively.

A control flask was also prepared that did not contain any cells but were treated with the same procedure. A total of 69 proteins were identified in the negative control flask, which represented FCS-derived proteins from incomplete washing of the tissue culture flasks. These proteins were removed from the list of identified CM proteins of each cell line and were not considered further in the data analysis.

Furthermore, to empirically determine false positive rate (FPR) of peptide detection, the dataset was searched against a reversed IPI Human v3.16 database using the same search parameters and Scaffold cut-offs. A total of 4, 4, and 6 proteins from the PC3, LNCaP and 22Rv1 cell lines, respectively were observed. Each protein was identified by one peptide with Mascot and with no peptides identified by X!Tandem (0% FPR). A FPR of <1% for all 3 cell lines was calculated by Mascot.
4.3.2 Identification of Internal Control Proteins

As an internal control, we sought to identify by LC-MS/MS three proteins that were known to be secreted by these cell lines and were monitored by ELISA during our cell culture optimization. The approximate initial concentrations of these proteins in CM are given in brackets, below. We confidently identified PSA in the CM of LNCaP (~550 µg/L) and 22Rv1 (~3 µg/L) with several peptides (Table 4.1). PC3 cells do not secrete any detectable PSA by ELISA and, as expected, this protein was not identified in its CM. One KLK6 peptide from the CM of PC3 (~1.5 µg/L) was identified, but no peptides from KLK5 were identified by MS. Together, these data suggest that the detection limit of our MS-based method for protein identification is in the low µg/L range.

4.3.3 Reproducibility between Replicates

Next, we investigated the reproducibility of our method by culturing each cell line in triplicate. The Mascot and X!Tandem results from the PC3, LNCaP and 22Rv1 cell lines are shown in Figure 4.2. In general, a 56% overlap of identified proteins in all three replicates from both Mascot and X!Tandem for each of the cell lines was observed. Approximately 20% of proteins were found in two replicates and 24% were exclusive to one replicate. This data highlights the ionization efficiency of the mass spectrometer, and thus the need for replicate analysis of samples by MS to obtain a more comprehensive list of the secretome of these cell lines.
Table 4.1: Known prostate biomarkers identified in the conditioned media of PC3, LNCaP, and 22Rv1 cell lines

<table>
<thead>
<tr>
<th>Protein</th>
<th>PC3 Mascot</th>
<th>PC3 GPM</th>
<th>LNCaP Mascot</th>
<th>LNCaP GPM</th>
<th>22Rv1 Mascot</th>
<th>22Rv1 GPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>hK2</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSA</td>
<td>9</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hK11</td>
<td></td>
<td></td>
<td>7</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACPP</td>
<td>7</td>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mac-2BP</td>
<td>15</td>
<td>13</td>
<td>7</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zn-α2-GP</td>
<td>7</td>
<td>10</td>
<td>15</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSMA</td>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NGEP</td>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Protein name and number of peptides identified by Mascot and X!Tandem for each cell line are listed. KLK2 – Human kallikrein 2, PSA – Prostate specific antigen, KLK11 – Human kallikrein 11, ACPP – Prostatic acid phosphatase, Mac-2BP – Mac 2 binding protein, Zn-α2-GP – Zinc alpha 2 glycoprotein, PSMA – Prostate specific membrane antigen, NGEP – New gene expressed in prostate.
Figure 4.2: Overlap of the 3 replicates from PC3, LNCaP and 22Rv1 conditioned media: Number of proteins identified from each cell culture replicate that overlapped with other replicates for both Mascot and X!Tandem results are depicted. Each circle represents a replicate.
4.3.4 Differences in Proteins Identified Between Cell Lines

Following this, the proteins identified from each of the three replicates of each cell line were combined to form a non-redundant list per cell line. These three lists were compared to determine their overlap. The results are shown in Figure 4.3. About 54% of proteins were unique to one of the cell lines, 21% were common in all 3 cell lines, with 24% were identified in two of the cell lines. This data highlights the heterogeneity of CaP cell lines and the need to investigate multiple cell lines, to obtain a comprehensive picture of the CaP proteome. The overlap among the cell lines for extracellular and membrane proteins, rather than total proteins yielded similar heterogeneity (Figure 4.4).

4.3.5 Genome Ontology Distributions of Proteins

As shown in Figure 4.5, 12% of the proteins identified were classified as extracellular, 18% as membrane and 12% as unclassified. The remainder of the proteins identified in the CM were classified as intracellular, nucleus, golgi, endoplasmic reticulum (ER), endosome or mitochondria (58%). Classification by cellular localization is redundant since a protein can be classified in more than one compartment. A similar distribution was found with the cellular component distribution of each cell line (Figure 4.6).
Figure 4.3: Overlap of proteins identified between each cell line: Each circle represents a cell line.
Chapter 4: Comparative Proteomic Analysis of the Conditioned Media of Three Prostate Cancer Cell Lines

Figure 4.4: Overlap of the extracellular and membrane proteins identified in each cell line: Extracellular and membrane proteins respectively were compared from each cell line. The numbers of proteins common and unique to each cell line are depicted.
**Figure 4.5**

Classification of proteins by cellular location: Each protein identified after MASCOT and X!Tandem searching was classified by its cellular location using Genome Ontology classifiers (www.geneontology.org).

- Intracellular, 821, 29%
- Nucleus, 458, 16%
- Membrane, 507, 18%
- Endosome, 103, 4%
- ER, 108, 4%
- Golgi, 57, 2%
- Mitochondria, 82, 3%
- Extracellular, 330, 12%
- Unclassified, 350, 12%
Figure 4.6: Classification of proteins by cellular location for each cell line: Each protein identified after MASCOT and X!Tandem searching was classified by its cellular location using Genome Ontology classifiers (www.geneontology.org). Distribution of proteins identified in the conditioned media of (A) PC3, (B) LNCaP, and (C) 22Rv1 cell lines.
4.3.6 Secreted and Membrane Proteins

After GO annotation, 329 proteins were classified as extracellular, 504 as membranous and 339 as unclassified. From the list of extracellular and membrane proteins several previously known CaP biomarkers were identified. These included PSA(364), KLK2(337,338), KLK11(365), PAP(366), Mac-2BP(367), zinc-α-2-glycoprotein(368), a new gene expressed in the prostate (NGEP)(369) and PSMA(370) (Table 4.1). Comparing all extracellular, membrane and unclassified proteins with the Plasma Proteome Database yielded 93 membrane proteins, 98 extracellular proteins, and 27 unclassified proteins as being found previously in serum.

4.3.7 Overlap with Previous Data

Given that we had previously analyzed the CM of the PC3(AR)6 cell line(98), the overlap between this set of data and our previous one was examined. In brief, the current data yielded 5-fold more proteins owing in part to the improvements in sample preparation, the type of mass spectrometer used (more sensitive) and the powerful bioinformatics (higher confidence of protein probability).

4.3.8 Overlap with Seminal Plasma Proteins

Recently, Pilch and Mann(362) conducted an in-depth proteomic survey of the seminal plasma proteome and identified 932 unique proteins. Comparing our list of extracellular, membrane and unclassified proteins found in the CM of prostate cancer cell lines with those reported in a biological fluid such as seminal
plasma (362) resulted in 108 extracellular, 120 membranous, and 40 unclassified proteins as being common.

### 4.3.9 Biological Network Analysis

The extracellular and membrane proteins from each cell line were analyzed using Ingenuity Pathway Analysis software v5.51 (Ingenuity Systems) and functional networks were developed. Cancer, cellular movement and pathways involved in tissue development were listed among the top networks for each of the cell lines. However, in the case of 22Rv1 these were listed secondary to the “housekeeping” functional networks. Next, the functions and disease association of each of the four candidates selected were examined. Follistatin displayed the most direct connections to cancer such as prostate (371,372), colon (373), and ovarian cancer (374) (Figure 4.7). Chemokine (C-X-C motif) ligand 16 (CXCL16) displayed connections to inflammation and chemotaxis (375) (Figure 4.8A). Pentraxin 3 (PTX3) showed multiple connections to tissue and embryonic development (376) (Figure 4.8B). Few studies have been conducted on spondin 2 (SPON2); it was shown to be involved in neuronal guidance and lung cancer (377) (Figure 4.8C).
Figure 4.7: Molecular functions related to diseases associated with Follistatin: Web diagram depicting the biological functions that follistatin is associated with, in the context of disease. Diagram generated through Ingenuity Pathway Analysis software (Ingenuity Systems).
Chapter 4: Comparative Proteomic Analysis of the Conditioned Media of Three Prostate Cancer Cell Lines

Figure 4.8: Molecular functions related to diseases associated with Chemokine (C-X-C motif) ligand 16, Pentraxin 3 and Spondin 2: Web diagram depicting biological functions associated with in the context of disease for CXCL16 (A), PTX3 (B) and SPON2 (C). Diagram generated through Ingenuity Pathway Analysis software (Ingenuity Systems).
Chapter 4: Comparative Proteomic Analysis of the Conditioned Media of Three Prostate Cancer Cell Lines

4.3.10 Overlap with Breast Cancer Secretome

The proteins identified by Kulasingam et al. (363) in the CM of two breast cancer cell lines were compared with those identified in our analysis of CaP CM. There was an overlap of 256 proteins that were present in the BT474, MDA468, PC3, LNCaP and 22Rv1 cell line CM. These proteins were then subjected to biological network analysis in a similar manner as stated above. Pathways related to cancer and cell growth were among the top identified.

4.3.11 Validation of Follistatin, Chemokine (C-X-C motif) ligand 16, Pentraxin 3 and Spondin 2

From the list of proteins identified, pre-clinical validation of four candidates were performed in the CM of PC3, LNCaP and 22Rv1 cell lines (Figure 4.9) and in 42 serum samples from patients with or without CaP (Figure 4.10). The concentration of each candidate in the CM correlates semi-quantitatively with the number of unique spectra (shown in brackets) identified from each peptide of each candidate after database searching by Mascot/X!Tandem: Follistatin (PC3, 100/128; 22Rv1, 0/1); CXCL16 (LNCaP, 3/4, 22Rv1, 2/2); PTX3 (PC3, 45/52); SPON2 (LNCaP, 152/149, 22Rv1, 2/2). A significant difference (Mann-Whitney Test) in sera of patients with or without CaP in all four candidates (Figure 4.9A, C, E, G) by ELISA was observed. In addition, the correlation between PSA levels and candidate levels in serum of patients with CaP was significant and positive by Spearman analysis (Fig 4.9B, D, F, H).
Figure 4.9: Concentrations of each candidate in the conditioned media of each cell line and the control flask. Levels of Follistatin (A), CXCL16 (B), Pentraxin 3 (C), Spondin 2 (D) in conditioned media is shown with corresponding number of spectra identified for each candidate by Mascot/X!Tandem.
Figure 4.10: Validation of Follistatin, Chemokine (C-X-C motif) ligand 16, Pentraxin 3 and Spondin 2 in serum: Levels of Follistatin, CXCL16, PTX3 and SPON2 measured in serum of patients with or without CaP (A, C, E, G respectively). Median values are shown by a horizontal line. p values were calculated using a Kruskal-Wallis test. Correlations of each candidate with PSA levels (B, D, F, H), p values calculated by the Spearman correlation. (r = Spearman correlation coefficient)
4.4 Discussion

LC-MS/MS analysis allows for the elucidation of the identity of thousands of proteins in complex mixtures, in a high-throughput fashion. This technology has been applied to cancer biomarker discovery, where biological fluids, tissues or cell cultures have been analyzed for differences in protein expression(97,352,378-380). However, the dynamic range of current LC-MS/MS methods is not adequate to identify all proteins in a complex mixture such as serum. Even with depletion of highly abundant proteins and extensive fractionation, this is a major challenge still faced today(296).

Prostate cancer, when diagnosed early, is associated with favourable clinical outcomes(381). Our objective was to identify proteins secreted by three tumourigenic CaP cell lines of differing origin and phenotype. The purpose of examining the secretome of cell lines of differing origin was to obtain a more complete picture of the proteome of CaP since CaP is a heterogeneous disease(382,383) and it requires a diverse model system for biomarker discovery. Thus, we chose to focus on shed and secreted proteins from three CaP cell lines, since this approach is amenable to MS and these proteins will most likely be produced by the tumour in measurable amounts to be detected via a blood test.

In our previous study of the CM of the PC3(AR)_6 cell line(98) we found a large number of intracellular proteins. In another study, similar data were seen with the LNCaP cell line (97). We sought to reduce the amount of intracellular proteins by optimizing the cell culture (Figures 3.1, 3.2, 3.3). Yet our current data (Figure 4.5) revealed a similar distribution of proteins by cellular component. From this, we
deduce that cell death during culture is an unavoidable contaminant when analyzing CM by proteomics. However, previous studies from our laboratory utilizing a similar cell culture-based approach found that the proteins identified in cell lysates do not contain as many extracellular proteins as the CM for that cell line(363). Furthermore, the extracellular proteins identified in the cell lysate displayed minimal overlap to the proteins identified in the CM illustrating that analyzing the CM, despite the amount of cell death occurring in SFM, leads to a significant enrichment of secreted proteins which may be novel serological markers. We also believe that our method may include microsomal proteins that were not removed during centrifugation of the CM after harvesting.

From our previous work we determined that replicate analysis expanded the coverage and increased the number of identified proteins(98). As can be seen in Figure 4.2, there are proteins that are uniquely identified by only one replicate. This is most likely due to the incomplete ionization of certain peptides during an LC-MS/MS run. Many studies have shown that cell lines do represent the tumour from which they originated. Hence, the proteins that they secrete should reflect the genetic alterations that they harbor. Given that the biological triplicates yielded a more complete coverage of the secretome for a cell line, it is highly probable that the differences in the proteins identified among the 3 cell lines (Figure 4.3) indeed reflects the heterogeneity of that cell line.

We confirmed the presence of two internal control proteins (PSA and KLK6) in the CM at µg/L concentrations as measured by ELISA and identified by MS (Figures 3.1, 3.2, 3.3). We were not able to identify by MS KLK5 which was also
present in PC3 at about the same concentration as KLK6 (2 µg/L). It is possible that
the stringent peptide and protein probability cut-offs utilized in this study was too
strict to allow identification of KLK5 by MS. Alternatively, it is possible that a
concentration of 2 µg/L is close to the detection limit of our methodology and hence
it was not identified. We thus conclude that our method can identify proteins in CM of
approx. low µg/L or higher, a detection limit which nevertheless is 2-3 orders of
magnitude better than the ones achieved by using serum (384). Furthermore, based
on the currently used biomarkers in the clinic, this is the expected concentration
range that potential tumour markers should be observed in serum.

The use of multiple search engines has been shown to increase confidence,
as well as expand coverage (385). In this study, Mascot and X!Tandem were used
since these search engines use different algorithms to determine if a mass spectrum
matches an entry in the database. The use of both search engines served to provide
an independent confirmation of the results. The use of peptide (386) and protein
prophet algorithms (387) contained within Scaffold allowed for increased confidence
of the protein identification probabilities. To further increase confidence, we
performed a search against a reversed IPI human database (388) and obtained FPR
of <1% for the cell lines by Mascot. The low FPR highlights the fidelity of the search
approach used. In addition, to eliminate contaminants left over from the FCS we
processed a control flask that did not contain any cells and deleted from the list of
proteins identified from the CaP cell lines the FCS-derived proteins.

Moreover, from the extracellular, membrane and unclassified proteins
identified in this study, 98, 93, and 26 proteins in extracellular, membrane and
unclassified protein lists, respectively were found in the plasma proteome. It is possible that the other proteins are of low abundance and have not yet been identified in circulation. This list was also compared against the list of proteins found in seminal plasma(362). We reasoned that if a protein identified from our CaP cell lines is also detected in seminal plasma, it is likely to be secreted or shed at relatively high concentration by prostate cells. We found 108 extracellular, 120 membranous and 40 unclassified proteins in our CM, as well as in the seminal plasma proteome.

To discern functional differences between the “secretome” of the three cell lines, we performed a biological analysis of the extracellular and membrane proteins identified from each cell line’s CM. First, we compared the extracellular and membrane proteins identified between each cell line (Figure 4.4). We found 18% and 20% of proteins identified as extracellular and membrane, respectively, were common to all three cell lines. This is a significantly lower overlap than the overall number of proteins which were common between the cell lines, indicating that there are significant differences specifically within the “secretome” from each cell line. We further investigated functional differences between the cell lines using the Ingenuity Pathway Analysis. The top ranked networks were those involved with cell movement, signalling, cancer and the cell cycle. Furthermore, 22Rv1 was not represented as having any extracellular proteins involved with cellular movement, whereas PC3 and LNCaP did. This is interesting and relevant since 22Rv1 was derived from an organ-confined prostate tumour while PC3 and LNCaP were derived from metastatic tumours. In addition, each cell line differs with its sensitivity to
androgens, a major component to prostate development: PC3 does not express the AR and is a model for androgen insensitive tumours, 22Rv1 expresses the AR but is considered AR insensitive, and LNCaP, which expresses AR, is considered androgen dependent. In addition, we also investigated the common proteins in breast CM identified by Kulasingam et al.(363) with those identified in this study to look for similarities between these two hormonally regulated cancers. Here we see that cancer and cell proliferation networks are among the top pathways identified to be common. This commonality between the cell lines highlights that these tumours utilize similar processes during carcinogenesis and warrants further study to delineate biomarkers that could be useful for diagnosis.

Using information from this study, we developed criteria to narrow down our list of candidate biomarkers: (a) we considered proteins that showed relatively prostate-specific mRNA expression by searching through the UniGene expressed sequence tag online database; (b) we selected proteins that were classified as extracellular or membrane and have been identified in seminal plasma(362); (c) we performed literature searches to ensure that these proteins have not been validated before as CaP biomarkers and highlighted proteins that have shown biological links to CaP and other cancers; (d) we then selected candidates that had commercially available immunoassays. The list of narrowed down candidates are listed in Table 4.2.

From these, follistatin, CXCL16, PTX3 and SPON2 were chosen for further validation based on the above criteria. Each candidate showed several links to CaP(371,389-391) or carcinogenesis(377,392-397). Each candidate showed a
significant difference between patients with or without CaP. In addition, a positive correlation with increasing PSA was also observed (Figure 4.10 B, D, F, H). From these results, we speculate that these candidates show an association with CaP progression. Future studies will determine if in combination they can improve the specificity of the PSA test.

To determine the biological association with CaP we profiled each of the candidate’s links to functions and diseases (Figure 4.7, 4.8). With the exception of SPON2, each of the candidates displayed several links to cancer development, tissue development, inflammation or chemotaxis. All of these processes have shown to play a role in the malignant development of tumours. However, the involvement of each candidate with respect to CaP pathobiology will need to be further studied to elucidate their role during progression.

In summary, we present a robust method of proteomic analysis of cell culture CM and bioinformatics for new biomarker discovery. The four candidates validated have not been previously shown to be serum markers for CaP but require further study to fully elucidate their roles in CaP progression. Additional candidates from this large database are worth validating in the future.
### Table 4.2: List of candidate biomarkers selected based on selection criteria.

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>64 kDa protein</td>
<td>C1orf116</td>
</tr>
<tr>
<td>Abhydrolase domain containing 14A</td>
<td>ABHD14A</td>
</tr>
<tr>
<td>ADAMTS-1 precursor</td>
<td>ADAMTS1</td>
</tr>
<tr>
<td>Cartilage intermediate layer protein 2 precursor</td>
<td>CILP2</td>
</tr>
<tr>
<td>CEGP1 protein</td>
<td>SCUBE2</td>
</tr>
<tr>
<td>c-Myc-responsive protein Rcl</td>
<td>C6orf108</td>
</tr>
<tr>
<td>Creatine kinase B-type</td>
<td>CKB</td>
</tr>
<tr>
<td>chemokine (C-X-C motif) ligand 16</td>
<td>CXCL16</td>
</tr>
<tr>
<td>Epiplakin</td>
<td>EPPK1</td>
</tr>
<tr>
<td>Epoxide hydrolase 2</td>
<td>EPHX2</td>
</tr>
<tr>
<td>GAGE-2 protein</td>
<td>GAGE2</td>
</tr>
<tr>
<td>Galanin precursor</td>
<td>GAL</td>
</tr>
<tr>
<td>glucosamingly (N-acetyl) transferase 2, I-branching enzyme isoform A</td>
<td>GCNT2</td>
</tr>
<tr>
<td>Glycoprotein endo-alpha-1,2-mannosidase (Fragment)</td>
<td>MANEA</td>
</tr>
<tr>
<td>Group 3 secretory phospholipase A2 precursor</td>
<td>PLA2G3</td>
</tr>
<tr>
<td>Growth-regulated protein alpha precursor</td>
<td>CXCL1</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td>RPL22L1</td>
</tr>
<tr>
<td>Hypothetical protein C1orf80</td>
<td>C1orf80</td>
</tr>
<tr>
<td>Secretogranin-1 precursor</td>
<td>CHGB</td>
</tr>
<tr>
<td>Secretogranin-3 precursor</td>
<td>SCG3</td>
</tr>
<tr>
<td>Semaphorin 3F variant (Fragment)</td>
<td>SEMA3F</td>
</tr>
<tr>
<td>Splice Isoform 1 of Follistatin precursor</td>
<td>FST</td>
</tr>
<tr>
<td>Splice Isoform 1 of Heme-binding protein 2</td>
<td>HEBP2</td>
</tr>
<tr>
<td>Splice Isoform PTPS of Receptor-type tyrosine-protein phosphatase S precursor</td>
<td>PTPRS</td>
</tr>
<tr>
<td>Spondin-2 precursor</td>
<td>SPON2</td>
</tr>
<tr>
<td>von Willebrand factor A domain-related protein isoform 1</td>
<td>VWA1</td>
</tr>
<tr>
<td>glucosamingly (N-acetyl) transferase 2, I-branching enzyme isoform A</td>
<td>GCNT2</td>
</tr>
<tr>
<td>Splice Isoform 1 of Zinc transporter SLC39A6 precursor</td>
<td>SLC39A6</td>
</tr>
<tr>
<td>Neurexin-3-alpha precursor</td>
<td>NRXN3</td>
</tr>
<tr>
<td>Tumor-associated calcium signal transducer 2 precursor</td>
<td>TACSTD2</td>
</tr>
<tr>
<td>protein tyrosine phosphatase, receptor type, sigma isoform 3 precursor</td>
<td>PTPRS</td>
</tr>
<tr>
<td>latent transforming growth factor beta binding protein 1 isoform LTBP-1L</td>
<td>LTBP1</td>
</tr>
<tr>
<td>Pentraxin-related protein PTX3 precursor</td>
<td>PTX3</td>
</tr>
<tr>
<td>Splice Isoform 1 of Kallikrein-15 precursor</td>
<td>KLK15</td>
</tr>
<tr>
<td>Interferon-induced 17 kDa protein precursor</td>
<td>ISG15</td>
</tr>
<tr>
<td>Lysyl oxidase homolog 1 precursor</td>
<td>LOXL1</td>
</tr>
<tr>
<td>Spondin-2 precursor</td>
<td>SPON2</td>
</tr>
<tr>
<td>64 kDa protein</td>
<td>C1orf116</td>
</tr>
<tr>
<td>CEGP1 protein</td>
<td>SCUBE2</td>
</tr>
<tr>
<td>ATP-binding cassette, sub-family C (CFTR/MRP), member 10</td>
<td>ABCC10</td>
</tr>
<tr>
<td>Hypothetical protein LOC255101 isoform 1</td>
<td>CCDC108</td>
</tr>
<tr>
<td>PREDICTED: ephrin receptor EphA6</td>
<td>EPHA6</td>
</tr>
<tr>
<td>Similar to Hepatocyte nuclear factor 1a dimerization cofactor isoform</td>
<td>PCBD2</td>
</tr>
<tr>
<td>von Willebrand factor A domain-related protein isoform 1</td>
<td>VWA1</td>
</tr>
<tr>
<td>Tetraspanin-1</td>
<td>TSPAN1</td>
</tr>
<tr>
<td>NGEP long variant</td>
<td>TEMEM16G</td>
</tr>
<tr>
<td>Sorbitol dehydrogenase</td>
<td>SORD</td>
</tr>
</tbody>
</table>
CHAPTER 5:

VALIDATION OF CANDIDATE BIOMARKERS IN PROSTATE CANCER PATIENT SAMPLES
Chapter 5: Validation of Candidate Biomarkers in Prostate Cancer Patient Samples

5.1 Introduction

In addition to the four candidates validated in the previous chapter there were four additional candidates that were selected for further validation. Based on the criteria outlined in the previous chapter for the selection of candidates identified in the CM of the CaP cell lines, a preliminary validation was performed on chemokine (C-X-C motif) ligand 5 (CXCL5), lipocalin-2 (LCN2), chemokine (C-X-C motif) ligand 1 (CXCL1) and tumor necrosis factor receptor superfamily, member 6b, decoy (TNFRSF6B). However, the preliminary results did not qualify these candidates for further study. Their association to cancer and pathological conditions are summarized below.

Chemokine (C-X-C motif) ligand 5 also known as epithelial neutrophil-activating protein 78 is a member of the CXC chemokine family and has been shown to act as an inflammatory chemoattractant and activator of neutrophil function. It has been seen to be elevated in serum of patients with rheumatoid arthritis, inflammatory bowel conditions such as Crohn’s disease and ulcerative colitis. High expression of CXCL5 was also seen in tissues of patients with chronic pancreatitis(398). Unlike other chemokines, expression of CXCL5 is not limited to immune cells and is seen in epithelial, endothelial and fibroblastic cell types. Expression of CXCL5 has been associated with both CaP and gastrointestinal malignancies and related inflammatory conditions(399-402).

Lipocalin 2 also known as neutrophil gelatinase-associated lipocalin is a 25 kDa secreted glycoprotein that is part of the lipocalin family of proteins and has been shown to function in cell proliferation and survival. An immunohistochemical study
has shown the expression of LCN2 in a variety of tissues including tumour tissues of the prostate, lung, colon, pancreas and kidney(403). In addition, LCN2 has been studied as a marker for kidney injury(404) and ovarian cancer(405) and has shown to be a poor prognostic indicator in breast cancer(406). The role of LCN2 in carcinoma progression has also been studied and has shown to be promote breast cancer and esophageal squamous cell carcinoma invasion by stabilizing extracellular matrix enzymes(407,408) and inhibit cell invasion in colon cancer cells(409).

Chemokine (C-X-C motif) ligand 1 also known as growth-related oncogene-α like CXCL5 is part of the CXC family of chemokines. As with CXCL5, CXCL1 is also elevated in serum of patients with inflammatory bowel conditions(410). Studies in several different cancers have revealed that CXCL1 is a mediator of angiogenic and tumourigenic activity in CaP cells(400,411) as well as has shown to promote the malignant phenotype in colon cancer(412).

Tumor necrosis factor receptor superfamily, member 6b, decoy, also known as decoy receptor 3, is a soluble receptor that is part of the tumour necrosis factor receptor family. This soluble receptor has been shown to be overexpressed in some cancers and thought to compete with binding Fas-L death receptors, thus inhibiting apoptosis(413). Amplification of the TNFRSF6B gene has been shown in lung and colon cancers(414), lymphomas(415), gastric cancer(416), glioblastoma(417) and hepatocellular carcinoma(418). Serum levels of the soluble decoy receptor have been measured in several malignancies and been shown to have diagnostic and prognostic ability(397,419,420).
An extended validation of follistatin was also performed in patient serum and tissue samples. From all the candidates previously validated, follistatin was determined to be the strongest. Here we measure its levels in serum of patients with biopsy confirmed CaP, BPH, PIN, inflammatory conditions and disease free. In addition, we show immunohistochemical staining of follistatin in prostate, colon and lung tumours.

5.2 Materials and Methods

Four candidates were validated in the biological fluids of patients with and without CaP as well as other disease conditions. In addition, an extended validation of follistatin was conducted in a larger patient cohort, as well as its levels in CaP and other cancer tissues were determined by immunohistochemistry.

5.2.1 Conditioned Media, Serum Samples and Tissue Specimens

Aliquots from the CM of the PC3, LNCaP, 22Rv1 and the control flask were used for measuring of CXCL5, LCN2, CXCL1 and TNFRSF6B. The cell lines were grown at their optimized seeding densities as stated in the previous chapter.

Serum samples used for measurement of CXCL5, LCN2, CXCL1 and TNFRSF6B were collected from patients at the Toronto Medical Laboratories, Toronto, Canada. The study was carried out after Institutional Review Board approval. Samples were screened based on PSA levels and levels above 10ng/mL were grouped in the CaP group, while PSA levels below 1ng/mL were considered normal. Serum samples used for measurement of follistatin, KLK2, PSA, fPSA and
KLK11 were collected from patients admitted to The Prostate Biopsy Clinic, in the Prostate Centre at Princess Margaret Hospital, Toronto, for prostate biopsy. Informed consent was obtained from each patient for participation in this study and blood samples were collected prior to biopsy in yellow top serum separator tubes. Whole blood was allowed to clot for 30 minutes at room temperature and subsequently aliquoted and stored at -80°C until needed. Patients were diagnosed with either CaP, BPH, PIN, inflammation or free of disease (Benign) based on biopsy results. Data regarding age, PSA level, Gleason score, number of biopsy cores taken, reason for biopsy, tumour location and if the patient had initial or recurrent CaP was recorded. The mean ages for each group were: Benign – 60, BPH – 63, CaP – 64, PIN – 60, Inflammation – 60. The majority of the patients in each group had PSA levels in the gray zone (4-10ug/L).

Paraffin embedded tissue samples of prostate, lung and colon carcinomas were obtained from the Toronto General Hospital, Toronto, Canada pathology laboratory after institutional review board approval.

5.2.2 Quantification of Candidates in Biological Fluids

Chemokine (C-X-C motif) ligand 5

The CXCL5 sandwich ELISA kit was purchased from R&D Systems (Cat#: DX000). The assay was performed using the kit protocol. Briefly, 200ul of assay diluent was added to each pre-coated well in a 96 well plate followed by 50ul of standard or sample in duplicate. The plate was incubated at room temperature for 2 hours followed by washing 3 times using an automated plate waster. Two hundred
microlitres of the secondary detection antibody was added to each well and the plate was incubated at room temperature for 2 hours. Two hundred microlitres of substrate solution was added to each well and the plate was covered with foil and left for 30 minutes at room temperature. Fifty microlitres of stop solution was then added to each well and the absorbance was read by a plate reader at 450nm wavelength with a correction of 570nm.

**Lipocalin-2**

The human lipocalin-2 sandwich ELISA kit was purchased from R&D Systems (Cat#:DLCN20) and was performed according to the kit protocol. The methodology is similar to CXCL5 with some minor differences.

**Chemokine (C-X-C motif) ligand 1**

The human CXCL1 ELISA kit was purchased from R&D systems (Cat#:DGR00) and was performed according to the kit protocol. The methodology is similar to that of CXCL5 and LCN2.

**Follistatin**

The human follistatin ELISA kit was purchased from R&D systems (Cat#:DFN00) and was performed according to the kit protocol. The methodology is similar to that of CXCL5, LCN2 and CXCL1.

**Tumor necrosis factor receptor superfamily, member 6b, decoy**

The human tumor necrosis factor receptor superfamily, member 6b, decoy (TNFRSF6B) ELISA kit was purchased from R&D systems (Cat#:DY142) and was performed according to the kit protocol with some minor modifications. Briefly, 200uL of the capture antibody was coated overnight at a concentration of 4ug/mL.
The antibody was aspirated off and 100uL of standard or sample was aliquoted in duplicate and incubated shaking at room temperature for 2 hours. The plate was then washed 6 times and the biotinylated detection antibody was added and incubated for 2 hours shaking at room temperature. The plate was washed 6 times and a streptavidin-alkaline phosphatase conjugate was added and incubated for 15 minutes. The plate was washed 6 times again and 100uL of a diffusional phosphate solution was added for 10 mins followed by 100uL of developing solution. The plate was read by time-resolved fluorescence.

Prostate specific antigen, Kallikrein-related peptidase 2, Kallikrein-related peptidase 11

The assays for human PSA, KLK2 and KLK11 were conducted using in-house developed ELISAs and were conducted as published previously(198,331,421).

Free PSA

Free PSA was measured using the Elecsys modular analytic system by Roche at Mt. Sinai Hospital, Toronto. Briefly, 20uL of sample and two monoclonal antibodies, one biotinylated and one labelled with a ruthenium complex react to form a complex. Streptavidin-coated microparticles are added to the mixture and bind the complex. The mixture is aspirated and dispensed into a measuring cell where the microparticles are capture by a magnet onto an electrode. Unbound substances are removed and a voltage is applied to the electrode inducing chemiluminescence that is measured by a photomultiplier. The duration of the assay is 18 minutes.
5.2.3 Immunohistochemistry of Follistatin

Paraffin sections were dewaxed in 5 changes of xylene and brought down with water through graded alcohols. Sections were then heat-retrieved in Tris-EDTA buffer at pH 9.0 for 2 minutes at 120°C using a decloaking chamber (Biocare). Sections were allowed to cool off at room temperature for 20 minutes before washing well in running tap water. Endogenous peroxidase and biotin activities were blocked respectively using 3% hydrogen peroxide and an avidin/biotin blocking kit (Lab Vision, Cat# TA-015-BB). Sections were then treated for 10 mins with 10% normal horse serum (Vector Labs, Cat# S2000) before incubated overnight with the mouse monoclonal anti-Follistatin antibody (R&D, Clone 85918, Cat# MAB669) at 1/200 in a moist chamber. This was followed by 30 mins each with biotinylated horse anti-mouse IgG (Vector Labs. Inc. Cat# BA-2001) and HRP-conjugated Ultra Streptavidin (ID Labs. Inc. Cat# BP2378). Colour development was done with freshly prepared NovaRed solution (Vector Labs. Inc: Cat# SK-4800) and counterstained with Mayer’s hematoxylin. Finally, sections were dehydrated through graded alcohols, cleared in xylene and mounted in Permount (Fisher: Cat# SP15-500).

5.2.4 Statistical Data Analysis

To determine if there were significant differences between the median serum levels measured for normal and cancer patient groups tested for CXCL5, LCN2, CXCL1 and TNFRSF6B the Mann-Whitney non-parametric test was employed using 95% confidence intervals and two-sided P values. To determine if there were differences between the multiple groups measured by follistatin, KLK2, PSA, fPSA
and KLK11 the Kruskal-Wallis test was employed using 95% confidence intervals. A Dunn’s multiple comparisons test was performed post-hoc to determine if there were any differences between individual groups. A Spearman correlation was utilized with two-tailed P values and 95% confidence intervals to determine if there was an association between follistatin with age, PSA or Gleason score. All tests were conducted using GraphPad Prism software (GraphPad Software Inc.).
5.3 Results

5.3.1 Chemokine (C-X-C motif) ligand 5

The CXCL5 protein was identified by MS in the CM of the PC3 cell line with both Mascot and X!Tandem proteomic search engines. Two unique peptides and 8 unique spectra were identified by each search engine. To confirm our MS results we measured the concentration of CXCL5 in the CM of each cell line and control flask by ELISA (Figure 5.1A). As can bee seen our ELISA data corroborates with our MS data as CXCL5 was only detected in the PC3 CM by both methods. An initial validation of CXCL5 was also conducted in serum samples from patients considered to not harbour CaP and have CaP (Figure 5.1B). From this preliminary validation we observed no statistical difference in median values between the two groups (p=0.7493). As a result, this candidate was not evaluated further.
Figure 5.1: Measurement of CXCL5 in CM of CaP cell lines and serum of CaP patients and normals. (A) CXCL5 is measured by ELISA in CM of each cell line and only identified in PC3 CM. Number of spectra identified by Mascot and X!Tandem are also shown. (B) Levels of CXCL5 in serum of normal and CaP patients. Horizontal bars represent median values. P value calculated by Mann-Whitney test.
Chapter 5: Validation of Candidate Biomarkers in Prostate Cancer Patient Samples

5.3.2 Chemokine (C-X-C motif) ligand 1

Another candidate preliminarily validated was CXCL1. Similarly CXCL1 was only identified in the CM of the PC3 cell line by MS. The MS results were validated by measuring CXCL1 concentrations by ELISA (Figure 5.2A). The number of unique spectra identified by Mascot and X!Tandem were 30 and 28 respectively, with 3 and 4 unique peptides identified by each search engine. The concentrations of CXCL1 were measured in the serum samples of men with and without CaP (Figure 5.2B), where there was no significant difference determined by the Mann-Whitney test (p=1.0000). Likewise this candidate was not studied further.
Figure 5.2: Measurement of CXCL1 in CM of CaP cell lines and serum of CaP patients and normals. (A) CXCL1 is measured by ELISA in CM of each cell line and only identified in PC3 CM. Number of spectra identified by Mascot and X!Tandem are also shown. (B) Levels of CXCL1 in serum of normal and CaP patients. Horizontal bars represent median values. P value calculated by Mann-Whitney test.
5.3.3 Lipocalin 2

Lipocalin 2 was identified in the CM of two of the CaP cell lines, PC3 and LNCaP by MS. This was validated by measuring the concentration of LCN2 in the CM of each cell line and the control flask. Once again the spectral counts derived from Mascot and X!Tandem agreed with our ELISA data (Figure 5.3A). In addition, LCN2 was also a protein identified by Pilch et al. in seminal plasma(362). We examined the serum concentrations of LCN2 in the serum of patients with and without CaP (Figure 5.3B). No significant difference was observed between groups and the protein was not examined further.
Figure 5.3: Measurement of LCN2 in CM of CaP cell lines and serum of CaP patients and normals. (A) LCN2 is measured by ELISA in CM of each cell line and identified in PC3 and LNCaP CM. Number of spectra identified by Mascot and X!Tandem are also shown. (B) Levels of LCN2 in serum of normal and CaP patients. Horizontal bars represent median values. P value calculated by Mann-Whitney test.
5.3.4 Tumor necrosis factor receptor superfamily, member 6b, decoy

The final candidate investigated, TNFRSF6B was identified solely in the CM of the PC3 CaP cell line by 1 peptide and 7 unique spectra each by Mascot and X!Tandem. Results from the ELISA measurement of TNFRSF6B in the CM of each cell line and control flask also agree with our MS data (Figure 5.4A). Serum samples from patients with and without CaP were measured for TNFRSF6B by ELISA and the results are shown in Figure 5.4B. As can be seen no significant difference was observed between the median values of each group and this candidate was not investigated further.
Figure 5.4: Measurement of TNFRSF6B in CM of CaP cell lines and serum of CaP patients and normals. (A) TNFRSF6B is measured by ELISA in CM of each cell line and identified in PC3 CM. Number of spectra identified by Mascot and X!Tandem are also shown. (B) Levels of TNFRSF6B in serum of normal and CaP patients. Horizontal bars represent median values. P value calculated by Mann-Whitney test.
5.3.5 **Extended Validation of Follistatin**

Measurement of follistatin in the serum of patients pre-biopsy was assessed (Figure 5.5). Median levels of follistatin in each group are shown by horizontal lines and were analyzed by the Kruskal-Wallis non-parametric one-way ANOVA test and no significant difference was observed between the groups (p = 0.9521) (Figure 5.5A). To determine if there was a correlation of follistatin in patients diagnosed with CaP with PSA, age or Gleason score a Spearman correlation was performed (Figure 5.5B, C, D respectively). No significant correlation was observed. In addition, post-hoc analysis of the 95% confidence intervals showed overlapping intervals for each median value, thus further confirming the validity of our result.

The concentrations of PSA, %fPSA, KLK2 and KLK11 were also measured in the same patient cohort (Figure 5.6). The Kruskal-Wallis non-parametric one-way ANOVA test was performed. The only significant difference that was observed was with KLK11 (Figure 5.6D). When examined further by post-hoc pair-wise analysis of the groups using the Mann-Whitney test it was determined that there were significant differences between the benign and CaP group (p=0.0100) and the BPH and CaP group (p=0.0085).

Follistatin expression was also preliminarily investigated in the tissues of prostate, colon and lung cancers by immunohistochemical staining (Figure 5.7, 5.8, 5.9). In the prostate, staining was observed in the luminal epithelial cells and not in the stroma of the benign, PIN, BPH and cancerous glands. All glands stained positive however, staining of prostatic carcinoma (Figure 5.7A) in particular high grade carcinoma (Figure 5.7E) was seen to be more intense relative to benign and
BPH glands (Figure 5.7A, B, D). Interestingly, certain basal cells of benign (Figure 5.7B) and BPH glands displayed a more intense staining (Figure 5.7D). In addition, PIN lesions were seen to stain equally to low grade carcinoma (Figure 5.7C).

Follistatin staining was also evaluated in lung tissue and was localized to the cytoplasm of normal bronchial epithelial cells (Figure 5.8A), pulmonary carcinoma (Figure 5.8C) as well as pulmonary squamous carcinoma (Figure 5.8D, E). Macrophages in the lung were also seen to stain positive (Figure 5.8B). In addition, one out of four specimens from different patients of pulmonary squamous carcinoma was seen to stain intensely (Figure 5.8E) with staining seen in the cytoplasm as well as in the surrounding membrane of the epithelial cells. Tissue specimens from normal and cancerous colon were also investigated for follistatin staining (Figure 5.9). Here we see a more intense stain in the luminal mucosal cells of the colon carcinoma (Figure 5.9C, D) versus the normal mucosal epithelium (Figure 5.9A, B).
Figure 5.5: Follistatin serum levels and correlations with PSA, Gleason score and age. (A) Concentrations of follistatin in patients with benign, BPH, Inflammation, PIN and CaP. Horizontal bars represent median levels. P value calculated by Kruskal-Wallis one-way ANOVA test. (B, C, D) Correlation of follistatin with PSA Gleason score and age in CaP patients, respectively. $r = $Spearman coefficient, P values are calculated using the non-parametric Spearman correlation.
Figure 5.6: PSA, %fPSA, KLK2 and KLK11 serum concentrations in patient serum. (A, B, C, D) Serum concentrations of PSA, %fPSA, KLK2 and KLK11 respectively in serum of patients with benign, BPH, inflammation, PIN and CaP disease. Horizontal lines represent median values. P values are calculated using the Kruskal-Wallis non-parametric one-way ANOVA test.
Figure 5.7: Prostate tissue specimens stained for follistatin. (A) CaP specimen containing benign gland and cancerous gland with perineural invasion (CaP-PNI). (B) Benign and cancerous prostate glands, basal cell staining is evident. (C) CaP specimen displaying PIN lesions and cancerous glands. (D) Prostate specimen displaying a BPH gland with evident basal cell staining (E) Specimen of high grade carcinoma (HG-CaP).
Figure 5.8 Lung tissue specimens stained for follistatin. (A) Normal lung bronchial epithelium (BrEp). (B) Normal lung alveoli, macrophages (Mac) are seen to stain positive for follistatin. (C) Specimen of pulmonary carcinoma. (D, E) Specimens of pulmonary squamous carcinoma, Intense staining of follistatin is seen in the latter specimen, with concentrations of intense membranous staining.
Figure 5.9: Colon tissue specimens stained for follistatin. (A, B) Normal colon mucosa. (C, D) Specimens of colon cancer. Staining of follistatin is evident in the cytoplasm of the columnar mucosal epithelium.
5.4 Discussion

There have been significant efforts to stop the progression of CaP by detecting it at an early stage, but these have been curtailed by lack of a sensitive and specific marker for the disease when it is organ-confined. In addition, there is a lack of an effective therapy that does not involve significant side effects and cause morbidity. Improvements in the prognosis of patients with organ-confined disease are needed to determine those patients that require active management. Therefore, the importance in delineating the role of growth and regulatory mechanisms in CaP translates into targeting the disease earlier through innovative therapies and improved diagnostics.

While the role of androgens in stimulation of CaP is well studied, the roles of associated growth factors are still emerging. Growth factors are seen as playing almost independent roles to androgen stimulation and provide potential targets for therapy for AIPC. One family of proteins, the transforming growth factor β (TGFβ) family have been implicated in progression of many cancers and are key in acting as tumour suppressors(422). Development of resistance to this pathway is an important event in malignant transformation(423). While the involvement of TGFβ has been well determined, the involvement of other members of the family such as follistatin is less well known.

Follistatin is a glycosylated secreted protein that is part of a larger group of proteins with a conserved “follistatin” domain that is rich in cysteines. Follistatin has been shown to be alternatively spliced with two isoforms of proteins FS288 and FS315. The biological function of follistatin was first shown to act to inhibit the
secretion of FSH by binding to proteins known as activins(424). The activins along with inhibins regulate production of gonadotropins with activins stimulating production and inhibins inhibiting production. While follistatin also binds to inhibin it does so in a weaker fashion than to activin where the binding is irreversible(425). The action of follistatin is not limited to the FSH secretion axis as it has been shown to be expressed in a variety of tissues(426) as well as shown to bind other molecules such as bone morphogenic proteins(427). With respect to CaP follistatin has been shown to suppress the growth inhibiting effects of activins.

The localization of follistatin in prostate tissues has been shown in normal, BPH and CaP. Using a polyclonal antibody Thomas et al.(390) were able to localize follistatin in the stroma, basal and epithelial cells of high grade CaP. Expression in BPH was also determined by the same group to be present in the stromal tissue(428). While our study results do not show stromal staining, our results do confirm basal cell staining in the prostate of some benign and BPH glands (Figure 5.7B, D) as well as staining of the epithelial cells in high grade CaP (Figure 5.7E). While the immunohistochemical staining of follistatin in the prostate does not show any obvious diagnostic significance, the interactions between its ligand activin has been shown to play a role in CaP progression.

Co-localisation of activin and follistatin in the prostate support the theory that follistatin reduces the growth inhibitory effects of activin(428). In addition, studies in cell lines have demonstrated the direct effects of follistatin’s ability to neutralise the action of activin(371,372). The PC3 cell line has been shown to secrete both activin and follistatin(371), which was confirmed by our MS and ELISA data. In the PC3 cell
line the inhibitory effect of activin is thought to be neutralised by follistatin in an autocrine manner that promotes cell growth. This was tested by treating PC3 cells with a neutralising antibody to follistatin and adding activin. This resulted in a 60% decrease in cell proliferation thus confirming the theory of follistatin’s neutralising ability over activin(371).

Follistatin has also been shown to have a heparin-binding region and have been shown to interact with the heparin-sulfate chains on proteglycans on the cell surface(429). This was thought to increase its ability to bind activin and sequester it and prevent it from binding its receptor(430). However, another mechanism showed the activin-follistatin complex being internalised and degraded by lysosomes(431). The localization of follistatin to the cell membrane was apparent in one specimen from lung cancer (Figure 5.8E).

While follistatin has been shown to be expressed in several tissues(426), immunohistochemical location in colon and lung tissues has not been shown. Studies with respect to lung cancer have shown follistatin to play an inhibitory role in the progression of metastasis of lung tumours(432). Similar to prostate, the cancerous epithelial cells of both lung and colon stain more intensely than the normal epithelium (Figure 5.8, 5.9). One specimen of lung cancer displayed an intense staining in the epithelial tumour cells. The prognostic significance of this will need to be evaluated by further study to elucidate the prognostic potential of follistatin as a lung tissue biomarker.

While there have been studies on the expression of follistatin in CaP there have not been any determining its potential as a prostate biomarker in serum given
that it has shown to play a role in promoting CaP progression. Our preliminary validation of follistatin in the serum of patients with and without CaP (Figure 4.5A) showed follistatin to have some discriminatory ability. However, this was not confirmed in an extended validation conducted with biopsy confirmed patients (Figure 5.5). The contradictory nature of the results is most likely due to the patients used in the preliminary validation were selected from those who had PSA values greater than 10ug/L. These patients in all probability harbour clinically advanced CaP and those with highly elevated PSA greater than 50ug/L are likely to have metastases. Measuring follistatin in this study set confirmed that follistatin is elevated in a patient population with elevated PSA and that it also correlates positively with increasing PSA (Figure 4.5B). The significance of these results were not seen in serum of patients with biopsy confirmed CaP. In the extended patient population the range of PSA values were mostly in the grey region of 4-10ug/L and the tumours identified by biopsy had Gleason scores in the range of 5-9. Correlations with PSA and Gleason score also did not show any significance with follistatin measured in serum (Figure 5.7B, C). This could be attributed to the fact that these are early stage tumours and levels of follistatin leaking into the circulation are not enough to be measured by ELISA. To ensure there was no bias with age a correlation of serum follistatin concentrations and age of CaP was shown to also be non-significant.

The challenge of discriminating CaP from benign and BPH is seen even with the use of PSA and %fPSA in this population (Figure 5.8A, B). There was no significant difference seen between the groups using PSA or %fPSA. In addition,
two well studied biomarkers for CaP, KLK2 and KLK11 were also measured (Figure 5.8C, D). In this case KLK11 was the only one to show a significant difference between the benign and CaP groups and BPH and CaP groups. However, the clinical relevance of this distinction is minimal considering there is significant overlap in the levels of KLK11 measured in each group.

In conclusion, we show that follistatin is unable to distinguish CaP patients from benign and inflammatory conditions in a patient cohort with PSA levels in the grey region. However, we show for the first time the expression of follistatin in tissues of colon and lung cancer with intense staining in one specimen of lung squamous carcinoma. The clinical significance of this finding will need to be elucidated with further studies in lung tissue specimens and serum samples of patients with and without lung cancer.
CHAPTER 6:

SUMMARY AND FUTURE DIRECTIONS
6.1 Summary

This thesis presents an optimized and validated approach to the discovery of novel candidate CaP biomarkers through proteomic analysis of CaP cell line CM. The use of CaP cell lines as model systems for the study of their secreted proteins by MS approaches resulted in the identification of hundreds of proteins. These proteins were assessed by literature and bioinformatics searches to elucidate their potential as candidates for validation. Select candidates were validated in the serum and tissues of patients with and without CaP.

6.2 Key Findings

1. Proof of Principle
   a. Established a procedure for long term growth of the CaP cell line PC3(AR)₆ in large volume roller bottle culture in SFM
   b. Processed CM in a suitable way for strong anion exchange fast performance liquid chromatography
   c. Prepared fractions for MS analysis by C₁₈ reversed phase HPLC coupled to a linear ion trap mass spectrometer.
   d. Identified 262 proteins from combining proteins identified in two replicates of culture by searching mass spectra with MASCOT
   e. Measured the novel candidate Mac-2BP in the serum of CaP, BPH and normal patients
2. Optimization of Culture Conditions
   a. Seeding density for each of the three CaP cell lines used (PC3, LNCaP and 22Rv1) was varied to optimize the amount of secreted proteins vs. intracellular proteins. Marker proteins for secreted and intracellular fractions were compared and total protein measured to determine optimal seeding densities.

3. Comparative Proteomic Analysis of CM of Three CaP Cell Lines
   a. Cultured and harvested the CM from each CaP cell line in triplicate
   b. Concentrated and prepared the CM for each replicate via peptide fractionation by HPLC SCX chromatography
   c. Analyzed each fraction by MS/MS via C₁₈ reversed phase capillary electrophoresis coupled online to a linear ion trap mass spectrometer
   d. Identified 2124 proteins by searching mass spectra with MASCOT and X! Tandem proteomic search databases and combining results through Scaffold
   e. Determined candidate proteins for validation through a set of criteria

4. Validation of Candidate Biomarkers
   a. Novel candidates were pre-clinically validated in serum of CaP and healthy males.
   b. Four novel candidates were selected for further pre-clinical validation in serums of CaP and healthy males, each showed a positive
correlation with PSA serum levels and a statistically significant increase in serum of CaP patients.

i. Chemokine (C-X-C) motif-16
ii. Spondin 2
iii. Pentraxin 3
iv. Follistatin
c. Follistatin was validated in an extended serum sample set along with KLK2, KLK11 and %fPSA. No discriminatory ability was seen with any marker.
d. Follistatin was evaluated as a tissue marker in the tissues of men with high and low grade CaP and PIN and BPH and was not shown to offer any discriminatory utility.
e. Evaluation of follistatin in the tissues of colon and lung cancers revealed and intense positive staining in one lung cancer specimen

6.3 Future Directions

The approach of this study yielded several novel candidates for validation as CaP biomarkers. The most promising of which was follistatin which showed slight discriminatory staining between CaP and BPH glands. However, even though the clinical utility of this as a tissue marker is not significant follistatin may play a role in CaP progression and further elucidation of its role in CaP is required to determine this. This would involve transcript levels and post translational modification analysis. In addition, follistatin was shown to be expressed in other tissues and we show
staining in other cancers. Thus, more information on the role of follistatin in these cancers is also required.

While we show that CXCL16, SPON2 and PTX3 show significant differences between CaP and healthy males, their roles in CaP progression have not been fully elucidated as well as their roles as tissue markers.

There is a need to examine the list of proteins produced in this study to a greater extent to extract more novel candidates that would require validation. There were several promising candidates that were not validated based on a lack of an ELISA. Developments of antibodies and an ELISA to these candidates would be of value. In addition, development of MS methods to quantify these candidates using techniques such as multiple reaction monitoring would aid in their validation.

Further study of more CaP cell lines would add to the coverage of the ‘secretome’ of CaP and would aid in uncovering more candidates for validation. It is likely that there will not be one biomarker for all tumours and thus a multi-parametric panel is more than likely required to properly assess risk of CaP. Quantitative MS proteomic comparisons of tumour tissue from matched tissue samples from tumour and normal tissue by using isotopic tagging reagents such as iTRAQ would aid in uncovering proteins that show quantitative differences between normal and CaP specimens.
REFERENCES
Reference List


70. Moore MJ, Osoba D, Murphy K, Tannock IF, Armitage A, Findlay B et al. Use of palliative end points to evaluate the effects of mitoxantrone and low-dose


129. Catalona WJ, Partin AW, Finlay JA, Chan DW, Rittenhouse HG, Wolfert RL, Woodrum DL. Use of percentage of free prostate-specific antigen to identify men at high risk of prostate cancer when PSA levels are 2.51 to 4 ng/mL and digital rectal examination is not suspicious for prostate cancer: an alternative model. Urology 1999;54:220-4.


133. Roehl KA, Antenor JA, Catalona WJ. Robustness of free prostate specific antigen measurements to reduce unnecessary biopsies in the 2.6 to 4.0 ng./ml. range. J Urol 2002;168:922-5.


173. Khan MA, Sokoll LJ, Chan DW, Mangold LA, Mohr P, Mikolajczyk SD et al. Clinical utility of proPSA and "benign" PSA when percent free PSA is less than 15%. Urology 2004;64:1160-4.


181. Kwiatkowski MK, Recker F, Piironen T, Pettersson K, Otto T, Wernli M, Tscholl R. In prostatism patients the ratio of human glandular kallikrein to free PSA improves the discrimination between prostate cancer and benign hyperplasia within the diagnostic "gray zone" of total PSA 4 to 10 ng/mL. Urology 1998;52:360-5.


References


References 180


279. Foster BA, Gingrich JR, Kwon ED, Madias C, Greenberg NM. Characterization of prostatic epithelial cell lines derived from transgenic


310. Stamey TA. The era of serum prostate specific antigen as a marker for biopsy of the prostate and detecting prostate cancer is now over in the USA. BJU Int 2004;94:963-4.


312. Scaros O, Fisler R. Biomarker technology roundup: from discovery to clinical applications, a broad set of tools is required to translate from the lab to the clinic. Biotechniques 2005;Suppl:30-2.


Ref Type: Abstract


401. Speetjens FM, Kuppen PJ, Sandel MH, Menon AG, Burg D, van de Velde CJ et al. Disrupted expression of CXCL5 in colorectal cancer is associated with


APPENDIX

ASSAY REPRODUCIBILITY AND PRECISION
Mac-2BP

Theoretical Concentrations (ng/mL) vs. Experimental Concentrations (ng/mL)

klk5

Experimental Concentrations (ug/L) vs. Theoretical Concentrations (ug/L)

n = 2
Appendix: Assay Reproducibility and Precision

**KLK6**

<table>
<thead>
<tr>
<th>Theoretical Concentrations (ug/L)</th>
<th>Experimental Concentrations (ug/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>1000</td>
</tr>
<tr>
<td>1000</td>
<td>10000</td>
</tr>
</tbody>
</table>

\[ n = 2 \]

**KLK2**

<table>
<thead>
<tr>
<th>Theoretical Concentrations (ng/L)</th>
<th>Experimental Concentrations (ng/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>100</td>
<td>1000</td>
</tr>
<tr>
<td>1000</td>
<td>10000</td>
</tr>
<tr>
<td>10000</td>
<td></td>
</tr>
</tbody>
</table>

\[ n = 2 \]
Appendix: Assay Reproducibility and Precision

**KLK11**

![Graph showing theoretical vs. experimental concentrations for KLK11, with n = 2.]

**PSA**

![Graph showing theoretical vs. experimental concentrations for PSA, with n = 2.]

Theoretical Concentrations (ng/L) vs. Experimental Concentrations (ng/L) for both KLK11 and PSA, indicating good reproducibility and precision with n = 2.
SPON 2

Theoretical Concentration (ng/mL)

Experimental Counts

n = 2

CXCL16

Theoretical Controls (ng/L)

Experimental Controls (ng/L)

n = 2
The expected and experimental concentrations for PTX3 and Follistatin are shown in the graphs. The expected concentrations (ng/L) are plotted against the theoretical concentrations (ng/L). The graphs indicate linearity with a slope of 1 for both PTX3 and Follistatin, suggesting good reproducibility and precision.

For PTX3:
- Theoretical Concentrations (ng/L): 0, 400, 800, 1200, 1600, 2000, 2400
- Expected Concentrations (ng/L): Corresponding values for expected concentrations
- Experimental Concentrations (ng/L): Corresponding values for experimental concentrations
- n = 2

For Follistatin:
- Theoretical Concentrations (ng/L): 0, 4000, 8000, 12000, 16000
- Expected Concentrations (ng/L): Corresponding values for expected concentrations
- Experimental Concentrations (ng/L): Corresponding values for experimental concentrations
- n = 2
CXCL1

Experimental Concentrations (ng/L) vs Theoretical Concentrations (ng/L)

n = 2

CXCL5

Experimental Concentrations (ng/L) vs Theoretical Concentrations (ng/L)

n = 2
Appendix: Assay Reproducibility and Precision

TNFRSF6B

Theoretical Concentrations
ng/L

Experimental Concentrations
ng/L

n = 2

Lipocalin 2

Theoretical Concentration
ng/L

Experimental Concentration
ng/L

n = 2