Whole Transcriptome Analysis Reveals Established and Novel Associations with TMPRSS2:ERG Fusion in Prostate Cancer

by

Anthony Chow

A thesis submitted in conformity with the requirements for the degree of Master of Science
Graduate Department of Laboratory Medicine and Pathobiology
University of Toronto

© Copyright by Anthony Chow 2012
Whole Transcriptome Analysis Reveals Established and Novel Associations with TMPRSS2:ERG Fusion in Prostate Cancer

Anthony Chow

Master of Science

Graduate Department of Laboratory Medicine and Pathobiology
University of Toronto

2012

Abstract

Shortcomings of current methods of prostate cancer detection draw attention to a need for improved biomarkers. The TMPRSS2:ERG gene fusion leads to the overexpression of ERG, an ETS family transcription factor, and is the most prevalent genetic lesion in prostate cancer, but its clinical utility remains to be defined. Two radical prostatectomy samples were analysed by next-generation whole transcriptome sequencing. The chosen samples differed in fusion gene status, as previously determined by RT-PCR. The involvement of novel and previously reported prostate cancer-related transcripts, Wnt signalling, p53 effector loss and several ETS-regulated pathways was identified in the prostate cancer cases examined. ERG was found to directly transactivate RhoGDIB, a gene associated with fusion-positive prostate cancer. Overexpression of RhoGDIB elicited spindle-shaped morphology, faster cell migration and increased cell proliferation, phenotypic changes suggestive of cancer progression. The present findings confirm the value of comprehensive sequencing for biomarker development and indicate avenues of future study.
Acknowledgments

2012 Graduation Night

(♫ to the tune of Paradise by the Dashboard Light – with apologies to Meat Loaf)

Well I remember every little thing / as if it happened only yesterday,
Getting off the bus / and there was Sunnybrook S-Wing in sight.
And I’d never seen a lab / lookin’ any better than Seth’s did,
And so I thought right then / that S224 would fit me right!

And now it’s three years later and here I write…
I feel my work is done then I decide – not quite…
And I’m reading PubMed papers choosing who I should cite,
Poring dully through my notebooks all through the night,
I yawn. Deadline’s tight. Then I yawn!
Deadline’s ti-i-ight~

Though it’s cold and lonely writing past midnight,
I can see 2012 graduation night!
Though it’s ti-i-ight~

Ain’t no doubt about it I am pretty stressed,
But I will work for my degree and I will do my best!

Ain’t no doubt about it;
These Acknowledgements will tout it,
Ain’t no doubt about it I’ve been doubly blessed,
With all of you supporting me,
So I can be my best!

Dr. Seth I’ll thank you, first,
You took a chance on me and took me in,
Tania, Wenyi, Aiguo:
For making me feel like a favourite son.
Valentina, Yutaka:
For how wise and patient that you’ve been,

And last but far from least –
dear Aida, Stephanie --
You’re both top-notch pals,
Well, you help keep cancer research fun!

And now it’s three years later and here I write…
I feel my work is done then I decide – not quite…
And I’m reading PubMed papers choosing who I should cite,
Poring dully through my notebooks all through the night,
I yawn. Deadline’s tight. Then I yawn!
Deadline’s ti-i-ight~

Ain’t no doubt about it;
These Acknowledgements will tout it,
Ain’t no doubt about it I’ve been doubly blessed,
With all of you supporting me,
So I can be my best!

Ooh~ sha-sha,
Sha-ooh~ sha-sha,
Sha-ooh~ sha-sha, sha-ooh~ sha-sha,
Ooh~ sha-sha,
Sha-ooh~ sha-sha,
Ooh~ sha-sha,
Sha-ooh~ sha-sha,
Sha-ooh~ sha-sha, sha-ooh~ sha-sha,
Ooh~
Ooh~
Ooh~
Though it’s cold and lonely writing past midnight, 
I can see 2012 graduation night! 
Though it’s cold and lonely writing past midnight, 
Writing past midnight! 
2012 graduation night.

You all did all that you can / to help me be ready for this test, 
Ain’t no doubt about it, I’ve been doubly blessed, 
With all of you supporting me, so I can be my –

I’ve gotta say all my thanks tonight, 
I’ve gotta say all my thanks – my future’s bright! 
I’ve gotta say all my thanks tonight, 
I’ve gotta say all my thanks – my future’s bright!

OK, here we go, I got a lot of people to thank right here, 
Mom, dad, Donald and Paul, no doubt, family comes first, 
I’ve had bad times, but there you were, just keeping me on target, watching me go. 
You really helped me fly!

I’ve got some friends I really couldn’t do without: PACERs/Milk in Bags, you’re key to my successes! 
And MLT and MegaCity, you sure entertained – and kept me sane!

Committee members: Drs. Vasundara, Yousef! 
Exam chair: Elsholtz! Thank you all so much, for letting me make things happen out there.

Next up: folks at Sunnybrook! Drs Nam and Sugar, our paper’s thanks to you! 
And thanks to S-Wing: neighbour labs and Management, Melanie! Thanks to you, things could work.

There’s also a pretty darn special gal: my one partner-ess in crime and other half, 
Duet-er, Mesmer, programming whiz, and it’s Moe! 
Moe who stays right by my side, whose charity knows no bounds.

Everyone, with you, I never had fear! 
I felt sure! You always had my back! 
Holy cow, how could I thank you –

Stop right there! 
Anthony Hang Fung Chow! 
Before you go any further – where’s your degree? 
Is your degree completed? Not that I see. 
Where’s your MSc~ee? 
Have you really forgotten an enormous loose end? 
Can you celebrate now / when you’ve yet to defend?

Where’s your degree? 
Is your degree completed? Not that I see. 
Where’s your MSc~ee? 
Have you really forgotten an enormous loose end? 
Can you celebrate now / when you’ve yet to defend?

Anthony Hang Fung Chow! 
Before you go any further – where’s your degree? 
Is your degree completed?
I’m all over it. / Trust me, really, I’m all over it.
I aim to impress my whole committee.

Trust me, really, I’m all over it.
I aim to impress my whole committee.

I’m all over it.
Trust me, really, I’m all over it.
I aim to impress my whole committee.

I’m all over it.
Trust me, really, I’m all over it.
I aim to impress my whole committee.

I aim to impress my whole committee.

Anthony Hang Fung Chow, / where’s your degree?
Is your degree completed? Not that I see.
Where’s your MSc~ee?
Have you really forgotten an enormous loose end?
Can you celebrate now / when you’ve yet to defend?

Let...let...let...
Let me get to it! / Leave it to me, let me get to it!
Let me get to it! / I will get a pass from my committee!

Let me get to it!
Leave it to me,
Let me get to it!

Let me get to it!
I will get a pass from my committee,
Committee, a pass from my committee!

Anthony Hang Fung Chow!
Where’s your degree?
Is your degree completed?
Not that I see!
Where’s your MSc~ee?
Have you really forgotten an enormous loose end?
Can you celebrate now / when you’ve yet to defend?

Let me get to it!

Is your degree completed?

Let me get to it!

Is your degree completed?
I can’t take it any longer, / Lord I’m aflame!
Oh look what torture prostate cancer research work became!
I will now study, make slides, practice, touch no video game…
And keep it up until the end of time!

So now I’m praying for the end of time…
To hurry up and arrive,
‘Cause if I take another look at these datasets,
I don’t think that I can really survive.

I’ll never rest until I’m known as Dr. Chow,
But God only knows I’m sick of this right now,
I’m praying for the end of time,
It’s all that I can do!
Praying for the end of time,
So I can end my time with ETS.

It was long ago and it’s to my dismay,
That it was so much better than it is today.
It was long ago and it’s to my dismay,
That it was so much better than it is today.
It was long ago and it’s to my dismay,
That it was so much better than it is today.
It was long ago and it’s to my dismay,
That it was so much better than it is today.
It was long ago and it’s to my dismay,
That it was so much better than it is today.
It was long ago and it’s to my dismay,
That it was so much better than it is today.
It was long ago and it’s to my dismay,
That it was so much better than it is today.
It was long ago and it’s to my dismay,
That it was so much better than it is today.

So now I’m praying for the end of time…
To hurry up and arrive,
‘Cause if I take another look at these datasets,
I don’t think that I can really survive.

I’ll never rest until I’m known as Dr. Chow,
I’m praying for the end of time,
It’s all that I can do~oo~oo!
Praying for the end of time,
So I can end my time with ETS~

I do transfections fine but nothing works quite right…
And I am reading PubMed papers and continue to write.
I feel my work is done then I decide – not quite…
And I am listing References choosing who I should cite.
I know with this degree my future could be bright…
And so I hope in 2012 my graduation’s in sight.
I do transfections fine but nothing works quite right…
And I am reading PubMed papers and continue to write.
I feel my work is done then I decide – not quite…
And I am listing References choosing who I should cite...
# Table of Contents

Acknowledgments........................................................................................................ iii
Table of Contents......................................................................................................... vii
List of Abbreviations .................................................................................................. x
List of Tables ............................................................................................................. xii
List of Figures ............................................................................................................ xiii

## Chapter 1 Introduction................................................................................................. 1
1 Prostate Cancer........................................................................................................... 1
   1.1 Prostate Cancer Metastasis ............................................................................... 1
   1.2 Management of Prostate Cancer ....................................................................... 3
       1.2.1 Active Surveillance ............................................................................... 3
       1.2.2 Medical Castration ............................................................................. 3
       1.2.3 Surgery ................................................................................................. 4
       1.2.4 Radiation Therapy and Chemotherapy ............................................... 5
   1.3 Heterogeneity of Prostate Cancer ...................................................................... 5
       1.3.1 Stage and Grade .................................................................................... 5
       1.3.2 Recurrence ............................................................................................ 6
       1.3.3 Androgen Sensitivity .......................................................................... 7

## Chapter 2 Biomarkers of Prostate Cancer.................................................................. 7
  2.1 Historical Biomarkers of Prostate Cancer.......................................................... 8
  2.2 Current Biomarkers of Prostate Cancer ............................................................... 8
  2.3 Emerging Biomarkers of Prostate Cancer............................................................ 9

## Chapter 3 ETS Transcription Factor Family.............................................................. 11
  3.1 ETS and PNT Domains ...................................................................................... 11
  3.2 ETS Involvement in Disease .............................................................................. 12
  3.3 ETS Involvement in Prostate Cancer ................................................................. 13
       3.3.1 ERG ..................................................................................................... 13
       3.3.2 TMPRSS2:ERG Gene Fusion ................................................................. 15

## Chapter 4 Next-Generation Sequencing ................................................................ 17
  4.1 SOLiD Sequencing .............................................................................................. 17
  4.2 Applications of Next-Generation Sequencing ..................................................... 18
4.3 Transcriptome Analysis .................................................................................................................. 19

4.3.1 Advantages of RNA Sequencing ................................................................................................. 19

4.3.2 Limitations of RNA Sequencing .................................................................................................. 20

4.4 Applications of RNA Sequencing ................................................................................................... 20

4.4.1 RNA Sequencing in Cancer Research ......................................................................................... 21

4.4.2 RNA Sequencing in Prostate Cancer Research ........................................................................... 21

5 Rationale, Hypothesis and Objectives ............................................................................................... 22

Chapter 2 Materials and Methods ........................................................................................................ 24

1 RNA Extraction and Preparation ......................................................................................................... 24

2 SOLiD System Whole Transcriptome Sequencing ........................................................................... 24

2.1 Sequencing Data Analysis .............................................................................................................. 25

3 Plasmid Construction .......................................................................................................................... 25

4 Cell Culture ......................................................................................................................................... 27

4.1 Stable Transfection ............................................................................................................................ 27

4.2 Luciferase Reporter Assay ............................................................................................................... 27

4.3 Wound Closure Assay ...................................................................................................................... 29

4.4 ³H-thymidine Incorporation Assay ................................................................................................... 29

Chapter 3 Results .................................................................................................................................... 30

1 Primary Analysis of Sequencing Data .................................................................................................. 30

2 Gene and Transcript Expression ......................................................................................................... 32

2.1 Expression of ETS and ETS-Targeted Genes ............................................................................... 33

2.2 Expression of Prostate Cancer-Related Transcripts ....................................................................... 34

3 Involvement of RhoGDIB in Prostate Cancer ................................................................................... 35

3.1 Effect of ERG on RhoGDIB Expression .......................................................................................... 36

3.2 Effect of RhoGDIB on PC3 Phenotype ........................................................................................... 38

4 Wnt Signalling and TMPRSS2:ERG+ Prostate Cancer ...................................................................... 41

5 p53 Variant and TMPRSS2:ERG– Prostate Cancer ............................................................................. 43
Chapter 4 Discussion........................................................................................................45
1 Utility of RNA-Seq in Prostate Cancer Research..........................................................45
   1.1 Detection of Prostate Cancer-Related Transcripts......................................................45
   1.2 Detection of Prostate Cancer Biomarkers..................................................................47
2 Significance of ETS Involvement in Prostate Cancer.....................................................48
   2.1 ETV1 ..........................................................................................................................48
   2.2 EGR1 ..........................................................................................................................48
   2.3 RhoGDIB .....................................................................................................................49
3 Significance of Wnt Signalling in Prostate Cancer..........................................................51
   3.1 PI3K Pathway ..............................................................................................................52
   3.2 AR ..............................................................................................................................53
   3.3 Other Components of Wnt Signalling.........................................................................54
4 Significance of p53 Variant in Prostate Cancer..............................................................55
   4.1 Cell Cycle Control Genes ..........................................................................................56
   4.2 Apoptosis-Inducing Genes .......................................................................................57
   4.3 p53 Polymorphism .....................................................................................................58
5 Future Directions.............................................................................................................59
6 Conclusion.........................................................................................................................61

Chapter 5 References........................................................................................................62
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABL</td>
<td>v-abl Abelson murine leukemia viral oncogene homolog 1</td>
</tr>
<tr>
<td>AMACR</td>
<td>α-Methylacyl coenzyme A racemase</td>
</tr>
<tr>
<td>AML1</td>
<td>acute myeloid leukaemia 1</td>
</tr>
<tr>
<td>APC</td>
<td>adenomatous polyposis coli</td>
</tr>
<tr>
<td>AR</td>
<td>androgen receptor</td>
</tr>
<tr>
<td>BI-1</td>
<td>Bax inhibitor-1</td>
</tr>
<tr>
<td>CD151</td>
<td>cluster of differentiation 151</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CDS</td>
<td>coding DNA sequence</td>
</tr>
<tr>
<td>CTSB</td>
<td>cathepsin B</td>
</tr>
<tr>
<td>DD3^PCA3</td>
<td>differential display code 3</td>
</tr>
<tr>
<td>DGKB</td>
<td>diacylglycerol kinase, β</td>
</tr>
<tr>
<td>EBS</td>
<td>Ets binding site</td>
</tr>
<tr>
<td>ECM</td>
<td>extracellular matrix</td>
</tr>
<tr>
<td>EGR1</td>
<td>early growth response 1</td>
</tr>
<tr>
<td>ELK4</td>
<td>ETS-domain protein (SRF accessory protein 1)</td>
</tr>
<tr>
<td>EMT</td>
<td>endothelial mesenchymal transition</td>
</tr>
<tr>
<td>ERG</td>
<td>Ets-related gene</td>
</tr>
<tr>
<td>ERK</td>
<td>extracellular signal-regulated kinase</td>
</tr>
<tr>
<td>ESRRA</td>
<td>estrogen-related receptor α</td>
</tr>
<tr>
<td>ETS</td>
<td>E twenty six</td>
</tr>
<tr>
<td>ETV1</td>
<td>Ets variant 1</td>
</tr>
<tr>
<td>FISH</td>
<td>fluorescent in situ hybridization</td>
</tr>
<tr>
<td>FLI-1</td>
<td>Friend leukemia integration 1 transcription factor</td>
</tr>
<tr>
<td>FOXP1</td>
<td>forkhead box protein P1</td>
</tr>
<tr>
<td>FZD</td>
<td>frizzled</td>
</tr>
<tr>
<td>GADD</td>
<td>growth arrest-DNA damage</td>
</tr>
<tr>
<td>GSK3β</td>
<td>glycogen synthase kinase 3β</td>
</tr>
<tr>
<td>GTPase</td>
<td>guanosine triphosphate hydrolase</td>
</tr>
<tr>
<td>IGFBP3</td>
<td>insulin-like growth factor-binding protein 3</td>
</tr>
<tr>
<td>IGF1</td>
<td>insulin-like growth factor 1</td>
</tr>
<tr>
<td>IL3/8</td>
<td>interleukin 3/8</td>
</tr>
<tr>
<td>KEGG</td>
<td>Kyoto Encyclopaedia of Genes and Genomes</td>
</tr>
<tr>
<td>LEF</td>
<td>lymphoid enhancer-binding factor</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>MALAT1</td>
<td>metastasis associated lung adenocarcinoma transcript 1</td>
</tr>
<tr>
<td>MIPOL1</td>
<td>mirror-image polydactyly 1</td>
</tr>
<tr>
<td>NCBI 1</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>NGS</td>
<td>next-generation sequencing</td>
</tr>
<tr>
<td>MAP</td>
<td>mitogen-activated protein</td>
</tr>
<tr>
<td>MDM2/4</td>
<td>murine double minute 2/4</td>
</tr>
<tr>
<td>MMP</td>
<td>matrix metalloprotease</td>
</tr>
<tr>
<td>MN1</td>
<td>meningioma (disrupted in balanced translocation) 1</td>
</tr>
<tr>
<td>PAP</td>
<td>prostatic acid phosphatase</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>pCMV</td>
<td>plasmid (cytomegalovirus)</td>
</tr>
<tr>
<td>PI3K</td>
<td>phosphoinositide 3-kinase</td>
</tr>
<tr>
<td>PIN</td>
<td>prostatic intraepithelial neoplasia</td>
</tr>
<tr>
<td>PNT</td>
<td>pointed (domain)</td>
</tr>
<tr>
<td>PSA</td>
<td>prostate specific antigen</td>
</tr>
<tr>
<td>Rac1</td>
<td>Ras-related C3 botulinum toxin substrate 1</td>
</tr>
<tr>
<td>Rho</td>
<td>Ras homolog</td>
</tr>
<tr>
<td>RhoGDIB</td>
<td>Rho GDP dissociation inhibitor β</td>
</tr>
<tr>
<td>RPM</td>
<td>reads per million</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>reverse transcriptase-polymerase chain reaction</td>
</tr>
<tr>
<td>RYBP</td>
<td>RING1 and YY1 binding protein</td>
</tr>
<tr>
<td>siRNA</td>
<td>small interfering RNA</td>
</tr>
<tr>
<td>SLC45A3</td>
<td>solute carrier family 45, member 3</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>SOLiD</td>
<td>sequencing by oligonucleotide ligation and detection</td>
</tr>
<tr>
<td>SPI1</td>
<td>spleen focus forming virus (SFFV) proviral integration oncogene</td>
</tr>
<tr>
<td>SOX2</td>
<td>sex determining region Y-box 2</td>
</tr>
<tr>
<td>TCF</td>
<td>T-cell factor</td>
</tr>
<tr>
<td>TGfβ</td>
<td>transforming growth factor β</td>
</tr>
<tr>
<td>TMPRSS2</td>
<td>transmembrane protease, serine 2</td>
</tr>
<tr>
<td>vWF</td>
<td>von Willebrand factor</td>
</tr>
<tr>
<td>ZEB1</td>
<td>zinc finger E-box-binding homeobox 1</td>
</tr>
</tbody>
</table>
List of Tables

Table 1. Characteristics of sequencing data for two radical prostatectomy samples as determined by whole transcriptome sequencing..............................................................30

Table 2. Expression of ETS transcription factors and their targets in two radical prostatectomy samples as determined by whole transcriptome sequencing.........................................................33

Table 3. Expression of prostate cancer-related genes in two radical prostatectomy samples as determined by whole transcriptome sequencing..............................................................34

Table 4. Expression of p53-regulated genes as determined by whole transcriptome sequencing.44
List of Figures

Figure 1. General structure of the TMPRSS2:ERG gene fusion ...........................................16

Figure 2. Human ERG coding DNA and corresponding protein sequences .................................26

Figure 3. Plasmid constructs used in luciferase reporter assay ................................................28

Figure 4. Scatterplot displaying expression of transcripts for two radical prostatectomy samples
as identified by whole transcriptome sequencing ........................................................................31

Figure 5. Correlation between TMPRSS2:ERG fusion and RhoGDIB (ARHGDI B) expression in
tumours obtained from a Toronto cohort .....................................................................................35

Figure 6. Multiple EBS are upstream of the RhoGDIB transcription start site ............................36

Figure 7. ERG positively regulated RhoGDIB promoter activity ..................................................37

Figure 8. Endogenous expression of RhoGDIB protein in various prostate cell lines ...............38

Figure 9. Effects of RhoGDIB overexpression on PC3 cell morphology and motility ..............39

Figure 10. Transient overexpression of ΔN-ERG in PC3 and PC3-7C ........................................40

Figure 11. Effect of RhoGDIB overexpression on PC3 cell proliferation ....................................41

Figure 12. Expression levels of Wnt signalling pathway genes in the fusion-positive sample
relative to the fusion-negative sample .........................................................................................42

Figure 13. Amino acid substitution attributed to rs1042522 .......................................................43

Figure 14. Mechanisms of action of RhoGDIB .........................................................................50

Figure 15. Model of positive feedback loop between Wnt and ERG .........................................54

Figure 16. Automated analysis of rs1042552 by PolyPhen-2 ......................................................58
1 Prostate Cancer

Prostate cancer is the most common solid tumour cancer among Canadian men. The Canadian Cancer Society estimates that 1 in 7 Canadian men will develop prostate cancer in their lifetime and 1 in 4 of those will die from it. Risk of developing prostate cancer increases dramatically with age (1) and family history (2) and may be positively correlated with components of Western lifestyle, such as diet (3). The high prevalence of the disease constitutes a significant burden on healthcare and society.

Prostate cancer arises in the epithelium of the gland, most frequently in the peripheral zone. It is sometimes preceded by the development of high-grade prostatic intraepithelial neoplasia (PIN) less than one decade before the onset of carcinoma. This condition is defined by the presence of abnormally proliferative prostate cells that remain associated with an intact or incomplete basal layer. Unlike normal prostate epithelium, in PIN most cell proliferation occurs in the luminal (as opposed to the basal) layer and cells karyotypically and phenotypically resemble prostate cancer (4, 5).

Proliferative lesions that cross the basal layer represent more serious disease and are considered malignant as they have the potential to invade adjacent tissues and then metastasise to distant sites. This most frequently occurs in the form of prostate adenocarcinoma, or invasive prostate cancer arising from the secretory epithelium, though in rare cases prostate cancer can arise from basal cells, endocrine cells and transitional epithelium (6). Prostate cancer often also arises as multifocal disease, complicating detection and treatment (7).

1.1 Prostate Cancer Metastasis

Prostate cancer most frequently metastasizes to local lymph nodes and bone, though liver and lung metastases are also common. Metastasis to local sites, such as the seminal vesicle and pelvic lymph nodes, can be explained by the “mechanical theory”: proliferative lesions expand beyond the prostatic capsule and directly affect adjacent tissues or reach nearby sites, such as the lumbar spine, by lymphatic dissemination via a porous structure called the Batson plexus. Spread
to distal sites, such as bone, liver and lung, is more aptly described by the “seed and soil theory”: cancerous cells achieve passage through endothelium and enter systemic circulation, sometimes with the involvement of angiogenesis to the primary tumour. Subsequent metastases, or daughter tumours, form in cellular microenvironments that are amenable to tumour reestablishment (8,9).

Several elements contribute to the suitability of a microenvironment for site-specific metastasis. First, paracrine release of chemoattractants like epidermal growth factor (EGF) (10) and osteonectin (11) by specific tissues can influence the migration of tumour cells. Second, such sites encourage the settlement of circulating seed cells. This can be achieved passively, as by dimensional constriction in narrow capillaries of the lung, or actively, as by expression of adhesion molecules on tumour cells, endothelial cells or even the basement membrane. For example, β1 integrin (12), collagen, laminin, fibronectin (13) and even cell surface-associated prostate-specific antigen (PSA) (14) are all conducive to prostate cancer cell attachment. Importantly, specific, adhesion molecule-mediated binding is more efficient than nonspecific entrapment at promoting metastasis (9).

Third, high levels of mitogenic factors favour the establishment of new tumours from arriving seed cells. Such elements are often differentially expressed between tissues and promote cell division in both normal cells and metastatic cells arriving from distant sites. Bone morphogenetic proteins (BMPs), for example, are highly expressed in bone and may facilitate the expansion of bone metastases from the prostate. Several mechanisms have been proposed for this effect, including the induction of growth factor release and osteoblastic changes that characterise bone metastasis of prostate cancer (15). Transforming growth factor (TGF)-β is also abundant in bone and cartilage and can affect metastasing cells alone or through such effectors as connective tissue growth factor (CTGF), a powerful promoter of endothelial cell adhesion, proliferation and migration (16).

Fourth, sites of metastasis are vulnerable to extracellular matrix (ECM) degradation and subsequent invasion by cancer cells. Tissues of the body vary in their ECM composition and some of these differences increase the probability of prostate metastasis formation: liver is rich in heparan sulphate proteoglycans while lung has high levels of collagen type IV, laminin and elastin (9). Proteases that target the matrix in prostate cancer metastasis include matrix
metalloproteinase (MMP)-9, which degrades collagens, fibronectin, elastin and laminin (17,18), and urokinase, an activator of the plasmin protease cascade (19).

Metastasis dramatically decreases the treatability of prostate cancer (8), so understanding the mechanisms of its spread is essential to preventing and managing the formation of tumours at secondary sites. Historically, both the mechanical and the seed and soil theories of metastasis guide prostate cancer research and therapy (9).

### 1.2 Management of Prostate Cancer

A variety of treatment modalities for prostate cancer exist. As with the management of any other disease, the selection of one of these modalities depends on the severity of the disease, the stability of the patient’s condition and patient and physician preference in terms of procedure invasiveness and adverse side effect risk (20).

#### 1.2.1 Active Surveillance

Despite the extreme prevalence of prostate cancer, less than 5% of cases eventually become the cause of death. Given these facts, active surveillance, also known as “watchful waiting” or “expectant management”, is a reasonable choice for patients with non-threatening disease or a desire to avoid common side effects of prostate cancer therapy, such as sexual and urological dysfunction. More involved intervention can always be initiated at a later time if circumstances or preferences change. Of course, this clinical decision also carries the risk of accidentally delaying treatment to a point beyond which it is effective (21).

In a recent review of 88 studies about active surveillance of prostate cancer, active surveillance was invariably associated with high rates of survival and low rates of disease progression during the waiting period. Since 2006, organisations as diverse as the American, French and Brazilian Urological Associations have deemed active surveillance a viable therapeutic option; in the United Kingdom, it is even the only recommended option for low-risk disease (22).

#### 1.2.2 Medical Castration

Growth of the prostate organ is androgen-dependent and remarkable in that it continues throughout life. When prostate cancer exists, androgens may instead stimulate the proliferation
of tumour cells (23). For this reason, androgen deprivation, or castration, is a first-line choice for prostate cancer management (21).

Testosterone is the primary male androgen, and in the adult male the vast majority of it is produced in the testes (24). Production of testosterone is stimulated by the release of pituitary follicle-stimulating hormone (FSH) and luteinising hormone (LH), itself mediated by hypothalamic gonadotropin-releasing hormone (GnRH). Currently, castration can be achieved medically via targeting the physiological action of GnRH (23).

Many synthetic analogues of GnRH, like leuprolide, have been developed. These can be over 100 times more potent than natural GnRH (25) and act by overstimulating GnRH receptors and thus promoting their internalisation (26). This eventually leads to GnRH insensitivity and progressively lower levels of secreted LH and FSH over time (27). Initially, however, GnRH agonists precipitate painful and dangerous clinical flares, causing such symptoms as bone pain (28), nerve compression, ureteral blockage (23) or even a brief surge in tumour growth (21). To reduce the severity of flares, anti-androgens like flutamide are co-prescribed with GnRH agonist therapy. These directly block androgen receptors and have the incidental effect of inhibiting androgens derived from adrenal gland precursors (27).

GnRH antagonists produce the desired testosterone reduction without the undesired clinical flare, but sufficiently potent and bioavailable compounds are difficult to develop and produce. Still, antagonists like abarelix are available and successfully achieve medical castration without transient upregulation of LH or testosterone production (23).

1.2.3 Surgery

Surgery is indicated for tumour resection, symptom management and, less commonly, hormonal manipulation. Earlier detection of prostate cancer, enabled by modern screening practices, is giving rise to more and better opportunities for surgical intervention. Radical prostatectomy is often prescribed for patients with localised disease, with lymph node dissections performed concurrently or separately if there is a need to assess metastasis. As an alternative to radical prostatectomy, cryosurgery is useful for debulking a primary tumour while preserving a partial prostate and avoiding comorbidities, but carries a proportionate risk of incomplete cancer removal (21).
Bilateral orchiectomy elicits a degree of testosterone reduction comparable to that achieved by medical castration but is much less expensive since it does not require life-long management. Still, it is often not the treatment of choice because testes removal is irreversible, incurs surgical risks and imposes negative psychological impacts (29).

1.2.4 Radiation Therapy and Chemotherapy

Ionizing radiation preferentially damages the DNA and inhibits the growth of rapidly proliferating cells, such as tumour cells. Two dominant options exist for delivering radiation therapy while minimising the volume of healthy tissue affected: targeted and/or intensity-modulated external beam radiation or interstitial brachytherapy, the physical insertion of a radiation source into the prostate. Both methods are useful for treatment of localised disease and may be effective in combination with other therapies (21). Radiation therapy places adjacent anatomic structures, such as the bladder and rectum, at greatest risk of collateral toxicity (30).

Systemic chemotherapy becomes necessary for disseminated disease or when other modes of intervention have failed or become unavailable, though many such agents have been shown to grant little clinical benefit. All chemotherapeutic drugs are toxic and major side effects include muscle loss, gastrointestinal and cardiovascular toxicity, nerve damage and hair loss. In a review of 107 randomised trials of chemotherapy, only docetaxel stood out for significantly improving survival over regular disease management, though several other combination therapies were reported to be of various benefit (31). Pharmaceutical or nutritional regimes may also be used for prostate cancer prevention, though current evidence does not yet support widespread use of such interventions for all men (32).

1.3 Heterogeneity of Prostate Cancer

As the discussion of treatment options suggests, management of prostate cancer is complicated by its heterogeneity. Appropriate intervention ranges from no intervention at all to whole prostate removal to systemic cytotoxic pharmaceutical use and combination therapy. Instances of prostate cancer differ about several characteristics that are relevant to treatment decisions.

1.3.1 Stage and Grade

De-differentiation is one of the hallmarks of cancer and can be assessed by histopathological analysis. Prostate biopsy is the current gold standard for diagnosing prostate cancer and is a
minor enough procedure that it does not require sedation and can be routinely ordered for men suspected of having disease (33). Gleason score, a measure of disease grade, is found by determining and then summing the Gleason values of the two most representative regions on a histological slide. These values range from one to five, arranged in order of increasing anaplasia, giving a maximum Gleason score of 10 (34). Though Gleason score naturally varies from biopsy to biopsy and pathologist to pathologist, it remains the best predictor of cancer aggressiveness and treatment outcome and is an important tool for measuring recurrence risk (33,35).

Prostate cancer staging is intended to classify clinical instances of prostate cancer into internally similar groups that can be used for assessing patient risk, but these criteria can be subjective and controversial. As recently as 2006, the American Joint Committee on Cancer (AJCC) guidelines for prostate cancer staging were criticised as being insensitive for predicting patient outcome. This criticism was levelled along with a proposal to include pre-treatment PSA level and Gleason score, two risk stratification measures, among staging guidelines (36). Current AJCC criteria for staging prostate cancer, released in 2009, consider tumour size, lymph node status, extent of metastasis and histopathologic grade. Better prostate cancer screening practices have allowed clinicians to identify diseases earlier in their manifestation, as evinced by decreasing average stage at diagnosis (33).

1.3.2 Recurrence

Treatment failure occurs in approximately 30 – 40% of prostate cancer cases treated with radiotherapy (37). Hormonal therapy is predictably followed by re-emergence of disease within two years (38). Recurrence may arise locally, especially following local interventions like partial prostatectomy and targeted radiotherapy, or in the form of metastatic disease (39). Recurrence, if it occurs, usually necessitates more extreme treatment – salvage prostatectomy, for example, may follow failed tissue-preserving surgery (21). Risk of recurrence exists for all patients and necessitates a suitable follow-up period subsequent to initial treatment (28).

During this monitoring period, serum PSA level is often used as a surrogate indicator of disease recurrence. PSA levels are expected to drop and remain low following successful treatment of prostate cancer. Though specific standards differ, successive or large increases in post-treatment PSA levels, termed biochemical recurrence, may identify patients with clinically recurrent or further disseminated disease (37). Still, only approximately one in three patients with
biochemical recurrence actually go on to develop metastatic disease and the median time
between biochemical and clinical recurrence is eight years (40). Although regular post-treatment
prostate biopsies are much more accurate at determining clinical recurrence, they are seldom
performed due to cost, inconvenience and risk of complication (37).

1.3.3 Androgen Sensitivity

Hormone-refractory or androgen-insensitive prostate cancer is a form of the disease that can
progress even in the absence of androgens, rendering androgen deprivation therapy ineffective.
Primary prostate tumours are heterogeneous and composed of cells that may be androgen-
dependent, -sensitive or -independent (21). Mechanistically, insensitive cells are impaired in
their ability to undergo apoptosis and therefore do not require certain concentrations of
androgens to survive (41). This phenotypic change may arise as part of multifocal disease, due to
genetic instability or as a result of adaptation to hormonal manipulation (42).

Androgen insensitivity is particularly characteristic of prostate cancer that recurs following
hormonal manipulation. At this point, cell death elicited by androgen deprivation is overtaken by
proliferation of unaffected cancer (42). Though essentially all androgen-deprived prostate
cancers become androgen-insensitive over time, the use of intermittent rather than sustained
hormonal therapy may delay this development (41).

When androgen deprivation therapy fails, limited hormonal options still exist: these include
direct antiandrogens, estrogens, and adrenal suppressors. Generally, however, these measures are
ineffective and are only used as a temporary measure until chemotherapy can begin (21,28).

2 Biomarkers of Prostate Cancer

Prostate cancer is best approached with active surveillance as early as possible – at lower stages,
more treatments are available, side effects of intervention are less severe and patient outcome is
more optimistic. Due to the high prevalence and low disease-specific mortality of prostate
cancer, however, it is important to additionally distinguish between occurrences of disease that
are likely to lead to symptoms or become aggressive and those that are not. Unfortunately, there
is currently no agreed-upon criterion for identifying lethal prostate cancer. As a case in point, the
Epstein criteria for distinguishing clinically significant and insignificant disease are of disputed
accuracy (43,44). In the face of uncertainty, intervention tends to err on the side of caution, leading to overdiagnosis and overtreatment (45,46).

Biomarkers are biological molecules that can be detected in bodily fluids or tissues and are indicative of disease. They can be broadly categorised as being diagnostic, prognostic (indicative of likely disease course or outcome) or predictive (indicative of likely response to a particular intervention), though uses of biomarkers extend to risk stratification, screening (initial population-scale detection of disease) and post-treatment monitoring (47). The development of biomarkers capable of identifying any or certain subsets of prostate cancer has unsurprisingly long been a field of intense research, yet no existing markers can predict whether a particular instance of disease is likely to recur or metastasise (48).

2.1 Historical Biomarkers of Prostate Cancer

Historically, prostatic acid phosphatase (ACPP or PAP) was the first prostate cancer biomarker detectable in serum. Total serum AP, a measurement of phosphatases derived from tissues throughout the body, was also clinically useful. Elevation of serum PAP is positively correlated with metastasis to the bone, risk of eventual lymph node involvement and higher staging. Unfortunately, PAP is neither sensitive nor specific enough as a biomarker to guide prostate cancer management. Levels of serum PAP often remain low in men with disease and have been known to fluctuate suddenly, casting doubt on the reliability of any particular measurement (47). Conversely, though measurement of PAP on sectioned biopsy material is associated with poor survival, this test is too indiscriminate to be useful as almost all tumour sections are strongly immunoreactive to PAP (49).

Though no longer considered useful for diagnosis, PAP is under investigation as a biomarker of patient outcome following prostatectomy. Specifically, high post-surgery PAP in the serum is associated with a greater risk of biochemical recurrence (50).

2.2 Current Biomarkers of Prostate Cancer

PSA, also known as kallikrein 3, is a member of the kallikrein family of serine proteases that is secreted by the prostate and originally gained scientific interest as a forensic marker for human semen. The discovery that PSA was easily detectable in serum and present at higher levels in prostate cancer patients allowed it to eventually replace PAP as a biomarker for prostate cancer.
Measurement of serum PSA is currently the most frequently used method of prostate cancer detection (47,51).

Diminishing the usefulness of PSA is the fact that it is more accurately a surrogate marker of prostate size and integrity. Benign prostatic hypertrophy can increase serum PSA levels and can in fact imitate many classic symptoms of prostate cancer, such as urinary obstruction (8). In older men, the demographic most at risk of morbidity and mortality from prostate cancer, other non-malignant processes additionally confound PSA test interpretation (47). Even sexual stimulation or minimally invasive medical procedures, including a digital rectal exam, can cause an irrelevant rise in serum PSA (52). Owing to these shortcomings, PSA detects prostate cancer with less than 50% positive predictive value and almost 75% of biopsies performed as a result of a suspicious PSA measurement are negative for cancer (46).

Attempts to refine the interpretation of PSA measurements take into account the proportion of free to protein-bound PSA or the concentration of proPSA, a PSA precursor, both of which can help distinguish cancer from hypertrophy (47). Tracking PSA velocity, the rate of increase in serum PSA per year, may aid in differentiating between patients who require no, local or systemic intervention (53). Still, the utility of any of these parameters is contested (54-57) and to wit, the U.S. Preventive Services Task Force recently assigned PSA-based screening for prostate cancer a grade D recommendation, which represents the lowest possible confidence in the test’s ability to provide a net clinical benefit (58).

2.3 Emerging Biomarkers of Prostate Cancer

The discovery and development of prostate cancer biomarkers has been hastened by improvements to biotechnology and consequent advancements in the understanding of prostate cancer biology. Though none have yet become clinically established, their existence highlights the urgency with which biomarker research is being conducted (47).

α-Methylacyl coenzyme A racemase (AMACR) is a peroxisome- and mitochondria-localized racemase that functions in bile and fatty acid metabolism. It was first identified as overexpressed in prostate cancer through DNA microarray analysis and found to be very sensitive and specific at detecting prostate cancer in prostate needle biopsies. At the mRNA level, AMACR transcripts were detected exclusively in cancer as opposed to benign samples. At the protein level, high- and
moderate-intensity immunostaining sensitively and specifically corresponded to cases with cancer, even those that other advanced pathological techniques failed to confidently diagnose (59). AMACR staining can also aid in diagnosis when pathologists are confronted by atypical small acinar proliferation, a histological finding that is suspicious but by itself insufficient evidence of adenocarcinoma (60). The specificity of the assay can be improved with by simultaneously staining for basal cell markers, such as keratin 903 (34βE12) and p63, which are often lost by malignant prostate cells (61). Still, AMACR immunostaining is as yet mostly limited to post-screening biopsy analysis.

Early prostate cancer antigen (EPCA) is a structural component of the nuclear matrix first studied through its rat homologue, Am-1. Chromosomal alterations characteristic of prostate cancer are often reflected by changes in the configuration of the nucleus and consequent alterations in matrix protein levels (47). As its name implies, EPCA is theorised to be upregulated early in the development of prostate cancer as its expression can be detected in premalignant tissues and is not correlated with disease grade or stage (62). Interestingly, EPCA can be immunochemically detected in tumour and non-cancerous, tumour-adjacent biopsies but not from hyperplastic or otherwise benign tissues. In addition, tumour-adjacent EPCA expression tends to occur in potentially pre-cancerous lesions, such as PIN. As such, EPCA may be useful for identifying false negative biopsies, especially those which necessarily result from random sampling of prostate tissue (63). Even extended and saturation prostate biopsy schemes can miss approximately 40% of all cancers detectible by whole-prostate analysis, so any improvement to these approaches brought about by EPCA is much-needed (64). There is evidence that higher serum EPCA levels can be predictive of cancer. Elevated serum EPCA may also be predictive of subsequent cancer development in patients with high-grade PIN, but this remains to be conclusively shown in larger and more diverse cohorts (62).

Differential display code 3 (DD3PCA3) is transcribed from a gene on chromosome 9 and believed to be a non-coding, polyadenylated mRNA (65). It was first examined through differential display analysis of a cohort of radical prostatectomy specimens and found to be overexpressed in cancerous versus non-cancerous tissues from the same patients (66). When measured in urine, DD3PCA3 expression is more specific and sensitive than PSA measurement, but its usefulness is contingent upon clinical use of RT-PCR techniques, rapid sample processing, high voided urine
volume and potentially distressing prostate massage (65). Nonetheless, DD3\textsuperscript{PCA3} has shown promise for diagnosis when considered alongside serum PSA or urinary AMACR (47).

These and other emerging biomarkers of prostate cancer show promise but require additional experimental validation before they can be applied in the clinic. The high prevalence of prostate cancer requires any potential biomarker to be highly sensitive to lethal disease before it can be used to guide treatment decisions (47). In the meantime, the search continues for novel biomarkers that are non-invasively detected and have prognostic potential.

3 ETS Transcription Factor Family

The ETS (E twenty six) transcription factor family is of interest in biomarker research because its members are variously associated with many diseases, including prostate cancer (67). Genes of the family are named after the v-ets oncogene carried by the E26 avian erythroblastosis virus with which they share great sequence homology (68). In humans, 27 members of this transcription factor family have been identified, but ETS proteins exist in a variety of animal species.

3.1 ETS and PNT Domains

All ETS proteins contain a highly conserved 85-amino acid DNA-binding domain called the ETS domain that is central to their function as transcriptional activators and repressors. The ETS domain forms a winged helix-turn-helix motif composed of a four-strand beta sheet and three alpha helices, one of which is responsible for recognizing a GGAA/T sequence known as an ETS binding site (EBS). This recognition, however, is flexible and may be modified by flanking DNA sequences or other transcription factors. DNA binding can additionally be regulated by phosphorylation or intramolecular inhibition (68,69). Recently, the DNA binding profiles for all ETS proteins was determined via a high-throughput competitive DNA binding assay. Differences in binding affinity for four nucleotides within and immediately upstream of the EBS were used to organize ETS transcription factors into four classes (70). Unlike earlier classifications (68), these latest groupings can be clearly explained by amino acid differences in the ETS domain.

Of the 27 ETS proteins, 11 also contain a pointed (PNT) domain, whose secondary structure forms five helices. The PNT domain mediates protein-protein binding and oligomerisation (69) and is at least partly regulated by a nearby MAP kinase phosphorylation site. Sequence
variability in amino acid positions on the surface of the domain additionally contribute to variations in protein interaction patterns among ETS transcription factors (71).

3.2 ETS Involvement in Disease

ETS factors participate in diverse pathophysiological processes, including proliferation, differentiation, haematopoiesis, apoptosis, metastasis, angiogenesis and transformation. Unsurprisingly, alterations in ETS activity are correlated to or causative of many diseases.

Chromosomal rearrangements are a frequent cause of ETS dysregulation that lead to disease. Such events may lead to alterations in levels of ETS protein expression or the production of chimeric protein with altered function or ectopic expression. Several paediatric leukaemias, Down syndrome and synovial sarcoma involve chromosomal translocations at well-characterised break points on chromosomes 11, 21 and X, respectively, that result in translocations of ETS genes (69).

Ewing’s sarcoma is an aggressive cancer of bone and soft tissue that disproportionately affects the young (72). It is associated with a fusion between the genes EWS and FLI1, an ETS transcription factor, that results in a chimeric product with the transcription regulatory properties of the former but the DNA-binding properties of the latter. Other ETS genes, including ETS-related gene (ERG) and ETS variant 1 (ETV1), are less frequently observed in analogous fusions, but their involvement in Ewing’s sarcoma lends credence to the theory that dysregulation of ETS target proteins is responsible for the development of this cancer (68). Though this effect is expected to depend on altered transactivation, EWS-ETS fusions can also act independently of transcription regulation, including by inhibiting mRNA splicing (69).

Similarly, some leukaemias are correlated with genetic lesions that fuse the N-terminal region of TEL and the C-terminal regions of AML1, ABL and MN1. In these disorders, the ETS domain of TEL is lost, but the protein-binding PNT motif located near the N-terminus is preserved (68). The protein-protein interactions mediated by this motif can contribute to the development of leukaemia by aberrantly activating receptor tyrosine kinases (69) or recruiting histones that repress transcription (73).

Elevation of ETS protein expression via gene amplification, increased transactivation or activating mutations can also lead to malignant disease. ETS1 overexpression, for example, is
associated with leukaemia, lymphoma and a variety of aggressive cancers. Several point mutations in SPI1, another ETS transcription factor, are associated with acute myeloid leukaemia (69).

Besides cancer, ETS proteins are involved with a host of inflammatory or autoimmune diseases. ETS1 dysregulation, for example, is associated with lupus and inflammatory bowel disease. As a component of the vascular endothelial growth factor (VEGF) angiogenic pathway, ETS1 is more highly expressed in active than remission-state ulcerative colitis and Crohn’s disease, but it is unclear whether this is a repair response mechanism or a contributor to disease (74). An allelic polymorphism 3’ of the ETS1 gene can distinguish patients suffering from two subtypes of systemic lupus erythematosus (75), but it remains to be seen whether these alleles are linked to inheritance of altered ETS1 function or expression.

3.3 ETS Involvement in Prostate Cancer

Chromosomal aberrations affecting ETS protein expression are present in 50% to 70% of prostate cancers. Phylogenetically, the ETS genes associated with prostate cancer fall into either the ERG or the ETV4 subfamilies. Though these two subsets of ETS transcription factors share little sequence homology beyond the expected ETS DNA-binding domain (67), both contain Class I proteins, as defined by DNA-binding specificity (70). It is unclear what characteristics shared by these transcription factors differentiate them from the ETS family members that are not normally implicated in prostate cancer, but possible explanations include genetic localisation in regions of genomic instability, ability to activate mitogenic RAS pathway effectors or some unidentified common mode of protein function (67). The frequent involvement of ETS genes in prostate cancer has made them popular subjects of biomarker research and led to their experimental inclusion in prostate cancer gene panels (76).

3.3.1 ERG

One particular ETS member, ERG, is frequently involved in genomic lesions characteristic of diseases such as Ewing’s sarcoma, acute myeloid leukemia and prostate cancer. The ERG gene can be alternatively spliced or polyadenylated into at least nine isoforms – the longest of these, ERG isoform 3, features both the PNT protein-binding and ETS DNA-binding domains, though truncated and non-coding ERG transcripts are also produced. In humans, ERG expression is
normally limited to endothelial and haematopoietic tissue and is virtually undetectable elsewhere, including in prostate epithelium (77-79).

Physiologically, ERG plays a role in angiogenesis and endothelial development by regulating the expression of such genes as VEGF receptor 1 and 2, vWF and endoglin (80). ERG is also necessary for self-renewal and differentiation of adult haematopoietic cells, functions in embryonic development and inhibits interleukin 8 (IL 8)-mediated inflammation (79).

Related to its normal biological functions in angiogenesis and cell renewal, ERG also participates in the direct transactivation of genes mediating phenotypic changes like improved angiogenesis and epithelial-to-mesenchymal transition (EMT), both hallmarks of cancer. Indeed, ERG overexpression on its own has been shown to be sufficient to transform mouse fibroblast cells in vitro and in vivo, allowing for serum- and anchorage-independent growth and tumour development from subcutaneous injections to mice (77).

An adequate blood supply is essential to tumour survival, so hypoxia-inducing strategies like antiangiogenesis are important approaches to cancer therapy in general (81). Vascular endothelial (VE)-cadherin is a primary component of adherens junctions in blood vessels and functions in both structural and signalling capacities. Downregulation of ERG causes adherens junction disruption and apoptosis via loss of constitutive VE-cadherin expression, leading to impaired angiogenesis in vivo (82). siRNA-mediated downregulation of ERG also decreases the expression of RhoJ, a uniquely endothelial cell-specific Rho GTPase essential for blood vessel lumen formation (83). ERG therefore plays a vital role in promoting blood vessel growth, including within the context of prostate cancer.

EMT describes the set of phenotypic changes that epithelial cells, including those of the prostate, must undergo before becoming capable of metastasis. Metastatic cells must separate from the primary tumour, enter the bloodstream, exit it at a distant site and be able to survive all the associated changes in microenvironment. As such, EMT comprises such events as loss of intercellular contacts, cell polarity and epithelial cell markers and gain of focal adhesions and mesenchymal cell markers (84). In prostate cancer, EMT develops early in the process of cancer development. Increased ERG activity causes EMT in prostate cancer cells, including loss of epithelial (E)-cadherin and gain of vimentin and cadherins 2 and 11, by several mechanisms,
including direct upregulation of Frizzled-4 (FZD4), a receptor in the Wnt signalling pathway, and ZEB1, a known inhibitor of E-cadherin (85,86).

### 3.3.2 TMPRSS2:ERG Gene Fusion

The TMPRSS2:ERG gene fusion is present in about half of all prostate cancers and is the most frequently observed genetic lesion among all human solid cancers (46). Transmembrane protease, serine 2 (TMPRSS2) is a relative of matriptase and hepsin, two membrane-bound serine proteases that promote invasion and metastasis of cancer. Like similar proteases, it is activated by autocatalytic cleavage into two fragments bridged by a disulphide bond (87,88). Its physiological roles are still under investigation but may include sodium ion channeling, cell-cell signaling or reproductive facilitation as a component of seminal prostasomes (89). In prostate cancer, it is responsible for the cleavage and activation of PAR2, a regulator of MMPs that enables prostate cancer cells to metastasise (90).

Formation of the TMPRSS2:ERG fusion gene most frequently occurs via deletion of the 2.8 Mb region on chromosome 21 that separates the TMPRSS2 and ERG genes, though chromosomal translocation is also possible (46). Androgen signalling may mediate this process by attracting topoisomerase IIβ (TOP2B), a double strand break-inducing enzyme, to androgen receptor binding sites (91). TMPRSS2:ERG fusions are genetically heterogeneous and can exist in homozygous or heterozygous forms, owing to chromosomal events that occur at different locations within the TMPRSS2 and ERG genes (92). Even within the same sample, multiple coding or non-coding transcripts may be produced (93). Moreover, alternative splicing can allow multiple protein products to be produced from the same initial transcript. Though a true fusion protein containing five amino acids from TMPRSS2 and most of the wild-type ERG protein is possible, the predominant product of TMPRSS2:ERG is a truncated ERG protein produced from exon 1 of TMPRSS2 fused to exon 4 of ERG (94).
Figure 1. General structure of the TMPRSS2:ERG gene fusion.

TMPRSS2:ERG is a gene fusion between two genes on chromosome 21. At its longest, the fusion can produce a chimeric product that incorporates a small part of TMPRSS2 and all exons of ERG. The most common product is a truncated form of ERG that retains both the PNT protein-binding and the ETS DNA-binding domains. 0 and the dashed vertical line represent transcription and translation start sites, respectively.

Regardless of exact structure, the TMPRSS2:ERG fusion causes ERG expression to become controlled by the androgen-responsive element (ARE) in the TMPRSS2 promoter, leading to the ectopic overexpression of ERG in prostate cancer in response to ubiquitous androgen receptor activity (94). In addition to the physiological and pathological effects of wild-type ERG, the truncated ERG typical of TMPRSS2:ERG fusion can cause development of PIN or progression of PIN to prostate cancer via mediating cell invasion (95,96). ERG exon 11 appears to be crucial to many of these effects, possibly due to its proximity to the PNT domain in exon 10: fusion products including the exon are associated with aggressive cancer (94). Interestingly, ERG overexpression does not necessarily arise from the presence of the TMPRSS2:ERG gene fusion. In late-stage, androgen-insensitive prostate cancer, for example, ERG overexpression may be absent, though overexpression of other ETS proteins is detected (97).

The detection of TMPRSS2:ERG in tissue, urine and serum is being actively investigated for diagnostic and prognostic benefit. Immunohistochemical detection of truncated ERG from radical prostatectomy samples can achieve over 95% sensitivity and specificity for prostate cancer diagnosis (98) and the presence of TMPRSS2:ERG in prostatectomy or biopsy samples is associated with such outcomes as higher clinical stage, worse Gleason score, metastasis and lethal cancer (97). In urine, TMPRSS2:ERG mRNA can be isothermally amplified and measured with 92% agreement with FISH analysis of biopsy cores. Urinary TMPRSS2:ERG is positively correlated with markers of tumour volume, Gleason score and presence of disease (64). Detection of TMPRSS2:ERG in serum is possible via RT-PCR of genetic material from
circulating tumour cells, but currently of uncertain clinical use. Moreover, discrepancies between the various methods of TMPRSS2:ERG detection emphasise that the heterogeneous and multifocal nature of prostate cancer render it resistant to easy categorisation (99).

The specific TMPRSS2:ERG fusion product involved or the level at which it is produced may also be correlated with such characteristics as seminal vesicle invasion and early biochemical recurrence (100). Still, there is evidence that contests the consistency of the predictive ability of TMPRSS2:ERG (101-104).

4 Next-Generation Sequencing

Next-Generation Sequencing (NGS), also known as Massively Parallel Nucleotide Sequencing, is a set of recent technologies that enables the production of terabytes of DNA sequence data in a shorter time frame and at lower cost than traditional, first-generation modes of sequencing. Several platforms, collectively considered to be of the second generation of DNA sequencing, are currently available and all are constantly improving in terms of throughput and cost. These differ in their approach to the task, but all such sequencing relies on the production of many short sequence “tags” or “reads” from a fragmented polynucleotide sample and then either aligning them along a known genome template or assembling them de novo into continuous sequences. This is enabled by groundbreaking technology that allows many miniature sequencing reactions, each seeded by one sample fragment, to proceed and be recorded in parallel. The simultaneous conduction and measurement of many sequencing reactions confers an enormous time advantage upon next-generation sequencing, rendering reasonable larger sequencing projects that were impractical by traditional methods (105,106).

4.1 SOLiD Sequencing

In Sequencing by Oligonucleotide Ligation and Detection (SOLiD), the genetic material to be sequenced is fragmented and coupled to beads. These beads allow each fragment to be clonally amplified by PCR. Beads associated with clusters of identical DNA fragments are then deposited onto a glass slide and reacted with sequencing primers, each attached to a fluorophore that encodes its sequence specificity. Through automated incubation and wash steps, primers hybridise with and are ligated to the fragmented template, and each time, the glass slide is imaged. After many cycles, the order in which fluorophores are detected from each bead can be
Gene expression can be quantified as reads per million (RPM), a measure of relative transcript abundance that takes into account the different number of total sequencing reads obtained in each sample. By literally relating the number of reads that map to a certain gene per million reads counted, RPM allows for direct comparison of the expression of the same gene from different samples. Though RPM is by definition biased towards longer genes and affected by counting efficiency (108), it still yields an approximate absolute expression level.

4.2 Applications of Next-Generation Sequencing

One important use of NGS is in the assembly of genomes without a reference template. For example, human (109), panda (110) and even Neanderthal (111) genomes have been produced in this way for downstream analyses of genetic variation, evolutionary genetics and the like. Similarly, NGS has been employed in the construction of animal mitochondrial and plant plastid genomes (112).

Classic chromatin immunoprecipitation (ChIP) experiments can also be augmented with NGS techniques. ChIP is a method of determining the DNA-binding properties of a protein by cross-linking fragmented chromatin with the protein of interest. After the protein is bound by a bead-associated specific antibody, free DNA is washed away, thereby enriching the resultant material for sequences that are bound by the protein of interest. While PCR performed with the appropriate primers can be used to determine whether specific genes are represented in the ChIP material, NGS can be used to identify essentially every fragment present. This latter method, also called ChIP-Seq, allows factor-DNA interactomes, the complete profiles of the DNA-binding affinities of proteins, to be constructed with relative ease (106,113).

Clinically, NGS is actively being investigated for such purposes as whole genome genotyping, carrier screening and heritable disease detection (114). Though its current cost precludes indiscriminate use for screening, NGS is rapidly and predictably becoming less expensive and will become increasingly attractive for use in medicine (115). Alternatively, screening can be performed via deep sequencing only of target loci for sequence variations connected to disease
Other creative applications of NGS include analysis of microRNAs and other small RNAs, chromosome rearrangements and chromatin interactions.

### 4.3 Transcriptome Analysis

Some NGS data can be treated as a representative sampling of total nucleotide content in the sample material, thereby giving information about relative abundance. RNA sequencing (RNA-Seq) is a powerful analytical technique based on this principle that is emerging as a tool for the study of the transcriptome, the collection of all RNA transcripts in a cell. Analysis of a transcriptome could take into account the relative frequency, start and end points, splicing pattern and post-transcriptional modification of transcripts. These properties decide a cell’s behaviour and can be used to characterise a cell type and physiological or pathological condition of interest.

Transcriptome analysis is a role that was previously played almost exclusively by hybridisation-dependent techniques like microarray systems. Microarray approaches to transcriptome analysis typically depend on the hybridisation of labelled, sample-derived cDNA fragments to a set of selected probes that represent genomic regions to be interrogated. The probes on a microarray could be focused narrowly on a field of interest, like prostate cancer, or cover the entire genome to the resolution of several base pairs, as is the case with genomic tiling microarrays. Major limitations of microarray techniques include the high background and low dynamic range caused by non-specific cross-hybridisation and the requirement for pre-existing knowledge of the target genome. It is also difficult and expensive to design microarrays that comprehensively include all sequences and detect all transcripts.

#### 4.3.1 Advantages of RNA Sequencing

As RNA-Seq does not depend on the use of pre-designed probes or primers, it is uniquely suited for the detection of unknown or otherwise unexpected transcripts, including those from novel genes, novel splicing events or novel polymorphisms and mutations. Compared to traditional methods based on Sanger sequencing, RNA-Seq does away with the need to insert sample DNA to intermediate plasmids. Freedom from designing a microarray also allows data from different RNA-Seq experiments to be more easily normalised and compared, an important consideration for large-scale, data-aggregating review studies.
In terms of performance, RNA-Seq can produce data of comparable quality and reproducibility to the densest gene expression-profiling microarrays and do so with no replicates (105). Moreover, RNA-Seq can be performed with minimal input RNA, again when compared to similarly comprehensive methods like genomic tiling microarrays (121).

Paired-end sequencing further extends the capabilities of RNA-Seq. Compared to Sanger-based sequencing or even single-end NGS, paired-end sequencing permits the alignment of long sequences by deriving three pieces of information from each DNA fragment sequenced: the sequence at both ends, determined by sequential reactions, and the approximate distance between them, defined by the fragment size. This ability is crucial for the generation of long, contiguous sequences (121).

### 4.3.2 Limitations of RNA Sequencing

Despite its many strengths, RNA-Seq comes with several caveats. The large quantities of data generated by RNA-Seq methods demand careful automated and manual analysis for proper interpretation and therefore remain relatively expensive in terms of computational power and human resources (120). RNA-Seq is also, like its predecessors, unable to definitively map sequence fragments that share 100% sequence homology with each other or with the template sequence, including those derived from tandem repeats and paralogous genes (121). In addition, RNA-Seq generally requires a surprisingly high sequence depth, or number of reads at each locus, in order to be effective because its output is derived from random sampling. In one illuminating example, over 50% of exon reads derived from a sample mapped to fewer than 1% of the exons in the template (105).

### 4.4 Applications of RNA Sequencing

Challenges notwithstanding, RNA-Seq provides a unique perspective from which to consider transcriptomes. In genomics, the 5’ and 3’ ends and splicing possibilities of existing genes and the locations of novel genes have been clarified through the analysis of RNA-Seq data – gene boundaries, for example, appear in the form of sudden drops in signal strengths and new genes present as strong signals mapping to unannotated regions of the genome (120).
4.4.1 RNA Sequencing in Cancer Research

When used to analyse biological specimens, RNA-Seq enables transcriptome comparisons that lead to improved characterisation of the biochemistry underlying normal and diseased conditions (120). Importantly, differences between these conditions can be used to guide the development of targeted therapies. Just as earlier DNA sequencing technologies allowed for the identification of many cancer-associated mutations and copy number changes, RNA-Seq has the potential to identify all these and other types of genomic and transcriptomic alterations at an even finer resolution (122).

In a direct comparison between microarray and RNA-Seq for analysing copy number changes in breast and lung cancer cell lines, RNA-Seq detected more gain and loss events and determined their breakpoints and the magnitudes of their copy number changes to a greater accuracy (123). A more recent study in ovarian cancer demonstrated the role of RNA-Seq in biomarker discovery by identifying a novel gene fusion of ESRRA and C11orf20 as a pathogenic biomarker of ovarian cancer (124).

RNA-Seq also performs well outside biomarker research. The sensitivity of RNA-Seq enabled the discovery, via subtractive transcriptome analysis, of Merkel cell polyomavirus, which acts in Merkel cell carcinoma analogously to human papillomavirus in cervical cancer (125). The combination of ChIP-Seq and comprehensive miRNA clarified the mechanistic involvement of SOX2 in glioblastoma multiforme (126) and CD151 in ovarian cancer metastasis (127).

4.4.2 RNA Sequencing in Prostate Cancer Research

Prostate cancer research and biomarker development are also being improved by RNA-Seq. Unsurprisingly, gene fusion research enabled by RNA-Seq is featured heavily in the literature. In an analysis of the transcriptomes of two prostate cancer and one benign prostate cell line, at least ten novel prostate-cancer associated chimeric transcripts were found in the samples. In addition to TMPRSS2:ERG, two novel ETS transcription factor-amplifying events, a MIPOL1:DGKB gene fusion and a SLC45A3-ELK4 read-through, were identified in the cancer cell lines (128). Other studies identified further novel, pathogenic gene fusions in prostate cancer either in the presence (129) or absence (130) of TMPRSS2:ERG. These fusions show promise as therapeutic targets and biomarkers and emphasise the importance of genomic instability in many prostate cancers.
Other types of prostate cancer biomarkers have also been found by RNA-Seq. Comprehensive comparison of non-coding RNAs in 102 prostate cancer samples or cell lines led to the discovery of over 100 such species that were differentially expressed between benign, localised and disseminated cancer (131). Selective RNA-Seq analysis of 218 prostate cancer tumours led to the identification of NCOA2 as an oncogene and FOXP1, RYBP and SHQ1 as potential tumour suppressors (132).

5 Rationale, Hypothesis and Objectives

The high prevalence of prostate cancer presents significant challenges to the many it affects and the healthcare systems charged with caring for them. Though many manifestations of the disease are readily treatable and a variety of therapy options exist, the great heterogeneity of prostate cancer precludes easy management. Shortcomings of current methods of prostate cancer screening, diagnosis and prognosis lead to rampant overdiagnosis and overtreatment while still missing cases of lethal disease. Overcoming these challenges requires the re-evaluation of current biomarkers and the identification of new ones.

The ETS family of transcription factors is a major focus of current prostate cancer biomarker research. Its members are responsible for myriad physiological processes and, when dysregulated, participate in the development of disease, including cancer. The TMPRSS2:ERG gene fusion leads to the overexpression of ERG, an ETS family transcription factor, and is the most frequently observed genetic lesion in human solid cancers. Its prevalence and biochemical interactions with androgen signalling have fuelled its popularity as a subject of research, and indeed, its detection in tissue and bodily fluids is being trialed for clinical use.

Next-generation sequencing technologies have been instrumental to breakthroughs in many areas of research from basic biology to medical prognostics. In the emerging field of transcriptomics, NGS outperforms first-generation sequencing and competing technologies, like microarrays. Recent applications of NGS to prostate cancer research inspired by the TMPRSS2:ERG fusion gene have found much success both in identifying novel fusion events and in elucidating the mechanisms by which they contribute to pathology.

For these reasons, we employ NGS to interrogate tumour samples that are positive and negative for TMPRSS2:ERG. We hypothesise that TMPRSS2:ERG fusion-positive and -negative tumours
have different gene expression profiles, differ in important signalling pathways and harbour different sequence polymorphisms, and that these differences are detectable by transcriptome analysis and can lead to biomarker discovery and development. Our aim is to use differential transcriptome analysis to identify single transcripts, signalling pathways and sequence variations that are associated with prostate cancer while evaluating their relationship to TMPRSS2:ERG and their possible association with disease.

Our specific objectives are as follows:

1. To identify within the entire transcriptomes of radical prostatectomy samples a subset of genes that are expressed at high levels or in correlation with TMPRSS2:ERG and survey it for ETS or prostate cancer-related genes,

2. To examine, via in vitro methods, a transcript of interest for possible regulation by TMPRSS2:ERG and contribution to cancer,

3. To organise genes along curated ontological pathways and survey them for differential expression between TMPRSS2:ERG-positive and -negative cancer; and

4. To search these transcriptomes for polymorphisms or mutations that may be associated with prostate cancer.
Chapter 2
Materials and Methods

1 RNA Extraction and Preparation

Of 139 radical prostatectomy samples that had previously been determined to differ in fusion

gene status by RT-PCR (133) and analysed by Illumina's 502-gene Human Cancer Panel (134),
two were selected for whole transcriptome analysis. Materials and information were collected in

accordance with the requirements of the Sunnybrook Research Ethics Board. Specimens were

snap-frozen in liquid nitrogen upon collection, quadrisected, and sectioned into 5-μm slices. One

section from each prostate was stained with haematoxylin and eosin and analysed at Sunnybrook

Health Sciences Centre by pathologist Dr. L. Sugar, who marked regions of tumour involvement.

RNA was extracted from the corresponding regions of separate, untreated sections using Trizol

Reagent (Invitrogen). RNA concentration and integrity were determined via Nanodrop 1000

spectrophotometry and Agilent 2100 Bioanalyzer with RNA6000 Nano chip, respectively,

according to manufacturers’ instructions. Polyadenylated mRNA was then selectively enriched

from the total RNA using the MicroPoly(A) Purist kit (Life Technologies). This enrichment was

performed in two rounds to minimize rRNA contamination in the poly(A) RNA fraction. The

absence of 18S and 28S rRNA was confirmed using the Agilent 2100 Bioanalyzer with

RNA6000 Pico kit.

2 SOLiD System Whole Transcriptome Sequencing

Whole transcriptome next-generation sequencing was performed by Dr. Y. Amemiya on the

Applied Biosystems SOLiD System (Life Technologies). Poly(A) RNA from the two samples

was fragmented using RNase III, ligated to P1 and P2 adaptors and reverse-transcribed into a

cDNA library. This library was then size-selected using the Agencourt AMPure XP reagent

(Beckman Coulter) and amplified by PCR with barcode primer using the SOLiD RNA

Barcoding Kit (Applied Biosystems). The concentrations of libraries were quantified in triplicate

using the SOLiD Library TaqMan Quantitation Kit (Applied Biosystems). Resultant barcoded

cDNA libraries were pooled together in equal concentrations, driven onto beads to generate bead

clones by emulsion PCR and then deposited on a glass slide. Paired-end (50 bp + 35 bps)

sequencing was conducted using the SOLiD 4 system.
2.1 Sequencing Data Analysis

Resultant data was filtered for colourspace quality, depleted of non-mRNA reads and mapped to the NCBI B36 human genome assembly using the Geospiza software package (www.geospiza.com). Secondary analysis normalised read counts to reads per million (RPM), identified sequence variations and matched data with Ref Seq ID. Genesifter software version 4.1 (www.genesifter.net) was used for statistical analysis in terms of relative expression, alone or organised into gene ontologies and Kyoto Encyclopedia of Genes and Genomes (KEGG)-curated pathways.

3 Plasmid Construction

A 1234-bp region immediately upstream of the RhoGDIB translation start site was amplified from VCaP genomic DNA using Phusion High-Fidelity DNA Polymerase (New England Biolabs, NEB) and cloned using XhoI and BglII restriction sites into pGL4.28 (Promega), a luciferase reporter vector, to generate pGL-RGBpro-Luc. A full-length RhoGDIB coding DNA sequence (CDS) was amplified and cloned using BamHI and EcoRI restriction sites into pCMV-3Tag 6 (Agilent), a mammalian protein expression vector with FLAG tag, to generate pCMV-RhoGDIB. An expression construct designed to produce the TMPRSS2:ERG fusion protein ΔN-ERG (95) was cloned using HindIII and XhoI restriction sites into pCMV-3Tag6, yielding pCMV-ΔNERG. The following DNA primers were used for pre-cloning amplification:

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>RhoGDIB promoter forward + XhoI</td>
<td>5’-GACTCTCGAGGTTTCAAGGCTGTATTCTG-3’</td>
</tr>
<tr>
<td>RhoGDIB promoter reverse + BglII</td>
<td>5’-GACTAGATCTTGAGTACTGCGAGTTAAGACAG-3’</td>
</tr>
<tr>
<td>RhoGDIB forward + BamHI</td>
<td>5’-GACTGGATCCATGACTGAAAAGCCCA-3’</td>
</tr>
<tr>
<td>RhoGDIB reverse + EcoRI</td>
<td>5’-GACTGAATTCTTATCATCCTGTCACCTCTT-3’</td>
</tr>
<tr>
<td>ΔN-ERG forward + HindIII</td>
<td>5’-GACTAAGCTTATGAGCCCCACGCCT-3’</td>
</tr>
<tr>
<td>ΔN-ERG reverse + XhoI</td>
<td>5’-GACTCTCGAGTTAGTAAGTGCCACATG-3’</td>
</tr>
</tbody>
</table>

pCMV-mutERG was generated from pCMV-ΔNERG by removal of the 255 bp sequence that codes for the EBS. SalI restriction sites were formed at both ends of the target sequence using PfuTurbo DNA polymerase (Stratagene) amplification of the pCMV-ΔNERG template.
The following DNA primers were used for sequential site-directed mutagenesis:

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGCAGT to GTCGAC forward</td>
<td>5’-CCGCTTTGCAAATCCAGTCGACGCCAGATCCAGCTTTTG-3’</td>
</tr>
<tr>
<td>GGCAGT to GTCGAC reverse</td>
<td>5’-CAAGCTGAGATCTGGCGTGACTGGATTTGCAAGGCCG-3’</td>
</tr>
<tr>
<td>TTCGAC to GTCGAC forward</td>
<td>5’-GCTACGCTAAGAGTCGACTTTCCAGG-3’</td>
</tr>
<tr>
<td>TTCGAC to GTCGAC reverse</td>
<td>5’-CGCTGGAGCTGAGCTTTG-3’</td>
</tr>
</tbody>
</table>

Products of the reaction were treated with DpnI (NEB) to digest the template, then ligated using the Quick Ligation Kit (NEB). The ERG constructs designed are described in Figure 2. All plasmid inserts were verified by DNA sequencing.

Figure 2. Human ERG coding DNA and corresponding protein sequences.

Shown for human ERG isoform 3 is the CDS, above in lowercase, and protein sequence, below in uppercase. The sequences in black and red correspond to ΔN-ERG, an N-terminally truncated form of the protein characteristically produced by the TMPRSS2:ERG fusion. The sequence in red corresponds to the 85-amino acid ETS domain (135); it was removed to yield mutERG by mutating the sites highlighted in yellow to 5’-GTC GAC-3’.
4 Cell Culture

HEK 293 (human embryonic kidney cells), selected for high transfection efficiency and ease of handling, were grown in Dulbecco’s modified Eagle medium (DMEM) supplemented with 10% foetal bovine serum (FBS) and 1× antibiotic-antimycotic containing penicillin, streptomycin and amphotericin B; PC3 (human prostate cancer metastasis to bone cells) do not harbour TMPRSS2:ERG (136); they and derivatives were grown in DMEM/F12 (1:1) supplemented with 10% FBS and 1× antibiotic-antimycotic. Cells were kept in humidified incubators maintained at 37°C with 5% CO₂. Cell harvest was achieved by washing cells with phosphate-buffered saline (PBS) and then treating them with 0.25% trypsin, 1mM ethylenediaminetetraacetic acid (EDTA). The above cell culture reagents were all purchased from Life Technologies.

4.1 Stable Transfection

PC3 was selected for its low endogenous expression of RhoGDIB, as determined by immunoblot with the following primary antibodies:

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Dilution</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>rabbit anti-human D4GDI (ab88317)</td>
<td>1:1000</td>
<td>Abcam</td>
</tr>
<tr>
<td>mouse anti-human β-tubulin (T5293)</td>
<td>1:1000</td>
<td>Sigma-Aldrich</td>
</tr>
</tbody>
</table>

PC3-7C were derived from PC3 (prostate bone metastasis cells) via transfection with pCMV-RhoGDIB using LipoD 293 transfection reagent (SignaGen). Stable selection was performed by culturing cells in complete medium with 300 µg/mL geneticin beginning 48 hours post-transfection. Surviving colonies were separately grown in a 96-well plate and assayed for Flag-RhoGDIB expression by immunoblot. β-tubulin expression was assessed as a loading control. The following primary antibodies were used:

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Dilution</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>rabbit anti-DYKDDDDK tag (PAB0900)</td>
<td>1:1000</td>
<td>Abnova</td>
</tr>
<tr>
<td>mouse anti-human β-tubulin (T5293)</td>
<td>1:1000</td>
<td>Sigma-Aldrich</td>
</tr>
</tbody>
</table>

4.2 Luciferase Reporter Assay

Two days before assay, 5×10⁴ HEK 293 cells were plated to each well of a 24-well plate and cultured overnight. Cells were then transfected with either 250 ng each pGL-RGBpro-Luc and
one of pCMV, pCMV-ΔNERG or pCMV-mutERG (Figure 3) using LipoD 293, as above. Successful transfection was confirmed by immunoblot using the following primary antibody:

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Dilution</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>rabbit anti-human ERG (2805-1)</td>
<td>1:1000</td>
<td>Biocare</td>
</tr>
<tr>
<td>mouse anti-human β-tubulin (T5293)</td>
<td>1:1000</td>
<td>Sigma-Aldrich</td>
</tr>
</tbody>
</table>

Luciferase activity was assayed the following day using the Luciferase Assay System (Promega) according to manufacturer’s instructions. Briefly, cells were lysed in 40 µL Passive Lysis Buffer. 20 µL of this lysate was then added to 100 µL Luciferase Assay Reagent. Luciferase activity from the reaction was measured over 10 seconds with a Lumat LB 9501 luminometer (Berthold Technologies). Differences were assessed by one-way analysis of variance (ANOVA) and Tukey’s post-hoc test using GraphPad Prism 5; P<0.05 was considered significant.

Figure 3. Plasmid constructs used in luciferase reporter assay.

To ensure adequate expression of luc2CP, an engineered luciferase perotein, pGL-RGBpro is designed such that the RhoGDIB promoter and minimal promoter both drive expression. The pCMVs are designed to produce proteins fused to a 3 × FLAG tag. The treatment conditions in this assay consisted of transfections with pGL-RGBpro-Luc and one of the three pCMVs.
4.3 Wound Closure Assay

One day before assay, 2.4×10^5 PC3 or PC3-7C cells were plated to each well of a 6-well plate and cultured overnight to approximately 80% confluence. Cells were then scratched with a 10 µL micropipette tip, rinsed once with complete medium, and cultured in 2 mL complete medium with 6 µL 500 mM hydroxyurea. Images were captured using AxioVert software (Carl Zeiss) over the subsequent 48 hours. Adobe Photoshop CS3 software was used to allow resultant images to be displayed across the full grayscale range.

4.4 ³H-thymidine Incorporation Assay

PC3 and PC3-7C cells were transiently transfected with pCMV-ΔNERG as described above and used for experiments after 48 hours. Transfection efficiency was estimated by immunoblot using the following antibody:

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Dilution</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>rabbit anti-human ERG (2805-1)</td>
<td>1:1000</td>
<td>Biocare</td>
</tr>
</tbody>
</table>

In a 96-well plate, 5×10^3 PC3 and PC3-7C cells, with or without pCMV-ΔNERG transfection, were plated in triplicate in 100 µL complete medium and cultured for 24 or 72 hours, after which time 10 µL 1µCi/mL ³H-thymidine in PBS was added to each well. After six hours, cells were harvested and scintillation counts were performed. Differences were assessed by one-way analysis of variance and Tukey’s post-hoc test using GraphPad Prism 5; P<0.05 was deemed significant.
Chapter 3
Results

1 Primary Analysis of Sequencing Data

To investigate transcriptional characteristics of TMPRSS2:ERG fusion-dependent and -independent prostate cancer, RNA from the two samples were subject to whole transcriptome sequencing using the SOLiD System. Over 50 million reads were obtained from each sample and 75.99% of these were mapped to the NCBI B36 genome assembly; approximately 40% of all reads mapped to exons and introns, while only 7.43% of reads mapped to intergenic sequences (Table 1).

<table>
<thead>
<tr>
<th>ID</th>
<th>Fusion</th>
<th>Total</th>
<th>Mapped</th>
<th>Exon/Intron</th>
<th>Intergenic</th>
<th>Genes Mapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC20</td>
<td>+</td>
<td>68 188 410</td>
<td>57 782 015 (84.75%)</td>
<td>26 064 149 (38.22%)</td>
<td>4 634 171 (6.80%)</td>
<td>22 200</td>
</tr>
<tr>
<td>PC78</td>
<td>–</td>
<td>58 157 308</td>
<td>38 229 328 (65.73%)</td>
<td>25 216 019 (43.36%)</td>
<td>4 757 784 (8.18%)</td>
<td>22 177</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>126 345 718</td>
<td>96 011 343 (75.99%)</td>
<td>51 280 168 (40.59%)</td>
<td>9 391 955 (7.43%)</td>
<td>23 187</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of sequencing data for two radical prostatectomy samples as determined by whole transcriptome sequencing.

The number of sequence reads obtained for each sample is broken down into reads that did (Mapped) and did not (not shown) map to loci on the reference genome. Within mapping reads, counts for reads that mapped to exons or introns (Exon/Intron) or intergenic regions (Intergenic) are also listed. Unmapped reads include those that were discarded due to low sequencing quality. Mapped reads not included in the categories shown included those mapping to rRNA and snRNAs. In brackets are the percentages of total reads represented by each category. The Genes Mapped column shows the number of the 26 088 in the reference database that were represented by at least one sequenced read.
These parameters are in line with other transcriptome sequencing projects (105, 137, 138). Unmapped sequencing reads are expected to consist of SNP- or mutation-rich regions, repetitive elements or novel transcript and splicing products (139). Reads mapped to intergenic regions suggest a need to identify novel transcriptionally active regions or extend the boundaries of genes, as these intergenic transcripts are likely to be near already annotated genes (140).

Between the two samples, 23,187 genes were identified with at least one read, representing 88.8% of genes in the database. After filtering out genes with fewer than 10 mapped reads, 19,625 remained for further analysis. Of these, 5,885 were significantly differentially expressed by more than twofold between samples, as determined by Genesifter software (Figure 4).

![Figure 4. Scatterplot displaying expression of transcripts for two radical prostatectomy samples as identified by whole transcriptome sequencing.](image)

All 19,625 genes mapped to by more than 10 sequencing reads are displayed. The 5,885 genes that were differentially expressed by more than twofold between the fusion-positive and fusion-negative samples are marked with red dots if more highly expressed and green dots if less highly expressed in the fusion-positive sample. Genes of particular interest and referenced elsewhere in this thesis are labeled.
2 Gene and Transcript Expression

One primary advantage of NGS is its ability to quantify RNA transcription levels more comprehensively than microarrays and in a comparable capacity to quantitative RT-PCR (120). Studying genes that are expressed at high levels in prostate cancer tissue may lead to the discovery of biomarkers that have yet to be identified or add to the body of evidence supporting the use of known or promising biomarkers. Examining genes that are differentially expressed between the fusion-positive and -negative samples may clarify the transcriptional consequences of the TMPRSS2:ERG fusion.
2.1 Expression of ETS and ETS-Targeted Genes

Many ETS family members are implicated in prostate cancer (67) and in this study, they and several target genes were represented in the group of genes that were expressed at RPM >500 or differentially expressed. The expression of ERG and TMPRSS2, the two genes involved in fusion under study, varied more than 10-fold between the samples studied. This confirms that the TMPRSS2:ERG fusion resulted in mRNA-level expression changes in these cases.

ETV1 expression was higher in the TMPRSS2:ERG fusion-negative sample while EGR1 expression was higher in the fusion-positive sample – ETV1 is occasionally implicated in prostate cancer-associated gene fusions (93) and EGR1 is a direct transactivational target of ETS1 (141). RhoGDIB, also known as ARHGDIB, is a gene whose expression was found to be associated with the presence of TMPRSS2:ERG (134). Here, RhoGDIB stood out for being expressed 14.16-fold more in the fusion-positive sample (Table 2). These findings stress the involvement of ETS pathways in prostate cancer and provide a short list of ETS-associated genes and pathways that warrant further investigation as biomarkers or therapeutic targets.

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene Name</th>
<th>Expression (RPM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERG</td>
<td>v-ets erythroblastosis virus E26 oncogene homolog</td>
<td>36.63  3.40</td>
</tr>
<tr>
<td>TMPRSS2</td>
<td>transmembrane protease, serine 2</td>
<td>158.33  1984.87</td>
</tr>
<tr>
<td>ETV1</td>
<td>ets variant 1</td>
<td>14.86  1752.10</td>
</tr>
<tr>
<td>EGR1</td>
<td>early growth response 1</td>
<td>670.24  378.64</td>
</tr>
<tr>
<td>ARHGDIB</td>
<td>RhoGDIB / Rho GDP dissociation inhibitor β</td>
<td>429.82  30.36</td>
</tr>
</tbody>
</table>

Table 2. Expression of ETS transcription factors and their targets in two radical prostatectomy samples as determined by whole transcriptome sequencing.

RPM values are derived by dividing the number of reads obtained for one gene in one sample by the total number of reads obtained in that sample, then multiplying the result by 10⁶.
2.2 Expression of Prostate Cancer-Related Transcripts

Also represented in a survey of genes that were highly expressed in either sample or both was a host of genes and transcripts reportedly associated with prostate cancer, including MALAT1, TMBIM6, CTSB and FOS, and the prostate cancer biomarkers AMACR and ACPP.

MALAT1 and AMACR were the genes most frequently detected by this sequencing in the fusion-positive and fusion-negative samples, respectively. The difference in expression of some of these genes between the two samples was dramatic: in the case of AMACR, the difference was over 100-fold (Table 3).

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene Name</th>
<th>Expression (RPM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALAT1</td>
<td>metastasis associated lung adenocarcinoma transcript 1</td>
<td>9696.37</td>
</tr>
<tr>
<td>TMBIM6</td>
<td>BI-1 / transmembrane BAX inhibitor motif containing 6</td>
<td>1993.33</td>
</tr>
<tr>
<td>CTSB</td>
<td>cathepsin B</td>
<td>787.77</td>
</tr>
<tr>
<td>FOS</td>
<td>FBJ murine osteosarcoma viral oncogene homolog</td>
<td>869.92</td>
</tr>
<tr>
<td>AMACR</td>
<td>α-methylacyl-CoA racemase</td>
<td>63.41</td>
</tr>
<tr>
<td>ACPP</td>
<td>PAP / acid phosphatase, protease</td>
<td>1630.09</td>
</tr>
</tbody>
</table>

Table 3. Expression of prostate cancer-related genes in two radical prostatectomy samples as determined by whole transcriptome sequencing.

RPM values are derived by dividing the number of reads obtained for one gene in one sample by the total number of reads obtained in that sample, then multiplying the result by $10^6$. All genes listed were amongst the 25 most highly expressed in one sample.
3 Involvement of RhoGDIB in Prostate Cancer

Through *in vitro* experimentation, expression of RhoGDIB was found to be positively regulated by the activity of PKCα and its downstream target, ETS1 (142). Microarray studies of 139 radical prostatectomy specimens in Toronto and 455 others in Sweden (143) both found that RhoGDIB expression was correlated with TMPRSS2:ERG fusion-positive status (134, Figure 5). Concordantly, the current study found that RhoGDIB was much more highly expressed in the fusion-positive sample (Table 2). These findings motivated the study of RhoGDIB in the context of prostate cancer.

Figure 5. Correlation between TMPRSS2:ERG fusion and RhoGDIB (ARHGDIB) expression in tumours obtained from a Toronto cohort.

The TMPRSS2:ERG fusion gene is correlated with upregulation and downregulation of a set of genes. RhoGDIB, also known as ARHGDIB, is among the set of genes upregulated in tumours with the fusion gene. This figure is adapted from Barwick et al. (134).
3.1 Effect of ERG on RhoGDIB Expression

ETS family transcription factors bind to the core EBS sequence, GGA/T; ERG preferentially binds to sequences resembling ACCGGAAGT (70). Multiple putative EBS of this form exist upstream of the RhoGDIB gene (Figure 6), suggesting a potential role for ERG in regulating the expression of RhoGDIB.

![Figure 6. Multiple potential EBS are upstream of the RhoGDIB transcription start site.](image)

The genomic region immediately upstream of the RhoGDIB transcription start site includes many potential EBS. Sequences resembling ACCGGAAGT or its reverse complement, ACTTCCGGT, are highlighted in black.

To investigate this possibility, pGL-RGBpro-Luc was co-transfected with either pCMV, pCMV-ΔNERG or pCMV-mutERG into HEK 293 cells (Figure 7A). ΔN-ERG is a shortened form of ERG isoform 3 and approximates the protein product of the most common TMPRSS2:ERG fusion (95). mutERG was constructed by removing the ETS domain from ΔN-ERG and therefore lacks the ability to bind to EBS in the promoter of target genes, like RhoGDIB. Co-transfection with pCMV-ΔNERG caused an increase in luciferase reporter gene activity relative to pCMV
and pCMV-mutERG (Figure 7B), demonstrating that ERG protein is sufficient and the ETS domain is necessary for mediating RhoGDIB transactivation. Expression of mutERG resulted in a small increase in RhoGDIB promoter activity that was noticeable in every repeat of the experiment, but this difference was not statistically significant.

Figure 7. ERG positively regulated RhoGDIB promoter activity.

A. Immunoblot of protein expression in HEK 293. pCMV produces no protein product while pCMV-mutERG produces a smaller product than pCMV-ΔNERG. B. Luciferase reporter gene activity from HEK 293 co-transfected with pGL-RGBpro-Luc and one of pCMV, pCMV-ΔNERG or pCMV-mutERG. ΔN-ERG expression activated the RhoGDIB promoter, but this did not occur with mutERG, which lacks the ETS DNA-binding domain. Data show mean ± SE from octuplicate measures and are representative of multiple experiments; different lowercase letters indicates significant differences by one-way ANOVA and Tukey’s post-hoc test (P<0.0001).
3.2 Effect of RhoGDIB on PC3 Phenotype

Since ERG directly transactivates the RhoGDIB gene, it is plausible that RhoGDIB mediates some cellular changes attributable to TMPRSS2:ERG, including those associated with prostate cancer. To determine the consequences of RhoGDIB overexpression, PC3 was selected for its TMPRSS2:ERG fusion-negative status and low endogenous RhoGDIB expression (Figure 8). PC3-7C, a RhoGDIB-overexpressing cell line, was derived by cloning full-length RhoGDIB into pCMV-Tag 2B and stably transfecting the recombinant vector into PC3 cells (Figure 9A).

![Figure 8. Endogenous expression of RhoGDIB protein in various prostate cell lines.](image)

As determined by immunoblot, RWPE1 endogenously expresses large amounts of RhoGDIB while VCaP expresses little. There is little or no endogenous RhoGDIB production in LNCaP, DU145 and PC3 cells.

Visual inspection of PC3-7C cultured alongside parental PC3 under identical conditions revealed more spindle-shaped cells in the former (Figure 9B). A spindle-shaped morphology is suggestive of EMT, a phenotypic change that is considered to be a critical intermediary to cancer cell invasion (144). As such, a wound healing assay was performed to further assess this change in cell behaviour. Monolayers of PC3-7C and parental PC3 cells were scratched and allowed to recover to determine their translational migration ability. Cells were cultured in the presence of hydroxyurea to eliminate differential proliferation as a confounding variable. PC3 cells required over 30 hours while PC3-7C cells required about 24 hours to close gaps of comparable width (Figure 9C), suggesting that RhoGDIB expression contributes positively to cell migration.
Figure 9. Effects of RhoGDIB overexpression on PC3 cell morphology and motility.

A, Immunoblot of stable overexpression of Flag-tagged RhoGDIB in PC3-7C. B, PC3-7C and parental PC3 were visually assessed in cell culture. PC3-7C grew with a more pronounced, spindle-shaped phenotype while PC3 formed a cobblestone monolayer more typical of epithelial cells. C, PC3-7C and parental PC3 were grown nearly to confluence, scratched with a micropipette tip and cultured in the presence of 1.5 mM hydroxyurea to prevent proliferation; wounds were imaged at 0h, 6h, 24h and 30h. After approximately 24 h, PC3-7C had nearly bridged the gap while parental PC3 had not.
Another important contributor to cancer development is irregular growth, whether from increased proliferation signalling or unresponsiveness to death cues. In order to determine whether RhoGDI\textsubscript{B} affects cell proliferation, a $^3$H-thymidine incorporation assay was performed. $^3$H-thymidine is incorporated into newly replicated DNA, so its presence is a surrogate marker for mitosis. Cultured PC3-7C and parental PC3 were exposed to $^3$H-thymidine for 6 hours. Subsequent results showed that PC3-7C cells incorporated much more of the radioactive label and that this effect was amplified by transient transfection with ΔN-ERG (Figures 10, 11).

<table>
<thead>
<tr>
<th>Days Post-Transfection</th>
<th>1</th>
<th>4</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC3 + ΔN-ERG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC3-7C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC3-7C + ΔN-ERG</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Figure 10. Transient overexpression of ΔN-ERG in PC3 and PC3-7C.

PC3 and PC3-7C were transiently transfected with pCMV-ΔNERG. Subpopulations of the same cultures used for $^3$H-thymidine assays were analysed by immunoblot one, four and seven days post-transfection. ΔN-ERG was strongly and specifically expressed in the desired cell lines over the course of all experiments.
Figure 11. Effect of RhoGDIB overexpression on PC3 cell proliferation.

PC3-7C and parental PC3, with or without ΔN-ERG overexpression, were incubated with $^3$H-thymidine over six hours. PC3-7C overexpressing ΔN-ERG proliferated more rapidly than PC3-7C, which in turn proliferated more rapidly than PC3 in either condition. Data show mean ± SE from triplicate measures and are representative of multiple experiments; different lowercase letters indicate significant differences by one-way ANOVA and Tukey’s test ($P<0.001$).

4 Wnt Signalling and TMPRSS2:ERG+ Prostate Cancer

KEGG-curated pathway maps encompass a broad variety of cellular systems and are useful for the interpretation of large quantities of data, like those derived from RNA-Seq (145). With this purpose in mind, Genesifter software was used to identify KEGG molecular interaction and reaction pathways in which many genes were differentially expressed between the fusion-positive and fusion-negative samples. In the fusion-positive sample, this led to the discovery of a gene expression profile consistent with increased Wnt signalling (Figure 12). Wnt is an important component of growth and development control, and upregulation of Wnt signalling is associated with prostate cancer (146). Genes throughout the signalling cascade were found to be more highly expressed in the fusion-positive sample, with lymphoid enhancer-binding factor 1 (LEF1) being the most striking of these (35.18-fold higher). In concert with these differences, phosphoinositide-3-kinase, regulatory subunit 5 (PIK3R5) of the Wnt-enhancing PI3K pathway were expressed at higher levels while adenomatous polyposis coli (APC) itself was expressed at a lower level. In contrast, the expression of several Frizzled family receptors was different between the two samples, but not necessarily in positive correlation with each other.
Figure 12. Expression levels of Wnt signalling pathway genes in the fusion-positive sample relative to the fusion-negative sample.

Higher Wnt and PIK3R5 and lower APC expression is consistent with decreased GSK-3β activity. This leads to decreased β-catenin degradation, accumulation of β-catenin in the nucleus and finally increased expression of Wnt pathway targets such as TCF7 and LEF1, important effectors contributing to proliferation and evasion of apoptosis.
5 p53 Variant and TMPRSS2:ERG– Prostate Cancer

Analysis of the sequencing data revealed the presence of a well-known single-nucleotide polymorphism (SNP) of p53, rs1042522, in the fusion-negative sample only. This SNP represents a C to G substitution that results in arginine instead of proline at residue 72 of the p53 protein (Figure 11) and may be associated with prostate cancer (147).

Figure 13. Amino acid substitution attributed to rs1042522.

The SNP rs1042522, a C to G substitution in the TP53 gene, changes proline to arginine in the sequence of p53 as indicated. Proline is (exclusively) encoded by all codons CCX and arginine is encoded by all codons CGX (among others), where X is any nucleotide.

Interestingly, a large number of p53 target genes were found to be less strongly expressed in the fusion-negative sample (Table 4), including 14-3-3σ (7.31-fold lower) and IGFBP3 (6.82-fold lower). IGF1, a signaling molecule inhibited by IGFBP3, was expressed 2.68-fold higher. Notably, these differences in p53 effector gene expression occurred in the absence of a corresponding disparity in p53 expression.
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene Name</th>
<th>Function</th>
<th>Expression</th>
<th>TMPRSS2:ERG+</th>
<th>TMPRSS2:ERG−</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP53</td>
<td>p53 / tumor protein p53</td>
<td>Mediates transactivation in response to cell stress</td>
<td>72P variant</td>
<td>72P variant</td>
<td>72R variant</td>
</tr>
<tr>
<td>CDKN1A</td>
<td>cyclin-dependent kinase inhibitor 1A</td>
<td>Causes cell cycle arrest in G2</td>
<td>57.38</td>
<td>24.76</td>
<td></td>
</tr>
<tr>
<td>GADD45A</td>
<td>growth arrest and DNA-damage-inducible, α</td>
<td>Causes cell cycle arrest in G1</td>
<td>25.52</td>
<td>75.56</td>
<td></td>
</tr>
<tr>
<td>GADD45B</td>
<td>growth arrest and DNA-damage-inducible, β</td>
<td>Causes cell cycle arrest in G2</td>
<td>52.36</td>
<td>13.89</td>
<td></td>
</tr>
<tr>
<td>SFN</td>
<td>14-3-3σ / stratifin</td>
<td>Causes cell cycle arrest in G2</td>
<td>72.46</td>
<td>10.09</td>
<td></td>
</tr>
<tr>
<td>FAS</td>
<td>Fas</td>
<td>Activates caspase-8 upon binding by Fas ligand</td>
<td>32.95</td>
<td>14.38</td>
<td></td>
</tr>
<tr>
<td>CASP8</td>
<td>caspase 8</td>
<td>Proenzyme; Cleaves Bid</td>
<td>30.15</td>
<td>8.81</td>
<td></td>
</tr>
<tr>
<td>BID</td>
<td>BH3 interacting domain death agonist</td>
<td>Triggers cytochrome c leakage and apoptosis when cleaved</td>
<td>47.10</td>
<td>19.22</td>
<td></td>
</tr>
<tr>
<td>IGFBP3</td>
<td>insulin-like growth factor binding protein 3</td>
<td>Major IGFBP in serum; Sequesters and prolongs half-life of IGFs</td>
<td>199.28</td>
<td>29.74</td>
<td></td>
</tr>
<tr>
<td>IGF1</td>
<td>insulin-like growth factor 1</td>
<td>Induces proliferation</td>
<td>39.63</td>
<td>108.18</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Expression of p53-regulated genes as determined by whole transcriptome sequencing.

RPM values are derived by dividing the number of reads obtained for one gene in one sample by the total number of reads obtained in that sample, then multiplying the result by $10^6$. The p53 variant associated with each sample is listed; in the general population, 72P is more common at that locus. All genes listed were differentially expressed between the TMPRSS2:ERG-positive and –negative samples by at least twofold.
Chapter 4
Discussion

1 Utility of RNA-Seq in Prostate Cancer Research

In this research, we took advantage of the comprehensive nature of whole transcriptome sequencing to examine the expression of established prostate cancer-related genes and to form novel associations between the disease and oncogenes/tumour suppressor genes. One conclusion that can be drawn from primary analysis of the data is that RNA-Seq is, indeed, a powerful method for measuring transcript expression. Measured transcript expression spanned beyond 0.1 and 10 000 RPM, demonstrating the dynamic range of the technique. Over 85% of all annotated genes were detected with at least one mapped read. Compared to microarrays, which have a typical dynamic range below three orders or magnitude and are insensitive to genes expressed at extreme levels (120), RNA-Seq clearly delivers more opportunities for investigation of resultant data.

To wit, our results provide insights into both TMPRSS2:ERG fusion-associated and -unrelated prostate cancers, identifying key genes, pathways and polymorphisms. These findings will be discussed in the following section.

1.1 Detection of Prostate Cancer-Related Transcripts

A subset of genes that have been previously associated with prostate cancer was determined to be highly expressed in the prostate samples by our RNA-Seq analysis: MALAT1, BI-1, CTSB and FOS. This finding adds to the body of research indicating the involvement of these genes in prostate cancer, points to usefulness of these products as biomarkers, given the high levels at which they were expressed, and additionally indicate likely fruitful avenues of future study. The expression of many of these genes differed between the two samples by more than twofold. If the TMPRSS2:ERG gene fusion indeed underlies this difference, they may be investigated as potential upstream or downstream signalling partners with the TMPRSS2:ERG gene fusion.

MALAT1 is a non-coding RNA first identified through first-generation sequencing of metastatic and non-metastatic lung cancer. It is well-conserved among mammals. Its location on chromosome 11 is associated with pathological genomic lesions, including a translocation that
causes primary renal cancer. MALAT1 is associated with metastasis and poor survival in lung
cancer and is elevated in over 50% of colon, pancreas and breast cancers as well. It can be
detected in tissues using RNA antisense probes and has been shown in one study to be
significantly overexpressed in approximately 25% of prostate cancer samples. Remarkably, in
every instance MALAT1 was more highly expressed in a tumour sample than in the matched
normal counterpart (148,149). We found that MALAT1 was the most plentiful transcript in the
fusion-positive sample and present about one-third as much in the fusion-negative sample (Table
3). This is consistent with the previously reported easy detection and heterogeneous expression
of MALAT1 in prostate cancer.

Bax inhibitor-1 (BI-1) is a potent suppressor of physiologically and pharmaceutically mediated
apoptosis: besides inhibiting its namesake, Bax, it also prevents apoptosis in response to serum
deprivation, interleukin 3 (IL3)-depletion and treatment with chemotherapeutic agents (150). In
human prostate tumours, BI-1 is upregulated at least at the mRNA level and knockdown of BI-1
is sufficient for inducing apoptosis in a variety of prostate cancer cell lines (151). Our RNA-Seq
results showed that BI-1 was very highly expressed in both samples (Table 3), indicating that it
may contribute to prostate cancer development independently of fusion gene status. BI-1 also
shows promise as a biomarker, given its high expression, but it has so far proven difficult to
detect by immunohistochemistry in prostate tissue (151).

Cathepsin B (CTSB) is a cysteine protease implicated in intracellular signalling and ECM
degradation, both by cleaving and activating intermediaries like MMPs or by direct proteolysis
of fibronectin and other matrix proteins. Consistent with these activities, it is also associated with
metastasis in a wide array of solid tumours, including those of the prostate. Inhibition of CTSB
via siRNA transfection decreases proliferation, migration, invasion and angiogenesis of prostate
cancer cells and can induce Bax- and caspase-mediated apoptosis (152). In tissue, CTSB is
overexpressed in prostate cancer versus normal controls, but neither its expression nor the
expression of its inhibitor, stefin A, correlate with Gleason score, despite its theoretical role in
cancer progression (153). Here, we found that CTSB was highly expressed in both samples, but
moreso in the fusion-positive case (Table 3). Given the above, the relationship between CTSB
and TMPRSS2:ERG and its potential as a prognostic marker of metastasis merit further
investigation.
FOS forms part of the activator protein 1 (AP1) transcription factor in heterodimers with Jun proteins. AP1 is modulated by and modulates a wide range of cellular pathways, including those involved in cell differentiation, proliferation, apoptosis and transformation (154). Expression of AP1 constituent proteins has been associated with aggressiveness of disease in diverse cancers. In prostate cancer, tissue expression of FOS and JUN is positively correlated with disease progression and AP1 activity is increased in androgen insensitive cancer. AP1 may be particularly relevant in prostate cancer because it is important for the regulation of nuclear hormone receptors, including AR, and may be regulated by AR in return (155,156). In this study, FOS was, like MALAT1 and CTSB, expressed at a higher level in the fusion-positive sample (Table 3). Its known interactions with AR make it even more attractive as a biomarker or therapeutic target in prostate cancer.

1.2 Detection of Prostate Cancer Biomarkers

Also highly expressed were two prostate cancer biomarkers that are of current interest in the field: AMACR and PAP. AMACR was first identified as overexpressed in prostate cancer through DNA microarray analysis and found to be very sensitive and specific at detecting prostate cancer in prostate needle biopsies (59). It has since found important applications, when combined with other markers, in detecting small cancerous lesions (61) or deciding histologically ambiguous cases (60), but its uses are limited to post-screening biopsy analysis. We found that AMACR was the most highly expressed gene in the TMPRSS2:ERG-negative sample but was unremarkable in the TMPRSS2:ERG2-positive case (Table 3). This discrepancy highlights the heterogeneity of prostate cancer and the difficulty of assigning thresholds above which a biomarker may be interpreted as indicative of disease. Given the reported sensitivity of AMACR staining (157), this finding may also be an artefact of small sample size.

PAP was the first serum biomarker for prostate cancer. Screening for prostate cancer by PAP has been obsolete since the discovery of PSA (158), but serum levels of PAP are still under consideration for non-diagnostic uses. As a case in point, serum PAP has since been found to be predictive of biochemical recurrence following radical prostatectomy (159). This coincides with the present finding of increased PAP expression in the fusion-negative sample (Table 3), which was derived from a patient who suffered biochemical recurrence, and suggests that detection of PAP from biopsied tissue may also be a viable means of predicting post-operative outcome.
2 Significance of ETS Involvement in Prostate Cancer

Components of the ETS and TMPRSS2:ERG gene pathways were found to be dysregulated in this research. First, TMPRSS2 expression was dramatically lower in the fusion-positive sample, an expected consequence of the TMPRSS2 gene being either deleted or translocated from its promoter during the process of fusion gene formation (97,160); accordingly, ERG expression was higher in that sample (Table 2). This confirms that the instance of TMPRSS2:ERG under study has appreciable transcriptional consequences and allows the fusion gene to be considered a mediator of other differences observed between the two samples.

2.1 ETV1

Ets variant 1 (ETV1), an ETS gene, was much more highly expressed in the fusion-negative sample (Table 2). ETV1 fulfills physiological roles in development and self-renewal and can be fused to EWS in Ewing’s sarcoma. Furthermore, its full-length, truncated and fusion transcripts have all been associated with prostate cancer. Indeed, ETV1 overexpression can be effected by any number of fusions or no fusion at all. ETV1 is, after ERG, the second-most frequently overexpressed gene in prostate cancer, but TMPRSS2:ETV1 accounts only for a minority of these. In most known cases, ETV1 fusions result in an N-terminally truncated ETV1 protein (161), not unlike the case with ERG, but novel ETV1 fusion partners are regularly being identified (93). A previous microarray study found that ERG and ETV1 overexpression occurred in approximately half of all prostate tumours, but never both in the same sample (162); later studies discovered it was possible for ERG and ETV1 overexpression to simultaneously exist in separate tumour foci from one patient (163). The pattern of TMPRSS2, ERG, and ETV1 expression observed here is consistent with these findings and suggestive of ETS involvement even in the TMPRSS2:ERG-negative sample, possibly in the form of a different gene fusion.

2.2 EGR1

One ETS-targeted gene, early growth response 1 (EGR1), was highly expressed in both samples, though slightly more so in the fusion-positive case (Table 2). EGR1 is a transcription factor that participates in oncogenesis through its action on protooncogenes, mitogens and their receptors (141), but may play the reverse role and inhibit growth (164) or mediate radiation-induced apoptosis (165), depending on the context. It enables soft agar growth and Matrigel invasion in
prostate cancer cells via positively regulating the expression of interleukin 8 (IL8) in cooperation with NFκB. This effect can be reversed by siRNA-mediated inhibition of EGR1 and has obvious implications for therapy (166).

As a biomarker, EGR1 protein was found to be frequently overexpressed in prostate cancer primary tumours and expression was particularly intense in cancerous as compared to adjacent areas (165). In biopsies, EGR1 protein expression is restricted to cancerous regions (167). EGR1 mRNA levels are higher in prostate cancer versus normal tissue, especially in samples with Gleason scores >8 (168). Furthermore, overexpression of EGR1 also characterises a subset of prostate cancer cases that is more likely feature recurrence following radiation therapy (165).

In addition, at least ETS1 and FLI-1 were previously shown to bind to EBS within the EGR1 promoter region and promote its transcription (141). Given the similarities between ETS1, FLI-1 and ERG in terms of protein domains and DNA binding specificity profiles (70), the present findings suggest that ERG also participates in EGR1-associated cell survival and invasion. This also supports the consideration of both EGR1 and the TMPRSS2:ERG fusion gene as candidates for inhibition in the diagnosis, prognosis and management of prostate cancer.

### 2.3 RhoGDIB

Expression of RhoGDIB has been correlated with ETS1 expression (142), TMPRSS2:ERG fusion-positive status (134) and, in this study, ERG overexpression, making it an attractive subject of biomarker investigation. RhoGDIB is a GDP dissociation inhibitor normally restricted to haematopoietic tissues and lymphocytes. True to its name, it functions primarily as a negative regulator of Rho-family GTPase signalling, either by inhibiting the latter’s exchange of GDP for GTP or by sequestering it when GTP-bound (Figure 12), though RhoGDIB may also play a positive role in GTPase signalling by ensuring appropriate targeting and timing of downstream activation (169). Specific targets of RhoGDIB include RhoA, Cdc24 and Rac1. These GTPases are components of many cellular pathways and can promote proliferation, survival, angiogenesis and metastasis when aberrantly activated (170).
Figure 14. Mechanisms of action of RhoGDIB.

Rho GTPases are molecular switches that activate Rho proteins when they are bound to GTP. A, Guanine nucleotide exchange factors (GEFs) catalyse the exchange of GDP for GTP, thereby activating GTPases. RhoGDIB can inhibit the activity of GEFs. B, RhoGDIB can bind to and sequester activated GTPases, preventing their interaction with and activation of downstream targets.

The balance between the activity of RhoA, Cdc24 and Rac1 ultimately determines cytoskeletal arrangement and related outcomes, such as invasion ability (171), and indeed, RhoGDIB expression has been paradoxically shown to be stimulatory of breast cancer (172) and inhibitory of lung cancer (173) metastasis, though even this is uncertain (174). RhoGDIB expression has also been extensively studied in the context of cancer progression, but its involvements herein similarly diverge in a context-dependent manner. For example, RhoGDIB is positively correlated with increasing gastric cancer stage, on the one hand, but better survival in bladder cancer patients on the other (170). The activity of RhoGDIB in the prostate has not been as well characterised, but a recent tissue microarray study showed that tissue expression of RhoGDIB may be associated with colon and prostate cancer metastasis (175).

RhoGDIB has also been linked to ETS family transcription factors in breast cancer. Specifically, ETS1 directly binds to the promoter region of RhoGDIB and influences its expression. Moreover, ETS1 and RhoGDIB are frequently coexpressed at the mRNA level in breast cancer.
cell lines and primary tumours (174). Since ETS1 is closely related to ERG (70), our observation of markedly increased RhoGDIB expression in the TMPRSS2:ERG sample (Table 2) and direct ERG involvement in RhoGDIB transactivation (Figure 6) is consistent with these findings. This, in turn, suggests that the detection of ETS transcription factors and RhoGDIB should be evaluated for use as biomarkers of this cancer.

The slight but consistently observed increase in RhoGDIB transactivation effected by mutERG overexpression may be due to residual activity from two autonomous transactivation domains present in the ERG transcription factor (176). These are able to activate transcription of genes without binding to EBS and may play a minor role in the regulation of RhoGDIB expression, likely in concert with coactivators.

The present study subsequently found that PC3-7C, a derivative of PC3 stably overexpressing RhoGDIB, have a spindle-shaped cell morphology (Figure 9B). This morphological change is suggestive of cytoskeletal rearrangement and increased cell motility (144) and accordingly, RhoGDIB overexpression also increased the rate of translational cell migration as determined by wound healing assays (Figure 9C). Furthermore, according to \(^{3}\text{H}\)-thymidine incorporation assays, PC3-7C proliferate more readily than parental PC3, especially with concomitant ΔN-ERG overexpression (Figure 11). Taken together, these results point to specific contributions by RhoGDIB in the development and progression of prostate cancer, particularly under the influence of TMPRSS2:ERG gene fusion.

### 3 Significance of Wnt Signalling in Prostate Cancer

Organising the transcriptome data along KEGG pathways uncovered patterns that were indicative of dysregulation in the Wnt signalling cascade. This pathway is evolutionarily conserved and has diverse physiological roles, though perhaps with an emphasis on stem cell fate and self-renewal functions. In the prostate, for instance, Wnt signalling is necessary for embryonic development and adult organ maintenance through maintaining epithelial progenitor cells (146).

In the canonical Wnt pathway, Wnt ligand binds to the cell-surface receptor Frizzled, transmitting a signal to Axin via Dishevelled. This signal inhibits the formation of a complex between glycogen synthase kinase-3β (GSK-3β), APC and Axin, which are together otherwise
responsible for β-catenin degradation. In the absence of this complex, β-catenin accumulates in the cytoplasm and translocates to the nucleus. There, it activates downstream effectors via interacting with the TCF/LEF family of proteins. (177).

As their names indicate, these proteins were originally studied in lymphocytes and play important roles in their differentiation and proliferation (178). Though technically not transcription factors due to their complete dependence on the binding of coactivators, their primary role is to activate target genes by binding to C/TCTTTGA/TAT, a conserved sequence known as the Wnt response element (WRE). When β-catenin binds to TCF/LEF, it displaces transcriptional repressors, like Groucho, and recruits various activating cofactors, enabling TCF/LEF function (179).

Effectors governed by TCF/LEF include c-myc, cyclin D1 and survivin and are responsible for such outcomes as proliferation and apoptosis evasion (145,180), so dysregulation of this pathway is a feature of many cancers (177). Indeed, Wnt pathway components like Wnt1 and β-catenin are overexpressed in prostate cancer cell lines, aggressive prostate cancer tissue and in lymph node and bone metastases (146). By virtue of their contributions to this pathway, the increased expression of several Wnt proteins, FZD10, TCF7 and LEF1 and decreased APC expression observed in the fusion-positive sample (Figure 12) are all generally suggestive of a contributory role for Wnt signalling in TMPRSS2:ERG-related prostate cancer. Importantly, however, not all members of these protein families were differentially expressed between the samples studied – of the dozen or so human WNT genes, only four were overexpressed in the TMPRSS2:ERG case. As such, additional efforts must be made to identify the differences between these genes that underlie their specific behaviour.

3.1 PI3K Pathway

The PI3K pathway mediates tumour cell survival, proliferation and migration and has important roles in immune and inflammatory responses. Aberrations somewhere along the PI3K pathway are a feature of all human cancers and in fact, after Ras, mutations in PI3K itself are the most prevalent in cancer. Mechanistically, the primary action of PI3K is to phosphorylate phosphatidylinositol to generate lipids like phosphatidylinositol-3-trisphosphate (PI3P), which act as plasma membrane docking sites for downstream proteins. The protein kinase B (PKB)
family of proteins are bound and activated by PI3P and are some of the most important effectors of PI3K signalling. Like PI3K, PKBs are frequently amplified or mutated in cancer (181,182).

PKB participates in the positive regulation of Wnt signalling by phosphorylating and inhibiting GSK-3β, similar to the action of Dishevelled, though this has been reported to induce or inhibit cell migration and invasion in different experimental models (182). Regardless, the increased expression in the fusion-positive sample of PIK3R5 (Figure 10), a regulatory PI3K pathway gene, lends additional credibility to Wnt signalling as a mediator of fusion gene-associated prostate cancer.

3.2 AR

Clinically, androgen deprivation often leads to at least temporary prostate cancer remission (183) and in the laboratory, androgens are necessary for tumour induction. AR, when bound by an androgen, is a nuclear transcription factor and translates the presence of androgens into cellular outcomes by activating genes that contain an ARE. Accordingly, amplification and constitutive activation of AR is often a feature of hormone-refractory prostate cancer (184,185).

Wnt signalling can enhance AR activity on multiple levels. For one, β-catenin, whose protein level is controlled by canonical Wnt signalling, directly binds to and acts on AR and increases AR-mediated signalling by several mechanisms: in a cancer cell line model, overexpression of β-catenin has been shown to increase the responsiveness of AR to less potent androgens, like androstenedione; relax the ligand specificity of AR such that it responds to estrogen stimulation; and decrease the sensitivity of AR to a direct antiandrogen, bicalutamide (185). The effects of β-catenin on AR are additionally amplified with E-cadherin loss, such as that which occurs in metastatic prostate cancer (186).

For another, TCF/LEF directly transactivates AR, and GSK3β may also influence AR signalling (146). Given that overexpression of ERG in TMPRSS2:ERG-positive tumours is caused by the placement of ERG under androgen-sensitive transcription (94), the upregulation of Wnt pathway components observed in this study alludes to a potential positive feedback loop between gene fusion, Wnt pathway upregulation, increased β-catenin levels and AR modulation (Figure 15).
Overexpression of genes in the Wnt pathway mimics Wnt signalling, leading to β-catenin stabilisation. β-catenin binds to AR and increases its responsiveness to androgens, decreases its responsiveness to inhibitors and relaxes its substrate specificity, thus enhancing AR-mediated transactivation. This, in turn, induces the expression of ERG from the TMPRSS2:ERG fusion gene and feeds back into the Wnt pathway.

3.3 Other Components of Wnt Signalling

Aberrations in APC expression alone may have a strong influence on prostate cancer development. It has been shown in a murine model of prostate cancer that Cre-mediated
conditional loss of APC in prostatic epithelium leads to accumulation of cytoplasmic and nuclear β-catenin and, ultimately, androgen-insensitive prostate cancer (187). Our RNA-Seq data showed that APC expression was decreased in the fusion-positive sample (Figure 10), which is consistent with a role for APC in the development of TMPRSS2:ERG-mediated prostate cancer in either Wnt-dependent or -independent manners.

The increase in expression of FZD10 is consistent with Wnt signalling upregulation, but the decrease in expression of FZD3, FZD4 and FZD8 are not. It is possible that this loss of receptor expression is indicative of negative feedback regulation in the Wnt pathway, but this type of regulation has only previously been reported to occur via lysosomal degradation of Dishevelled protein, not repression of Frizzled transcription (188). In medulloblastoma tissues and cell lines, however, Frizzled receptor expression has been found to be inconsistently up- and down-regulated (189), so perhaps in prostate cancer, too, there is a similar need to clarify the significance of individual Frizzled receptors.

4 Significance of p53 Variant in Prostate Cancer

KEGG-assisted analysis drew attention to lower expression of many p53 target genes, particularly those involved in cell cycle arrest and apoptosis, in the fusion-negative sample (Table 4). The transcription factor p53 is credited with activating DNA repair, promoting senescence and initiating apoptosis upon detection of DNA damage and is therefore a crucial tumour suppressor. The high rate of turnover of p53 proteins belies the importance of its precise regulation in these functions. True to form, p53 is lost or disabled in virtually all human cancers and p53 expression can reverse tumour progression (190). Decreased expression of p53 targets is therefore consistent with a disease state that is less responsive to anti-proliferative signals.

In its inactive state, p53 is inhibited via proteasomal degradation mediated by murine double minute 2 (MDM2) and transcriptional repression by MDM4. Ataxia-telangiectasia mutated (ATM) and Rad3-related (ATR) respond to genomic double strand breaks and other sources of genomic stress, respectively, by catalyzing the efficient degradation of MDM2 and 4. The additional involvement of a host of recruited effectors allows ATM and ATR to fine-tune p53 and downstream responses to suit the nature and severity of genome damage. Incidentally, oncogene activity is also known to induce p53 activity, likely in a reactionary defense capacity, and these functions may be enabled by ATM and ATR or another intermediary, ARF (190).
Once freed from inhibition by MDM2 and MDM4, p53 binds to p53-responsive elements (p53-REs) to induce the transactivation of a wide variety of genes with functions in cell cycle arrest, apoptosis induction, angiogenesis inhibition and senescence activation (145,180).

4.1 Cell Cycle Control Genes

A remarkable proportion of the p53-controlled cell cycle genes differentially expressed in this study are specifically involved in pharmaceutically induced arrest in prostate cancer. The decreases in expression of these genes in the fusion-negative sample point to their potential use as predictive markers of p53-based therapy, but their concerted loss additionally points to p53 as an overarching governor of TMPRSS2:ERG-independent prostate cancer.

p21 is a cell cycle control gene that participates in stress signalling by initiating cell cycle arrest in G1 or G2. The promoter region of p21 contains two p53-REs that are important to its activation, but other stimulators of p21 expression include ERK and TGFβ. p21 has been shown to mediate arrest in G2 and consequent growth inhibition in prostate cancer cells in response to treatment with G-1, a potential anti-cancer agent. This action of p21 broadly suppresses G2 checkpoint regulators and is therefore a powerful component of the p53 response (191,192). Decreased expression of p21 in the fusion-negative sample is expected to have a correspondingly dire effect on prostate cancer progression and recovery.

Similarly, the GADD family of proteins are key motivators of cell cycle arrest as a result of cell stress. GADDs are direct targets of p53 signalling and in murine models, they have been shown to mediate DNA repair and apoptosis. In several human cancers, GADD genes are often repressed by promoter methylation, but contradictory roles for GADD overexpression have been reported, as is the case for pancreatic cancer (193). One particular member of the GADD family, GADD45A, causes G1 arrest and is responsible for the antitumor activity of fucoxanthin in liver and prostate cells. (194). Loss of GADD45A and GADD45B expression in the TMPRSS2:ERG-negative sample (Table 4), then, may frustrate tumour suppression mediated by p53 or drugs like fucoxanthin.

14-3-3σ, also known as stratifin, is an epithelium-restricted member of the evolutionarily conserved 14-3-3 protein family. Upon activation by p53, 14-3-3σ inhibits cell cycle progression beyond G2 in order to prevent the propagation of genomic damage (195). It has been shown to
be expressed at successively lower levels in normal prostate, PIN and adenocarcinoma. Interestingly, loss of 14-3-3σ expression is primarily caused by promoter hypermethylation and DNA demethylation drugs like 5-aza-2’-deoxycytidine can reverse this loss (196). In terms of chemotherapeutics, 14-3-3σ expression is decreased in gefitinib-resistant prostate cancer cell lines following gefitinib treatment, so monitoring of 14-3-3σ levels before and after gefinitib therapy may yield a predictive marker of treatment effectiveness (197). This can demonstrably be achieved via RNA-Seq methods.

4.2 Apoptosis-Inducing Genes

Several apoptosis-inducing effectors of p53 signalling were also less highly expressed in the fusion-negative sample, again with disproportionate representation from genes mediating drug action in prostate cancer. Programmed cell death is an important barrier to the uncontrolled proliferation of DNA-damaged cells characteristic of cancer. Decreased expression of apoptosis-associated genes has implications for disease and again points to an upstream disruption of p53 activity.

Fas, caspase-8 and Bid positively regulate the caspase cascade and induce apoptosis. Fas is a member of the tumour necrosis factor (TNF) superfamily that mediates cell death in conjunction with Bid and caspases in a wide variety of tissue-destroying diseases, such as multiple sclerosis (198). Fas-pathway apoptosis has been found to be essential to the anti-cancer action of 2-Methoxyestradiol through in vitro and in vivo models of prostate cancer (199). Though only Fas itself is a direct target of p53 signalling (200), the accompanying decreases of caspase-8 and Bid expression in the fusion-negative sample (Table 4) point to significant impairment of Fas-mediated apoptosis.

IGFBPs are generally responsible for sequestering IGFs and preventing their activation of target receptors, leading to an attenuation of growth signals. IGFBP3 is a p53-regulated member of this family that can perform the above functions and activate apoptosis pathways as well (201); this latter function it can perform independently, by activating the caspase cascade (159), or in concert with its IGF1-sequestering effect. In prostate cancer, dependency on IGF1 may replace androgen-dependence in hormone refractory cancer. Applying 5-fluorouracil (5-FU) to prostate cancer cells causes an increase in IGFBP3 expression, especially in androgen-independent cell lines, and inhibition of IGFBP3 decreases the effectiveness of 5-FU treatment (202). Our RNA-
Seq data showed that IGFBP3 was less and IGF1 was more highly expressed in the fusion-negative sample; this is consistent with the above and with the recent finding that this pattern of IGFBP3 and IGF1 dysregulation may be associated with advanced prostate cancer (203), though this assertion is controversial (204). Regardless, the IGF axis of prostate cancer regulation deserves further consideration as a substitute for androgen response.

4.3 p53 Polymorphism

The observed decreases in p53 target gene expression detailed above together imply a loss of p53 signalling, yet p53 expression was not significantly different between the fusion-positive and -negative samples. Transcriptome sequencing, however, revealed that the missense SNP rs1042522, encoding a P72R variant of p53, was present in the fusion-negative sample. Though automated phylogenetic and structural analysis by PolyPhen-2 predicts this polymorphism to be benign (Figure 16) (205), the 72R form has been reported to be much more efficient at inducing apoptosis (206) but less active at transactivation of target genes (207). These assessments are not consistent with each other and may reflect the peculiar consequences of the P72R change: it modifies p53 protein structure in such a way as to noticeably speed its movement through polyacrylamide and impair downstream transcriptional regulation without affecting its DNA-binding specificity and affinity. Differences in cofactor binding between 72P and 72R p53, however, may explain their differential transactivation behaviour (207).

Figure 16. Automated analysis of rs1042552 by PolyPhen-2.

The polymorphism rs1042552 was deemed likely benign by Polyphen-2 analysis. Polyphen-2 considers the evolutionary conservation of a sequence variant, its translation consequences (i.e. silent, missense or nonsense), whether it affects known functional domains in the protein product and its probable effect on 3D structure.
The expression profile of cell cycle and apoptosis control genes in the fusion-negative sample is consistent with the weakened transactivation ability of the 72R variant of p53. Taken together, these findings suggest that the rs1042522 polymorphism impairs the ability of p53 to act as a tumour suppressor in TMPRSS2:ERG-negative prostate cancer. The involvement of affected downstream genes in mechanisms of anti-cancer drug action additionally stresses the clinical relevance of this polymorphism in terms of both predictive marker and targeted drug discovery. These findings also demonstrate the usefulness of transcriptome sequencing for investigation of mutations and polymorphisms.

5 Future Directions

In this study, we sequenced the entire transcriptomes of two prostate radical prostatectomy specimens that differed in TMPRSS2:ERG fusion status. We discovered that RNA-Seq was powerful enough to detect a host of genes that are important to prostate cancer development in either or both samples and that Genesifter software was very useful for drawing attention to genes and sets of genes that warrant second looks. This naturally implies two primary dimensions of further study.

One, obviously, is to further mine the enormous dataset generated for additional points of interest. The research described here has only scratched the surface of the available data, focusing on the most overexpressed, most differentially expressed and the most physiologically prominent gene entries. Comprehensive study and discussion of the available information is already far outside the scope of this thesis, to say nothing of additional associations and relationships that would arise if our RNA-Seq data were considered alongside further replicates, different controls or RNA-Seq experiments conducted by other groups. Studying more samples, in particular, would lend resultant findings more much-needed epidemiological weight. Furthermore, the current study was designed primarily to interrogate differences between TMPRSS2:ERG-positive and -negative prostate cancer. Taking the same analytical approach with sequenced transcriptomes of normal prostate tissue or material from metastases, on the other hand, could provide information about transcriptomic changes underlying prostate cancer progression in general. Given the consistent deflation of NGS cost and the emergence of third-generation sequencing, one can only imagine that these types of tasks would become more and more feasible.
Another path of study is to follow one of the research leads identified in this communication:

First, our findings should be validated by quantitative PCR. The high but varied expression of prostate cancer-associated transcripts and biomarkers, like MALAT1 and AMACR, suggests that precise determination of their expression in a larger cohort of prostate cancer samples could lead to discoveries about the mechanisms that regulate their transactivation and for which they can be useful biomarkers. In particular, the alternating expression of some of these genes in correlation with fusion status suggests that they could measured together to more sensitively or specifically provide prognostic cues.

Second, the precise interaction between ERG, RhoGDIB and other ETS transcription factors and effectors could be further clarified. ChIP experiments with RhoGDIB promoter mutants could reveal which EBS are involved in ERG-mediated transactivation. Assaying the effects of RhoGDIB overexpression or siRNA-mediated depletion in additional cell lines or murine models would determine the universality of the phenotypic changes observed herein. Examination of ERG and ETV1 overexpression in more samples could further test the hypothesis that they occur mutually exclusively and advance our understanding of the multifocal, heterogeneous nature of prostate cancer. For targets like EGR1, which are regulated by multiple ETS factors and regulatory of cancer genes like p53, further study of their use as therapeutic targets is warranted.

Third, the proposed positive feedback loop between increased Wnt signalling, ß-catenin accumulation, AR activity enhancement and TMPRSS2:ERG transactivation requires verification. In particular, our study has only shown that fusion gene expression is correlated with increased expression of Wnt pathway members and has yet to establish causation, much less determine the underlying mechanism. Transactivation studies like those performed and proposed for RhoGDIB above can confirm or refute our hypothesis that fusion gene expression is responsible for the observed changes in Wnt gene expression.

Last, we have shown that the p53 axis may be fertile ground for therapeutic target and biomarker discovery in prostate cancer. Almost every p53 target gene considered has already been identified as a necessary mediator of popular and emerging anticancer drugs. The missense SNP rs1042522 could be studied as a predictive marker of patient response to these drugs and similar approaches. Furthermore, given the correlation between p53 target and fusion gene expression, the TMPRSS2:ERG fusion may also be considered as a predictive marker in the same capacities.
6 Conclusion

Our research demonstrates the utility of whole transcriptome sequencing for confirming established associations of genes with disease and for forming new associations of this kind. New involvements of TMPRSS2:ERG, ETS transcription factors, ETS targets, the Wnt signalling pathway and a polymorphism in p53 were established in the context of prostate cancer. A focused investigation of RhoGDIB additionally led to confirmation that it is a direct transactivational target of ERG that can elicit phenotypic changes characteristic of cancer.

Further investigation of the genes and pathways identified in this manner will provide insights into mechanisms contributory to prostate cancer and the involvement of TMPRSS2:ERG gene fusion therein. This, in turn, can guide the discovery and evaluation of diagnostic and prognostic biomarkers and therapeutic targets for prostate cancer.
Chapter 5
References


55 Ross AE, Loeb S, Landis P, Partin AW, Epstein JJ, Kettermann A, Feng Z, Carter HB and Walsh PC: Prostate-specific antigen kinetics during follow-up are an


Ma J and Waxman DJ: Combination of antiangiogenesis with chemotherapy for more effective cancer treatment. Mol Cancer Ther 7: 3670-3684, 2008.


160 Hermans KG, van MR, van DH, Jenster G, van Weerden WM and Trapman J: TMPRSS2:ERG fusion by translocation or interstitial deletion is highly relevant


171 Sander EE, ten Klooster JP, van DS, van der Kammen RA and Collard JG: Rac downregulates Rho activity: Reciprocal balance between both GTPases


prostate cancer cell lines with or without resistance to androgen: a clue to
application of gefitinib to hormone-resistant prostate cancer. Oncol Rep 15: 1453-
1460, 2006.

198 Starling GC, Bajorath J, Emswiler J, Ledbetter JA, Aruffo A and Kiener PA:
Identification of amino acid residues important for ligand binding to Fas. J Exp

DA, Olumi AF and Kumar AP: Involvement of FLIP in 2-methoxyestradiol-
induced tumor regression in transgenic adenocarcinoma of mouse prostate model.

200 Thiery J, Abouzahr S, Dorothee G, Jalil A, Richon C, Vergnon I, Mami-Chouaib F and
Chouaib S: p53 potentiation of tumor cell susceptibility to CTL involves Fas and

201 Ingermann AR, Yang YF, Han J, Mikami A, Garza AE, Mohanraj L, Fan L, Idowu M,
Ware JL, Kim HS, Lee DY and Oh Y: Identification of a novel cell death receptor
mediating IGFBP-3-induced anti-tumor effects in breast and prostate cancer. J

202 Kawabata R, Oie S, Takahashi M, Kanayama H, Oka T and Itoh K: Up-regulation of
insulin-like growth factor-binding protein 3 by 5-fluorouracil (5-FU) leads to the
potent anti-proliferative effect of androgen deprivation therapy combined with 5-

203 Chan JM, Stampfer MJ, Ma J, Gann P, Gaziano JM, Pollak M and Giovannucci E:
Insulin-like growth factor-I (IGF-I) and IGF binding protein-3 as predictors of

204 Lehrer S, Diamond EJ, Droller MJ, Stone NN and Stock RG: Re: Insulin-like growth

205 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov
AS and Sunyaev SR: A method and server for predicting damaging missense

206 Dumont P, Leu JJ, Della Pietra AC, George DL and Murphy M: The codon 72
polymorphic variants of p53 have markedly different apoptotic potential. Nat

207 Thomas M, Kalita A, Labrecque S, Pim D, Banks L and Matlashewski G: Two
polymorphic variants of wild-type p53 differ biochemically and biologically. Mol