IDENTIFICATION OF METASTASIS-DRIVING GENES AS POTENTIAL THERAPEUTIC TARGETS / BIOMARKERS FOR METASTATIC PROSTATE CANCER

by

Yan Ting Chiang

B.Sc., the Hong Kong University of Science and Technology, 2009

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Interdisciplinary Oncology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2015

© Yan Ting Chiang, 2015
Abstract

Metastatic prostate cancer is currently incurable. Metastasis is thought to result from changes in the expression of specific metastasis-driving genes, leading to a cascade of activated downstream genes setting the metastatic process in motion. As such, metastasis-driving genes could provide effective therapeutic targets and prognostic biomarkers for improved disease management.

In search of potential metastasis-driving genes, genes with elevated expression in patient-derived metastatic LTL-313H prostate cancer tissues, as distinct from non-metastatic LTL-313B tissues, were identified. Among these genes, TIMELESS and DLXI were promising. Unfortunately, their silencing and overexpression in prostate cancer cells did not lead to inhibition of metastatic properties, indicating that they were not metastasis-driving genes.

A different, novel approach was used based on the notion that metastasis-driving genes can activate genes in an amplification cascade fashion. Accordingly, I used the IPA’s Upstream Regulator Analysis tool to analyze the differential gene expression profile of the metastatic and non-metastatic tissues to predict the upstream master regulatory (metastasis-driving) genes accountable for the differential expression. Six candidate genes were identified, including GATA2, a pioneer factor-encoding gene. Elevated GATA2 expression in clinical metastatic prostate cancer specimens correlated with poor patient prognosis. Furthermore, GATA2 gene silencing in human prostate cancer LNCaP cells led to marked reduction in cell proliferation, cell migration, tissue invasion, focal adhesion disassembly and a dramatic change in transcriptional activity, indicating that GATA2 plays a critical role in prostate cancer metastasis. As such, GATA2 could represent a metastasis-driving gene and a potential therapeutic target for inhibiting the growth and metastasis development in prostate cancer.
Further analysis of GATA2-regulated genes led to the development of a GATA2-based metastatic gene signature. Its prognostic value was confirmed using two prostate cancer patient cohorts. In addition, it was shown to be a prognostic factor for risk assessment of metastasis development, independent of the widely used D’Amico prognostic classification system. However, a thorough validation is critical and, if successful, the GATA2-based gene signature could lead to a paradigm shift in the management of early prostate cancer.

In conclusion, the findings of this study appear to be potentially useful for improved management of metastatic prostate cancer.
Preface

A version of Chapter 2 has been published. [Chiang YT], Gout PW, Collins CC, Wang Y. *Prostate cancer metastasis-driving genes: hurdles and potential approaches in their identification.* Asian J Androl, 2014. **16**(4): p. 545-8. As the lead investigator, I was responsible for the experimental design, data collection and analysis, as well as manuscript composition. Gout PW was involved in reviewing and writing the article. Collins CC was involved in analyzing data and reviewing the article. Wang YZ was the supervisory author of this project and was involved throughout the project in concept formation and manuscript design.

A version of Chapter 3 has been published. [Chiang YT], Wang K, Fazli L, Qi RZ, Gleave M, Collins C, Gout PW, Wang YZ. *GATA2 as a potential metastasis-driving gene in prostate cancer.* Oncotarget, 2014. **5**(2): p. 451-61. As the lead investigator, I was responsible for the experimental design, data collection and analysis, as well as manuscript composition. Wang K was involved in producing and analyzing bio-statistical data. Fazli L was involved in tissue microarray construction and evaluation. Qi RZ, Gleave M and Collins CC were involved in interpreting data and reviewing the article. Gout PW was involved in interpreting data, reviewing the article and writing the manuscript. Wang YZ was the supervisory author of this project and was involved throughout the project in concept formation and manuscript design.

A version of Chapter 4 is being prepared for publication. [Chiang YT], Crea F, Erho N, Alshalalfa M, Black P, Collins C, Davicioni E, Gout PW, Wang YZ. *A GATA2-based gene signature as a prognostic factor for metastatic progression of early stage prostate cancers.* As the lead investigator, I am responsible for the experimental design, data collection and data analysis, as well as manuscript composition. Crea F is involved in experimental design and data interpretation. Erho N and Alshalalfa are involved producing and analyzing statistical data. Black
P, Collins C, Davicioni E are involved in interpreting data and reviewing the manuscript. Gout P is involved in interpreting data, reviewing the article and writing the manuscript. Wang YZ is the supervisory author of this project and is involved throughout the project in concept formation and manuscript design.
Table of Contents

Abstract ....................................................................................................................................... ii

Preface .......................................................................................................................................... iv

Table of Contents .......................................................................................................................... vi

List of Tables .................................................................................................................................. xi

List of Figures ................................................................................................................................. xii

List of Abbreviations ..................................................................................................................... xiv

Acknowledgement .......................................................................................................................... xix

Dedication ........................................................................................................................................ xx

Chapter 1: Introduction .................................................................................................................. 1

1.1 The Prostate ............................................................................................................................. 1

1.2 Prostate Cancer ....................................................................................................................... 1

1.2.1 Screening and Diagnosis ..................................................................................................... 2

1.2.2 Risk Factors ....................................................................................................................... 3

1.2.3 Treatment ............................................................................................................................ 3

1.2.4 Prognosis ............................................................................................................................ 4

1.2.5 Genomic Changes in Prostate Cancer ................................................................................. 6

1.3 Metastasis ................................................................................................................................ 7

1.3.1 Multi-step Processes of Metastasis in Prostate Cancer ....................................................... 8

1.3.2 Focal Adhesions and Cell Migration .................................................................................... 11
Chapter 2: Identification of Potential Prostate Cancer Metastasis-driving Genes

2.1 Introduction

2.2 Materials and Methods

2.2.1 Materials

2.2.2 Comparative Analysis of Gene Expression Profiles

2.2.3 Clinical Relevance Analysis

2.2.4 RNA Isolation from OCT-embedded Tissues

2.2.5 Quantitative Real-time Polymerase Chain Reaction (qRT-PCR)

2.2.6 Cell Cultures

2.2.7 Plasmid Construction

2.2.8 Small Interfering RNA (siRNA) and Cell Transfection

2.2.9 Western Blotting

2.2.10 MTS Cell Proliferation Assay

2.2.11 Wound Healing Migration Assay

2.2.12 Flow Cytometric Cell Cycle Analysis

2.3 Results

2.3.1 Elevated TIMELESS and DLX1 Gene Expression in Metastatic Prostate Cancer Tissues
2.3.2 Identification of *TIMELESS* Function in Prostate Cancer Cell Lines .......................... 26
2.3.3 Identification of *DLX1* Function in Prostate Cancer Cell Lines .................................. 26
2.4 Discussion .............................................................................................................................. 26

Chapter 3: Identification of *GATA2* as a Potential Metastasis-driving Gene in Prostate Cancer

3.1 Introduction .......................................................................................................................... 36
3.2 Materials and Methods ....................................................................................................... 37
3.2.1 Materials .......................................................................................................................... 37
3.2.2 Cell Culture ..................................................................................................................... 38
3.2.3 Identification of Upstream Regulatory Genes ................................................................. 38
3.2.4 Clinical Relevance Analysis ............................................................................................ 38
3.2.5 Tissue Microarray (TMA) Construction and Immunohistochemistry ............................ 39
3.2.6 siRNA Transfection ........................................................................................................ 39
3.2.7 Western Blotting ............................................................................................................. 39
3.2.8 MTS Cell Proliferation Assay ......................................................................................... 40
3.2.9 Wound Healing Cell Migration Assay ............................................................................ 40
3.2.10 Modified Boyden Chamber Assays .............................................................................. 41
3.2.11 Focal Adhesion Disassembly Assay and Immunofluorescence Staining ....................... 41
3.2.12 Total RNA Isolation and Quantitative Real-Time PCR (qRT-PCR) .............................. 42
3.2.13 Gene Expression Data Profiling .................................................................................... 42
3.2.14 Gene Expression Data Analysis ................................................................. 43
3.2.15 Weighted Gene Co-expression Network Analysis (WGCNA) ..................... 43
3.3 Results ........................................................................................................... 44
3.3.1 GATA2 as a Potential Upstream Metastasis-driving Gene ......................... 44
3.3.2 Elevated Expression of GATA2 Correlates with Poor Prostate Cancer Prognosis .. 44
3.3.3 Expression of GATA2 in Prostate Cancer Cell Lines and Tissues ................. 45
3.3.4 GATA2 Gene Silencing Reduces in vitro Migration, Matrigel Invasion and Proliferation of Prostate Cancer Cells .......................................................... 45
3.3.5 A Role for the GATA2 Gene in Focal Adhesion (FA) Disassembly ................. 45
3.3.6 Changes in Gene Expression Induced by GATA2 Silencing .......................... 46
3.3.7 Functions of GATA2 as Indicated by Biostatistical Analysis ....................... 46
3.4 Discussion ...................................................................................................... 47

Chapter 4 : A GATA2-based Gene Signature as a Prognostic Factor for Metastatic Progression of Early Stage Prostate Cancers ....................................... 71
4.1 Introduction .................................................................................................. 71
4.2 Materials and Methods ................................................................................ 72
4.2.1 Data Sets Used .......................................................................................... 72
4.2.2 Pearson’s Correlation Analysis ................................................................. 73
4.2.3 Oncomine Analysis .................................................................................... 73
4.2.4 IPA Analysis .............................................................................................. 74
4.2.5 Development of a GATA2-based Gene Signature ........................................... 74
4.2.6 Statistical Analysis .......................................................................................... 74
4.3 Results .................................................................................................................. 75
4.3.1 Identification of Potential GATA2-regulated Genes ........................................ 75
4.3.2 Clinical Outcome-driven Analysis of the GATA2-regulated Genes ................. 76
4.3.3 Pathway-driven Analysis of the GATA2-regulated Genes ............................... 76
4.3.4 Development of a GATA2-based Gene Signature ........................................... 76
4.3.5 Performance of the GATA2-based Gene Signature in Training and Validation Data Sets .................................................................................................................. 77
4.3.6 The GATA2-based Gene Signature is an Independent Prognostic Variable ....... 77
4.3.7 High GATA2-based Gene Signature Scores Correlate with Increased Probability of Metastasis Development ................................................................. 78
4.4 Discussion ............................................................................................................. 78

Chapter 5 : Conclusions ............................................................................................ 90
5.1 Summary of Findings ........................................................................................ 90
5.2 Significance of the Research Findings ............................................................ 92
5.3 Ongoing Research: Silencing of GATA2 as a Potential Therapeutic Approach for Metastatic Prostate Cancer ................................................................. 92

References .................................................................................................................. 94
List of Tables

Table 2.1: qPCR primers sequences used ........................................................................................................ 29
Table 2.2: Gene expression changes of TIMELESS and DLX1 in prostate cancer tissue

xenografts .......................................................................................................................................................... 29
Table 3.1: qPCR primers sequences used ........................................................................................................ 51
Table 3.2: Potential candidates of prostate cancer metastasis-driving genes. ............................................. 52
Table 3.3: Number of genes showing expression changes following GATA2 silencing. ......................... 52
Table 3.4: Genes whose expression changed following GATA2 silencing .................................................. 53
Table 3.5: Module significance in prostate cancer metastasis and function annotation............................ 54
Table 3.6: Genes in brown module ................................................................................................................ 57
Table 3.7: Genes in yellow module. ................................................................................................................ 59
Table 4.1: Number of GATA2-regulated genes in prostate cancers associated with poor patient

outcome ......................................................................................................................................................... 80
Table 4.2: GATA2-regulated genes associated with cancer and disease-related processes ..................... 81
Table 4.3: Description of the 10 genes of which the GATA2-based gene signature is composed ............... 82
Table 4.4: Clinical and pathological features of the human prostate cancer cohorts used. ...................... 83
Table 4.5: Performance of the GATA2-based gene signature in training and validation sets.................. 84
Table 4.6: The GATA2-based gene signature is an independent prognostic variable ............................... 84
List of Figures

Figure 1.1: Multi processes of metastasis in cancer ............................................................... 16
Figure 1.2: Formation of FA complexes via integrins to ECM ............................................. 17
Figure 1.3: The assembly and disassembly of FA during cell migration .............................. 18
Figure 2.1: Master regulatory gene in the development of metastasis ................................. 30
Figure 2.2: Elevated expression of TIMELESS and DLX1 associated with metastatic prostate cancer (PCa) ........................................................................................................... 31
Figure 2.3: Analysis of TIMELESS function in PC3M prostate cancer cells. ....................... 33
Figure 2.4: Analysis of DLX1 function in PC3M prostate cancer cells ................................. 35
Figure 3.1: A two-step strategy used in identifying potential metastasis-driving genes in prostate cancer .......................................................................................................................................................... 60
Figure 3.2: Elevated expression of GATA2 associated with poor prostate cancer (PCa) patient prognosis ........................................................................................................................................................................... 62
Figure 3.3: Expression of GATA2 in prostate cancer cell lines and tissues ......................... 63
Figure 3.4: GATA2 silencing decreases proliferation of prostate cancer cells ...................... 64
Figure 3.5: GATA2 silencing decreases cell migration and matrigel invasion of prostate cancer cells ......................................................................................................................................................... 66
Figure 3.6: A role for the GATA2 gene in focal adhesion disassemble ................................. 68
Figure 3.7: Gene expression changes induced by GATA2 silencing in LNCaP cells ............ 69
Figure 3.8: Weighted gene co-expression network analysis (WGCNA) was used to identify genes that are potentially regulated by GATA2 ........................................................................................................... 70
Figure 4.1: Study outline .......................................................................................................... 85
Figure 4.2: Five GATA2-regulated genes with potential for predicting the risk of biochemical recurrence in early stage prostate cancer ..................................................................................... 86

Figure 4.3: Five GATA2-regulated genes with potential for predicting the risk of biochemical recurrence in early stage prostate cancer ..................................................................................... 87

Figure 4.4: Forest plots of univariable analysis showing the prognostic power of GATA2-based gene signature for the development of metastasis as endpoints........................................ 88

Figure 4.5: GATA2-based gene signature used in Kaplan Meier analysis with endpoint as development of metastasis ........................................................................................................... 89
List of Abbreviations

ADT: Androgen deprivation therapy
AHR: Aryl hydrocarbon receptor
ANG: Angiogenin
ATR: Ataxia telangiectasia and Rad3 related
AR: Androgen receptor
ARNT: Aryl hydrocarbon receptor nuclear translocator protein
ATCC: American type culture collection
AUC: Area under the ROC curve
AURKA: Aurora kinase A
AZGP1: Zinc-alpha-2-glycoprotein
BCA assay: Bicinchoninic acid assay
BCL3: B-cell lymphoma 3-encoded protein
BCR: Biochemical recurrence
BH: Benjamini-Hochberg
BIRC5: Baculoviral IAP repeat-containing 5
BMP6: Bone morphogenetic protein 6
BPH: Benign prostatic hyperplasia
CAM: Cell adhesion molecule
CAPRA: Cancer of the prostate risk assessment
CI: Confidence interval
CRPC: Castration-resistant prostate cancer
CRYAB: Crystalin alpha b
CXCL12: Chemokine (C-X-C motif) ligand 12
CXCR4: Chemokine (C-X-C motif) receptor 4
CYP3A5: Cytochrome P450 3A5
DAB: (3, 3'-diaminobenzidine) peroxidase substrate
DAPI: 4',6-diamidino-2-phenylindole nuclear stain
DLX1: Distal-less homeobox 1
DRE: Digital rectal examination
DST: Dystonin
ECM: Extracellular matrix
ECE: Extra-capsular extension
EGF: Epidermal growth factor
EIT: Epithelial immune cell-like transition
EMT: Epithelial-mensenchymal transition
EPAS1: Hypoxia-inducible factor-2alpha
EZH2: Enhancer of zeste homolog 2 (Drosophila)
E2F1: E2F transcription factor 1
FA: Focal adhesion
FAK: Focal adhesion kinase
FBS: Fetal bovine serum
FC: Fold change
FDR: False discovery rate
FGF: Fibroblast growth factor
FOXM1: Forkhead box M1
F2R: Coagulation factor II (thrombin) receptor
GAPDH: Glyceraldehyde-3-phosphate dehydrogenase
GATA2: GATA binding protein 2
GEO: Gene expression omnibus
GS: Gleason score
HIF1A: Hypoxia inducible factor 1, alpha subunit
HPRT: Hypoxanthine phosphoribosyltransferase 1
IGF: Insulin-like growth factor
IKZF1: DNA-binding protein Ikaros
IPA: Ingenuity pathway analysis
ITGA6: Integrin alpha-6
ITGB1: Integrin beta-1
KLK2: Kallikrein-2
LHRH: Luteinizing hormone-releasing hormone
LNI: Lymph node involvement
MIAME: Minimum information about a microarray experiment
MMP: Matrix metaloproteinase
MSKCC: Memorial Sloan Kettering Cancer Centre
MTPN: Myotrophin
MTS: 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium
MVA: Multivariable analysis
MYC: Avian myelocytomatosis viral oncogene
MYCN: V-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog
NaCl: Sodium chloride
NFKB2: Nuclear factor NF-kappa-B p100 subunit
NHT: Neoadjuvant hormonal therapy
NKX3.1: NK3 homeobox 1
NOD-SCID: Non-obese diabetic/severe combined immunodeficient
NUSAP1: Nucleolar and spindle associated protein 1
OCT: Optimal cutting temperature compounds
PBS: Phosphate buffered saline
PCa: Prostate cancer
PCR: Polymerase chain reaction
PDX: Patient-derived xenograft
PDX1: Pancreatic and duodenal homeobox 1
PIN: Prostatic intra-epithelial neoplasia
PPARA: Peroxisome proliferator-activated receptor alpha
PPARGC1B: Peroxisome proliferator-activated receptor gamma, coactivator 1 beta
PRC1: Protein regulator of cytokinesis 1
PSA: Prostate-specific antigen
PTEN: Phosphatase and tensin homolog
qRT-PCR: Quantitative real-time polymerase chain reaction
RB1: Retinoblastoma 1
RFC5: Replication factor c
RIN: RNA integrity number
RIPA buffer: Radioimmunoprecipitation assay buffer
ROC: Receiver operating characteristic
RPMM: Roswell Park Memorial Institute medium
SCID: Severe combined immuno-deficient
SDS: Sodium dodecyl sulfate
siRNA: Small interfering RNA
SM: Surgical margin
SNAI1: Zinc finger protein SNAI1
SRC: v-src avian sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog
SRCAP: Snf2-related CREBBP activator protein
STAT4: Signal transducer and activator of transcription 4
STMN1: Stathmin
SVI: Seminal vesicle invasion
TGFβ1: Transforming growth factor, beta 1
TIMELESS: Timeless homolog (drosophila)
TIMP: Tissue inhibitors of metalloproteinase
TMA: Tissue microarray
TMPRSS-ERG: Transmembrane protease, serine 2/ E26 transformation-specific family
TNM: Tumor/nodes/metastasis
TOM: Topological overlap measure
TP53: Tumor protein p53
TPX2: Targeting protein for XkIp2
TRIM24: Tripartite motif containing 24
TUBA1B: Tubulin alpha 1B chain
TURP: Transurethral resection of prostate
UCSF: University California of San Francisco
UHRF1: Ubiquitin-like with PHD and ring finger domains 1
uPA: Urokinase-type plasminogen activator
UVA: Univariable analysis
VACURG: Veterans Administration Cooperative Urological Research Group
VEGFB: Vascular endothelial growth factor B
WGCNCA: Weighted gene co-expression network analysis
WT1: Wilms tumor 1
XPO6: Exportin 6
Acknowledgement

Completing a PhD degree is one of the most challenging yet joyful journeys in my life. I feel fortunate to be given this opportunity by my supervisor, Dr. Yuzhuo Wang, who is one of the most influential people in my life. He is an inspiring and supportive advisor to me throughout my graduate school career. I owe him dearly for his trust in me, giving me the freedom to explore interesting ideas. I extend my gratitude to my committee members, Drs. Cheryl D. Helgason, Marianne D. Sadar, Christopher J. Ong and Stephen Lam for their guidance and valuable input in my research study.

I would like to thank all of the past and present members in the Wang lab for making the lab an interesting, creative and fun workplace. I give special thanks to Dr. Gout for his support and encouragement throughout my PhD program, and in particular while writing of this thesis. There was no way for me to complete my work without your helpful advices and supports. I should also extend my sincerest thanks to all of the members that I have gotten to know in the British Columbia Cancer Research Centre and the Vancouver Prostate Centre. It has been a great privilege for me to know you all by heart and our friendship will stay with me forever.

I wish to thanks my brother and sisters for your unconditional love that was my driving force. My husband, Sulin Cao, whose love and support that I lean on; I really appreciate that you were always there to share the highs and lows of my life. Last but not least, I would like to dedicate this work to my deceased parents, Yew Seng Chiang and Siew Kee Goh. I hope this work makes you proud.
To my husband, Sulin Cao,

and my dearest son, Ryder Yuxuan Cao

In memory of my parents
Chapter 1: Introduction

1.1 The Prostate

The prostate is a tubuloalveolar exocrine gland of the male reproductive system. It is located underneath the bladder and immediately in front of the bowel, encircling the neck of the bladder and urethra. The main function of the prostate is to produce and secrete a slightly alkaline (pH 7.3) solution for the protection of spermatozoa in the acidic environment of the vagina [1]. An adult prostate is slightly bigger than a walnut. Histologically, a normal prostate gland mainly consists of double layers of epithelial cells (separated from stroma by a basement membrane), i.e. a continuous layer of cuboidal basal cells covered by a layer of columnar luminal cells. The latter actively secrete prostate-specific proteins into the glandular lumen [2, 3]. Throughout the cuboidal basal cell layer, neuroendocrine cells are infrequently scattered; their functions remain unclear [3]. The prostate can be compartmentalized into four distinct regions, known as the peripheral zone, central zone, transition zone, and anterior fibro-muscular zone (stroma) [4]. Male sex hormones (androgens) and their receptor plays a major role in the growth and function of the prostate [5]. Common prostate disorders include prostatitis, benign prostatic hyperplasia (BPH), and cancer. BPH, the enlargement of prostate, occurs so frequently in aging males that it is regarded as part of the normal aging process [6].

1.2 Prostate Cancer

Prostate cancer is the most commonly diagnosed cancer and the second leading cause of cancer death in Western men [7]. According to Canadian Cancer Statistics, it has been estimated that more than 23,000 men will be diagnosed with the disease in Canada in 2014. One in eight Canadian men is expected to be diagnosed with prostate cancer during his life time and one in 28
will die of it. When the malignancy is localized to the prostate, surgery or radiation therapy can be curative. However, many treated patients will experience local recurrence or metastasis of the disease [8-10]. Advanced, metastatic prostate cancer is highly resistant to therapy and is currently incurable.

In the prostate, most cancers develop within the luminal epithelium of the peripheral zone, in contrast to non-malignant BPH that normally arises in the transition zone [3, 11]. It is commonly believed that prostate cancer progresses through multiple stages: from proliferative inflammatory atrophy, to prostatic intra-epithelial neoplasia (PIN), localized prostate cancer, metastatic prostate cancer and ultimately castration-resistant prostate cancer (CRPC) [12]. Prostate cancer is considered a heterogeneous disease, as multiple distinct cancer foci with different degrees of cellular dysplasia and altered genetic content are usually found in a single prostate [13, 14]. In contrast, metastatic prostate cancer is considered a monoclonal disease, as multiple metastases in a patient are often clonally related [15].

1.2.1 Screening and Diagnosis
Most prostate cancers are first suspected during a digital rectal examination (DRE) and via a prostate-specific antigen (PSA) blood test. When cancer is suspected, the actual diagnosis can only be made following a prostate biopsy. The stage of the disease is indicated by pathologists using histological and clinical staging systems. The most commonly used histological method is the Gleason grading system developed in the 1970s by Dr. Donald F. Gleason and a panel of the Veterans Administration Cooperative Urological Research Group (VACURG) [16, 17]. A Gleason grade of 1-5 is assigned on the basis of changes developed in the architecture of the gland (a grade of 4-5 indicates a poorly differentiated cancer with ill-defined boundaries). As cancers are often multifocal, a Gleason score (2-10) is determined by adding the grades from the
two most common cancer cell patterns within the tumor. A high Gleason score indicates an aggressive disease with poor patient prognosis. The clinical status of the disease is also characterized by a four-stage TNM system, evaluating tumor size (T), degree of spread to regional lymph nodes (N) and presence of distant metastases (M).

1.2.2 Risk Factors
While causes of prostate cancer remain elusive, several risk factors are known to increase a man’s chance of developing the disease. Primary risk factors are age, family history and race [18, 19]. The incidence of prostate cancer is very low for men younger than 45 years, but it increases exponentially with age. Sixty percent of all newly diagnosed prostate cancer cases occur in men 70 years of age and older. Having first degree family members (father, brother or son) with prostate cancer can increase the risk of getting the disease by two to three fold. Moreover, men from African American descent are more likely to be diagnosed with prostate cancer and have a mortality rate more than double that of the Caucasian male [19]. Although males in Asia in general have the lowest incidence of prostate cancer, the occurrence of the disease for Asian men living in western countries is almost identical to that for Caucasian men in America [20]. This suggests that environmental factors such as diet and lifestyle could be secondary risk factors for prostate cancer. The accumulated scientific evidences indicate that high consumption of meat and dairy products has been linked to a greater risk in developing prostate cancer. On the other hand, high level of physical activity is associated with decreased risk of prostate cancer development [21].

1.2.3 Treatment
Treatment of early stage prostate cancer (with a Gleason score ≤ 6) includes active surveillance, surgery or radiation therapy. Active surveillance is considered a good option for men whose
cancers are slow growing and symptomless, and/or who are elderly and suffer from other serious medical conditions. In active surveillance, patients are regularly tested for serum PSA levels, and subjected to DREs, transrectal ultrasound and needle biopsies to check the growth of their malignancies [22]. When the growth of localized cancers accelerates, surgery or radiation are the usual options for treatment. It has been estimated that 20-30% of patients who are treated will relapse after about five years [8-10]. In such a case, the recurring cancers are routinely treated with androgen deprivation therapy (ADT), i.e. depriving cancers of male sex hormones (androgens), as their growth in general is dependent on testosterone and dihydrotestosterone [23]. The ADT is mainly performed by administration of luteinizing hormone-releasing hormone (LHRH) agonist and antagonist [24]. After one to three years, however, most hormone-dependent cancers become refractory to ADT and progress to castration resistant prostate cancer (CRPC). In such a case, a secondary hormone therapy using anti-androgens or androgen receptor (AR) antagonist (e.g. bicalutamide, MDV3100) is used [25]. The success of such drugs in targeting CRPC indicates the dependency of these tumors upon the androgen receptor to drive tumor growth. Moreover, it has been shown that expression of steroidogenic enzymes (e.g. CYP17A1), which support de novo biosynthesis of testosterone in cancer cells, is enhanced in CRPC. Thus, inhibitors that block such androgen biosynthesis (e.g. abiraterone) are commonly used for CRPC treatment [26]. Chemotherapy (e.g. doxetacel, cabazitaxel) is normally used when the disease has metastasized. The development of therapy resistance is the major challenge in the management of metastatic prostate cancer, a disease that is currently incurable.

1.2.4 Prognosis

Since the PSA blood test has become a routine procedure for prostate cancer screening of men over the age of 50, more and more men have been diagnosed with early stage (low grade, organ-
confined) prostate cancer [27-29]. The majority of these men are not at risk of developing metastatic prostate cancer and should be spared from unnecessary interventions leading to morbidity. Conversely, men that are at risk of developing metastatic disease should be treated aggressively, as suggested by recent clinical trials [30, 31]. Unfortunately, it is currently not possible to differentiate which early stage prostate cancer patients are at risk of developing metastasis. Thus the primary means of determining the appropriate treatment for such patients still largely rely on Gleason scores, tumor stages and serum PSA levels, tools which generally fail to provide an accurate prediction for the progression of early stage prostate cancer [32-34].

Current classification systems for predicting prostate cancer progression:

1) *Partin tables* are used to predict whether a localized tumor will remain confined to the prostate, based on three variables: Gleason scores, serum PSA levels, and clinical TNM stages. The tables, originally developed by Dr. Alan Partin et al, are based on accumulated medical records of 703 patients with clinically localized disease who underwent a radical prostatectomy (without neoadjuvant therapy) between 1982 and 1991 at the James Buchanan Brady Urological Institute [35]. Care must be taken in interpreting the predictions since the tables were developed using data from patients in a specific institute, and may not be valid for the general patient population.

2) *D’Amico classification* is the most widely used tool in the clinic to evaluate the risk of prostate cancer recurrence [36]. It groups patients into three categories: low, intermediate, and high risk of recurrence, using parameters such as Gleason scores, PSA levels and clinical TNM stages. The major downside of this 3-level system is that it is not able to accurately evaluate the risk of recurrence in men having multiple risk factors (e.g., high Gleason score and high PSA level), rendering the system not that useful.
3) *Kattan nomograms* are a set of tools, developed by Drs. Michael Kattan and Peter Scardino, for clinicians to estimate the probability of disease recurrence and/or survival of a patient following surgery or radiation therapy [37]. The nomograms make use of data (PSA levels, clinical TNM stages, Gleason scores, numbers of positive and negative biopsy cores) available either at the time of diagnosis or after surgery.

4) The UCSF *Cancer of the Prostate Risk Assessment (CAPRA) score* is used to estimate a patient’s likelihood of developing metastasis and cancer-induced mortality. It is based on addition of points assigned to age at diagnosis, PSA level at diagnosis, Gleason scores of biopsy specimens, clinical TNM stage and percent of prostate needle biopsy cores containing cancer, leading to a CAPRA score of 0-10. An increase of 2 points roughly represents a doubling of the risk that the cancer will recur. The CAPRA score of patients can be easily calculated, and has a predictive power comparable to the best nomograms. Recently, it has been shown to predict clinical prostate cancer endpoints with good accuracy [38].

**Prognostic Biomarkers**

In view of the above, biomarkers for reliably predicting disease progression are urgently needed for better disease management. Identifying such biomarkers, however, is a major challenge as there is a lack of understanding of the molecular and cellular mechanisms underlying the development of metastatic prostate cancer.

**1.2.5 Genomic Changes in Prostate Cancer**

Extensive genomic analyses of prostate cancer tissues have indicated that the clinical heterogeneity of the disease is based on a complex molecular landscape of genomic rearrangements and rare point mutations leading to increased transcriptomic diversity [39]. High
frequency somatic alterations in prostate cancer have been shown to preferentially occur in a
number of chromosomal regions, i.e. chromosomal gains in 8q and losses in 3p, 8p, 10q, 13q,
and 17p [40-42]. Moreover, a number of oncogenes and tumor suppressor genes have been
mapped to these particular chromosomal regions showing marked changes in expression, e.g.,
elevated expression of MYC at 8q24; reduced expression of NKX3.1 at 8p21, PTEN at 10q23,
RB1 at 13q14 and TP53 at 17p13. In addition to chromosomal gains and losses, a number of
fusion genes have been identified, including TMPRSS2-ERG at 21q22, found in 50% of localized
prostate cancers [43, 44]. In contrast to other cancers, point mutations in prostate cancer are
relatively infrequent. Loss of activity of tumor suppressors (e.g., p53, PTEN and RB1) are more
frequent, due to alterations in copy number through chromosomal rearrangements rather than to
mutations [41].

1.3 Metastasis

Metastasis is the process by which cancers spread from the tissue of origin to other tissues [45,
46]. It is the principal cause of death from solid tumors. The two main routes of metastasis are
lymphatic and hematogenous dissemination. The spreading of cancer cells to regional lymph
nodes through lymphatic dissemination serves as a first clinical indication of metastasis. Yet the
spreading of prostate cancer cells to an anatomically distant site is more likely to happen via the
blood stream (hematogenous dissemination) [47]. Although the molecular mechanisms that
underlie the development of metastasis remain largely unknown, metastasis is generally
recognized to involve a multi-step process of complex, interrelated events, including cell
detachment from the primary tumor, tissue invasion, survival in blood or lymph vessels,
extravasation, adhesion and proliferation at a distant site [48-50].
1.3.1 Multi-step Processes of Metastasis in Prostate Cancer

**Loss of cell adhesion**

Carcinoma tissues are characterized by changes in architecture due to loss of intercellular adhesion. Such a loss is a prerequisite for the development of invasive phenotypes. It is facilitated by decreased expression of cell adhesion molecules (CAMs) such as epithelial (E)-cadherin, a transmembrane glycoprotein that has been widely investigated. Its extracellular domain binds to other cadherins on adjacent cells, forming bonds between cells allowing their interaction; binding of the cytoplasmic tail of E-cadherin to α, β, or γ-catenin is essential for the structure of epithelial cells [51, 52]. Loss of expression of E-cadherin and α-catenin, accompanied by increased expression of mesenchymal N-cadherin, is commonly observed in prostate cancer cells [53]. Such changes are thought to act as an “on” switch for epithelial-mesenchymal transition (EMT) by which epithelial cells lose their cell polarity and intercellular adhesion, while gaining migratory and invasive properties. As such, EMT is thought to be an important factor underlying the development of metastatic properties of prostate cancer cells. Moreover, down-regulation of E-cadherin in prostate cancer has been significantly correlated with increased tumor grade and poor patient prognosis [54-56].

**Local tissue invasion**

During metastasis, malignant cells invade through natural tissue barriers (e.g., basement membrane, interstitial matrix) which are formed by extracellular matrix (ECM), consisting of extracellular molecules (e.g., type IV collagen, laminin and heparan sulphate proteoglycan) secreted by cells, that provides structural and biochemical support to the surrounding cells [57, 58]. The breaching of the tissue barrier is facilitated by secretion by the cancer cells of proteases, such as matrix metalloproteinases (MMPs), serine proteinases, cysteine proteinases and aspartyl
proteinases [59]. Twenty-four MMPs have been described to date and elevated expression of specific MMPs has been reported to be associated with increased prostate cancer progression [60-62]. The function of MMPs is regulated by naturally occurring tissue inhibitors of metalloproteinases (TIMPs); increases in the MMP: TIMP ratio, due to either TIMP down-regulation or increased MMP production by cancer cells, can induce an invasive phenotype [63]. Urokinase-type plasminogen activator (uPA), an extensively studied member of serine proteinases, catalyses the conversion of inactive plasminogen to the very potent, broad-spectrum protease, plasmin, which can degrade many of the ECM components that are critical in the development of metastasis [64, 65].

**Survival of cancer cells in the circulation**

As solid tumors grow larger, increasing numbers of malignant cells will enter the blood circulation. However, only a small fraction of these cells will survive. Thus many of them are not able to withstand the shearing forces of the turbulence of the bloodstream [46, 66, 67], in contrast to a small portion of the cells that are present in clots formed with platelets and proteins (e.g., prothrombin, thrombin) [68]; such cells have enhanced expression of cell surface receptors (e.g., selectin), promoting adherence and aggregation with platelets [69-71]. In addition, cancer cells can be destroyed in the circulation by the anti-cancer immune response of the host. To overcome the anti-cancer immune activity, cancer cells appear to be able to acquire immune-suppressive properties via a transdifferentiation process, coined epithelial immune cell-like transition (EIT) [72]. Additionally, other processes can be used by cancer cells to evade immune destruction [73].
Metastatic spread of cancer cells to distant organs

It is well established that cancers have preferential metastatic sites: prostate cancer metastases are usually observed in regional lymph nodes and bone, as distinct from other tissues, suggesting that the metastatic spread (homing) of cancer cells is a non-random process. Various factors involved in this process are presented below:

1) Specific vascular flow patterns

There is a portal venous network between the prostate and the vertebrae [74-76], that promotes the flow of prostate cancer cells to lower lumbar vertebrae (five vertebrae between the rib cage and the pelvis) and then to the higher vertebrae (e.g., ribs, long bones, skull). The frequency of prostate cancer metastases in the vertebrae correlates with this flow pattern, i.e. there is a higher frequency of metastases in the lower lumbar spine than in the higher vertebrae, suggesting an upward metastatic spread along spinal veins after initial lumbar deposits [76].

2) Cell migration along a chemoattractant gradient

Bone tissue extracts have been reported to contain chemotactic factors (e.g., osteonectin) that can promote the chemotaxis and tissue invasion of prostate cancer cells [77, 78]. It has also been shown that TGFβ1, IGF1 and IGF2, secreted by bone cells, promote chemotaxis of prostate cancer cells to the bone [79, 80]. Furthermore, elevated expression of chemokines (e.g., CXCL12) and their receptors (e.g., CXCR4) in metastatic prostate cancer has been associated with enhancement of migratory capabilities of cancer cells to the bone and lymph nodes [81, 82].

3) Interaction of cancer cells with endothelium at the metastatic site

For cancer cells to develop into metastases, the circulating cancer cells must first come to a stop, adhere to the vascular endothelium, extravasate and invade the local ECM [83]. In the case of
prostate cancer cells, it has been hypothesised that their metastasis to bone is mediated, at least in part, by preferential adhesion to bone endothelium, involving a variety of CAMs [84-86].

**Colony formation in distant organs**

As proposed by the “seed and soil” hypothesis, originally proposed by Paget in 1889 and more recently reinforced by Fidler [48], the ability of metastatic cells (the “seeds”) to colonize a metastatic site depends on their ability to interact with the new microenvironment (the “soil”). In the case of prostate cancer cells, their interaction is known to significantly lead to the development of osteoblastic lesions by inducing expression of bone morphogenetic proteins (BMPs) that stimulate bone matrix formation [87]. BMP6, one of the BMPs, was detected in 50% of metastatic prostate cancer samples, but not in non-metastatic or benign samples [88]. As well, osteoblasts and bone stromal cells have been demonstrated to secrete many growth factors (e.g., TGFβ, FGF, EGF, IGF2) that stimulate the growth of prostatic carcinoma [80]. Moreover, bone components (collagen type 1 and osteopontin) have been reported to stimulate proliferation and anchorage-independent growth of prostate cancer cell lines [89, 90]. There is evidence showing that prostate cancer cells do not only stimulate bone matrix formation (osteoblastic activity), but also promote bone degradation through activation of osteoclasts and secretion of matrix-degrading enzymes [91]. Degradation of the bone matrix leads to release of embedded growth factors and cytokines that in turn may stimulate the proliferation of prostate cancer cells, a vicious cycle whereby the cancer cells stimulate bone turnover and bone turnover stimulates cancer cell proliferation [92].

### 1.3.2 Focal Adhesions and Cell Migration

Focal adhesions (FAs) are large, dynamic protein complexes at the cell perimeter through which the cytoskeleton of a cell connects to the ECM (Fig. 1.1). The formation of FAs is initiated by
binding of extracellular domains of integrins to the ECM leading to recruitment of various cytoplasmic proteins to the intracellular domains of the integrins [93, 94]. The cytoplasmic proteins involved include catalytic proteins [e.g., focal adhesion kinase (FAK), and e-Src] and adaptor proteins (e.g., vinculin, paxilin, talin, and α-actinin) [95, 96]. The recruitment of cytoplasmic proteins initiates signal transduction events such as phosphorylation of signaling components by focal adhesion kinase that promotes adhesion maturation [97].

During migration of cells, the FAs at their leading edges grasp the ECM and generate forces required to pull the cell forward. The FAs, at the rear end of cells, are then disassembled to release the cells at that end from the ECM. As such, migration of cells requires continuous and well-coordinated assembly and disassembly of FAs (Fig. 1.2). While FA assembly has been shown to be driven by GTP-bound Rho GTPases [98, 99], the molecular events leading to FA disassembly remain to be clarified. Recently, it has been shown that the disruption of FA disassembly can lead to metastasis [100, 101].

1.4 Prostate Cancer Tissue Xenograft Models
Cancer xenograft models are routinely established by transplanting human tumor cells into immunocompromised mice [e.g., nude (athymic) or severe combined immuno-deficient (SCID) mice] that do not reject human cells. In the case of prostate cancer, xenograft models based on cultured prostate cancer cell lines such as PC-3, LNCaP and DU145 are most widely used. While such cell line xenograft models provide valuable tools for studying basic aspects of cancer, they lack the molecular and cellular complexity of the malignancies (e.g., tumor heterogeneity) and hence are not adequate for elucidating the intricate mechanisms underlying the growth and progression of prostate cancers [102-104]. Furthermore, such models have minimal relevance to
the clinical disease, as a National Cancer Institute retrospective study has shown that the activity of anticancer drugs in cancer cell line xenograft models did not correlate with their activity against the same human cancer types in phase II clinical trials [105]. Other prostate cancer xenograft models such as CWR and LAPC cell-based systems have originally been established by injection of human primary cancer cell suspensions mixed with Matrigel into mice [106, 107]. Similar to cell line-based models, these models lack complex configurations of multiple, diverse and interacting cell types that are known to be important in prostate cancer progression [108].

Recently, our laboratory has developed a variety of transplantable prostate cancer tissue xenograft lines in NOD-SCID mice from histologically intact human primary prostatic adenocarcinomas using subrenal capsule grafting technology [109]. A comparison of three graft sites (subcutaneous, subrenal capsule and orthotopic) has demonstrated that the subrenal capsule graft site has superior properties [110]. Thus it has high engraftment rates allowing not only survival of high grade prostate cancers, but also of low grade cancers and even benign prostate tissue, a result of superior nutrient supply provided by the highly vascularized kidney. As well, the subrenal capsule grafting technique tends to preserve important properties of the original cancers, including histopathology, chromosomal aberrations and gene expression profiles [110-112]. In view of this, these subrenal capsule, patient-derived xenograft (PDX) models have relatively high clinical relevance [113]. As such, they are useful for studying various aspects of cancer growth and progression. For example, paired metastatic LTL-313H and non-metastatic LTL-313B prostate cancer tissue lines, developed from the same patient’s primary cancer tissue, can be used for studying mechanisms underlying the development of prostate cancer metastasis, including identification of potential therapeutic targets and metastatic biomarkers for metastatic prostate cancer as described in the following studies [114-116].
1.5 Objectives and Hypotheses

The objectives of the studies described in this thesis are to identify (i) metastasis-driving genes responsible for the development of metastasis in prostate cancer cells and (ii) to determine whether such genes, or their products, can be used as (a) potential therapeutic targets for prostate cancer (e.g., growth inhibition and prevention of metastasis development), and/or (b) metastatic biomarkers to reliably predict whether a given early stage prostate cancer will become metastatic.

The hypotheses underlying these studies are: (i) metastasis-driving genes can be identified via differential gene expression analysis of paired metastatic and non-metastatic patient-derived prostate cancer tissue xenografts; (ii) genes/gene products identified by this method provide potential therapeutic targets and/or metastatic biomarkers for improved management of the disease.

The specific aims of these studies are:

1. To identify potential prostate cancer metastasis-driving genes by analysis of differential gene expression of paired metastatic LTL-313H and non-metastatic LTL-313B patient-derived prostate cancer tissue xenografts

2. To determine if candidate genes identified in Aim 1 are differentially expressed in clinical prostate cancer samples (e.g., primary and metastatic specimens) and to check whether their differential expressions correlate with clinicopathological stages

3. To determine whether silencing of candidate genes can lead to inhibition of metastatic properties and reduced viability of human prostate cancer cell lines

4. To determine whether gene expression patterns of potential metastasis-driving genes (identified in Aim 3) or their downstream genes provide metastatic biomarkers (gene signatures) for risk assessment of metastasis development in early stage prostate cancers
5. To validate the usefulness of the gene signatures (identified in Aim 4) in various prostate cancer patient cohorts
Figure 1.1: Multi processes of metastasis in cancer (adapted from Fidler, I. 2003 [48]). See text session 1.3.1.
Figure 1.2: Formation of FA complexes via integrins to ECM. The cytoplasmic end of integrins recruits multiple cellular proteins to regulate the cell movement (adapted from Nagano, M. et al., 2012 [96]).
Figure 1.3: The assembly and disassembly of FA during cell migration. A higher ratio of assembly relative to disassembly of FAs leads to stable adhesion of cells to ECM (i); while a higher ratio of disassembly relative to assembly leads to unstable adhesion (ii). During cell migration, FAs assembly is required at the leading edge of migrating cell, whereas disassembly of FAs is predominant at the rear end of cells (adapted from Nagano, M. et al., 2012 [96]).
Chapter 2: Identification of Potential Prostate Cancer Metastasis-driving Genes

2.1 Introduction

Prostate cancer is the most commonly diagnosed non-cutaneous cancer and the second leading cause of cancer death for North American men [7]. When the malignancy is localized to the prostate, surgery and radiation therapy can be curative. However, many treated patients will experience local recurrence or metastasis [8-10]. Advanced, metastatic prostate cancer is highly resistant to conventional therapy and is currently incurable. Discovery of new therapeutic targets for more effective treatment of metastatic prostate cancer is urgently needed for improved disease management and patient survival [117-120].

Metastasis is thought to result from increases in the expression of specific metastasis-driving genes (master regulatory genes) that lead to activation of downstream genes mediating the metastatic process [121, 122] (Fig. 2.1). Metastasis-driving genes could therefore serve as therapeutic targets and provide metastatic biomarkers for better management of metastatic prostate cancer. The genes could be identified by comparing gene expression profiles of metastatic and non-metastatic human prostate cancer cells in patients’ specimens to pinpoint genes showing increased expression, followed by determining whether silencing of such genes can lead to inhibition of metastatic properties. A major hurdle using this approach, however, is that primary prostate cancer samples, the usual source of non-metastatic prostate cancer cells, do not consist of pure non-metastatic cells, but also contain cancer cells that have metastatic ability, making such comparison for the discovery of metastasis-driving genes not feasible. To overcome this problem, our laboratory has developed transplantable, metastatic LTL-313H and non-metastatic LTL-313B patient-derived prostate cancer tissue xenograft lines from one patient’s primary tumor in NOD-SCID mice. These xenograft lines closely resemble each other as
evidenced in their similar pattern of chromosomal copy number aberration, but one (LTL-313H) shows metastatic ability when grafted in NOD-SCID mice, and another (LTL-313B) does not. Thus comparison of gene expression profiles of such xenograft lines appears suitable for investigating the molecular basis of metastasis [109, 114].

In an effort to identify metastasis-driving genes in prostate cancer, I compared the gene expression profiles of the above metastatic LTL-313H and non-metastatic LTL-313B paired sublines. As presented below, this led to identification of a number of up-regulated genes, including TIMELESS, a gene that is required for maintenance of replication fork stability [123-126] and ATR-dependent DNA damage signaling [123] and DLX1, a family member of the drosophila distal-less homeobox family that plays a role in the development and differentiation of various mammalian tissues [127-131]. Elevated expressions of the TIMELESS and DLX1 genes have been reported for a number of human cancers, including colorectal cancer [132], breast cancer [133], head and neck squamous carcinoma [134], and acute myeloid leukemia [135]. However, elevated expression of these genes has not previously been associated with prostate cancer metastasis. Also, the function of TIMELESS and DLX1 genes in metastasis development are unknown and worth for further investigation.

2.2 Materials and Methods

2.2.1 Materials

Chemicals, solvents and solutions were obtained from Sigma-Aldrich, Oakville, ON, Canada, unless otherwise indicated.
2.2.2 Comparative Analysis of Gene Expression Profiles

Gene expression profiles of paired metastatic LTL-313H and non-metastatic LTL-313B prostate cancer tissue xenografts were obtained using microarray technology. Raw gene expression data were filtered for improved quality prior to analysis of differential gene expression. Specifically, probes without corresponding gene annotations and probes without detectable expression levels (less than 3 in log2 scale) were removed. The MIAME-compliant gene expression data are accessible through GEO Series accession number GSE41193 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41193).

2.2.3 Clinical Relevance Analysis

Gene expression profiles of prostate tissues from Memorial Sloan-Kettering Cancer Centre (MSKCC) (normal tissues, primary and secondary prostatic adenocarcinomas) and their associated clinical information [41] were downloaded from the CBio Cancer Genomics Portal website [136]. Correlations were sought between poor prognostic factors of the patients (MSKCC cohort) and the relative expression levels in their prostate cancer tissues of selected genes (identified in the xenografts) using Student’s t-test.

2.2.4 RNA Isolation from OCT-embedded Tissues

Prostate cancer specimens frozen in OCT blocks had previously been obtained from patients, with their informed written consent, following a protocol approved by a local ethics committee (Kelowna hospital, British Columbia, Canada) and examination by a pathologist [137]. Eight sections of 10 µm thick were sectioned from OCT-embedded tissues using a cryostat. Sections were immediately placed in a microcentrifuge and RNAs isolated following the manufacturer’s instructions. The integrity and quality of the total RNA samples were checked with the Agilent 2100 Bioanalyzer and NanoDrop ND-2000 UV-VIS spectrophotometer for A260/A280 and
A260/A230 ratios. Only samples with RNA Integrity Number (RIN) ≥ 8.0, A260/280 OD values between 1.8 and 2.0 and an A260/A230 OD value of 2.0 were used.

2.2.5 **Quantitative Real-time Polymerase Chain Reaction (qRT-PCR)**

Total RNA was isolated from xenograft tissues or cell lines using the RNeasy mini kit (Qiagen Inc, Hilden, Germany) following the manufacturer’s instructions. Total RNA (1 µg) was used to synthesize cDNAs using a QuantiTect Reverse Transcription Kit (Qiagen). qRT-PCR reactions using KAPA SYBR Fast Universal (Kapa Biosystems, Woburn, MA) were performed in a ViiA 7 Real-Time PCR system (Applied Biosystems, Foster City, CA). Data were averaged and normalized to a geometric mean[138] of the expression of three housekeeping genes (GAPDH, HPRT1, and TUBA1B) reported to be stably expressed in prostate tissue [139]. The primer sequences used can be found in Table 2.1.

2.2.6 **Cell Cultures**

Human PC3M prostate cancer cell line was obtained from the American Type Culture Collection (ATCC). Cultures were maintained in RPMI 1640 supplemented with 10% fetal bovine serum (FBS).

2.2.7 **Plasmid Construction**

To clone the human DLX1 gene into a pEGFP C1 vector, full length DLX1 cDNA was amplified from DNA isolated from LNCaP cells using PCR technology followed by purification using a PCR purification kit (Qiagen). The primer sequences used for the amplification can be found in Table 2.1. Subsequently, the PCR products and the pEGFP C1 vector (Clontech Laboratories, Mountain View, CA) were cleaved, using restriction enzymes (BamH1 and XhoI; Invitrogen, Carlsbad, CA), followed by their purification using the plasmid purification kit (Qiagen). Next,
the DLX1 cDNA was ligated into the vector using a ligation kit (Promega, Madison, MI). The ligated vectors were introduced into DH5α bacteria followed by selection of Ampicillin-resistant bacteria. The pEGFP-DLX1 plasmid was isolated from the bacteria using a plasmid isolation kit (Qiagen) and sequenced by Eurofins Genomics (Huntsville, AL) as a service to verify their fidelity.

A plasmid containing full length human TIMELESS cDNA, pcDNA3-TIMELESS-3FLAG, was a gift from Dr. Noguchi [140].

2.2.8 Small Interfering RNA (siRNA) and Cell Transfection

Small interfering RNA (siRNA) targeting TIMELESS (siTIM 5`-GUAGCUUAGUCCUUUCAAdTdT -3`[141]), DLX1 (siDLX1- siGENOME smart pool) and negative control (scrambled) siRNAs (siGENOME non-targeting siRNA #5) were purchased from Dharmacon (Lafayette, CO). To knockdown gene in vitro, the cells were transfected with 100-200 pmol of siRNA in lipofectamin 2000 reagent (Invitrogen) following the manufacturer’s instructions. After 48 hours of incubation, the cells were used for various functional assays.

2.2.9 Western Blotting

Cells were lysed using RIPA lysis buffer (50 mM Tris-Cl pH 7.4, 150 mM NaCl, 1% Igepal, 0.5% Na-deoxycholate, 0.1% SDS) supplemented with a protease inhibitor cocktail (Roche, Nutley, NJ). Whole-cell lysates (20 µg), whose protein concentrations were determined using the BCA protein assay (Thermo Fisher Scientific, Fremont, CA), were run on 8% SDS polyacrylamide gel for Western blotting. The following antibodies were used: anti-TIMELESS (a gift from Dr. Xiaoming Yang), and anti-β-actin (Cat No L002, Epitope Biotech Inc., Burnaby, BC, Canada).
2.2.10 MTS Cell Proliferation Assay

Cells were seeded onto 96-well culture plates (3000/well, around 30% of cell confluency) MTS (Promega) was used to determine the cell populations following the manufacturer’s instructions. The absorbance of formazan (reduced MTS) at 490 nm was measured daily until 100% of cell confluency of control cells were reached. Statistical significance was established using the Student’s t-test.

2.2.11 Wound Healing Migration Assay

Cells (8x10^5) were seeded onto 24-well culture plates and incubated at 37⁰C in a 5% CO₂ incubator; following cell attachment, the medium was changed to serum-free medium. The next day, a wound was created in the middle of a confluent cell monolayer using a pipette tip. Cell debris was removed by washing with 1xPBS (2-3 times) and the cells further incubated in RPMI medium supplemented with FBS (10%). Photographic images were taken using a Zeiss Axiovert 200M microscope (Carl Zeiss Inc., Oberkochen, Germany) immediately after generating the wound, and after 8 and 24 hours of further incubation [52]. The cell-recovered areas at different time points were measured to estimate the extent of cell migration using Adobe Photoshop (Adobe, San Jose, CA). Statistical significance was established using the Student’s t-test.

2.2.12 Flow Cytometric Cell Cycle Analysis

Cells were trypsinized from 6-well plates and washed with 1xPBS (1 ml). Next, cells were fixed with 70% ethanol and kept at 4⁰C overnight. Fixed cells were pelleted to remove ethanol and washed with PBS (1 ml). Cells were re-suspended in staining buffer (1 ml; 0.1% Triton X-100, 100 µg/ml RNaseA, 20 µg/ml propidium iodide) and incubated at room temperature for 30 minutes in the dark. The cell cycle analysis was done using FacsCanto II (BD Bioscience, Franklin Lake, NJ).
2.3 Results

2.3.1 Elevated TIMELESS and DLX1 Gene Expression in Metastatic Prostate Cancer Tissues

A comparison of gene expression data (GSE41193), obtained from paired metastatic LTL-313H and non-metastatic LTL-313B prostate cancer tissue xenografts, was used to identify genes in the metastatic LTL-313H tissue showing elevated expression (> 1.5-fold). From these genes, transcription factor-related genes (as annotated in Gene Ontology) were selected for further studies, as such genes are thought to play a critical role in metastasis [142, 143]. Subsequently, genes that had previously been associated with metastatic prostate cancer were excluded as I was interested in identifying novel metastasis-driving genes in prostate cancer. Of the remaining genes, TIMELESS and DLX1, whose elevated expression was validated by qRT-PCR (Table 2.2), were of particular interest, since they also showed elevated expression in secondary prostate cancer samples of a MSKCC cohort study [41]. As shown in Figure 2.2A, the expression level of the TIMELESS gene was significantly higher in secondary prostate cancer samples in comparison to normal prostate and primary prostate cancer samples. As well, elevated TIMELESS expression was positively associated with poor prognostic factors, including increased lymph node involvement, increased Gleason score and elevated PSA levels (Fig. 2.2B). In the case of DLX1, its expression level in the MSKCC cohort was significantly higher in both primary and secondary cancers when compared to normal prostate samples (Fig. 2.2C). This differential expression pattern of the DLX1 gene was validated by qRT-PCR analysis using a smaller pool of patients’ samples (6 normal prostate samples and 5 prostate cancer samples; Fig. 2.2D). However, elevated expression of DLX1 gene was not found to be associated with poor prognostic factors (data not shown).
2.3.2 Identification of TIMELESS Function in Prostate Cancer Cell Lines

siRNA-induced silencing of TIMELESS expression in PC3M cells led to a marked reduction in the level of TIMELESS protein (Fig. 2.3A) and markedly inhibited the migration of PC3M cells in a wound healing assay (Fig. 2.3D) without affecting their proliferation rate (Fig. 2.3B) or their cell cycle progression (Fig. 2.3C). However, overexpression of the TIMELESS gene in PC3M cells (Fig. 2.3E) did not lead to any significant changes in their migration rate (Fig. 2.3H), proliferation rate (Fig. 2.3F) or cell cycle progression (Fig. 2.3G).

2.3.3 Identification of DLX1 Function in Prostate Cancer Cell Lines

siRNA-induced silencing or overexpression of DLX1 did not affect the migration, proliferation or cell cycle progression of PC3M (Fig. 2.4A-H) and C4-2 cells (data not shown).

2.4 Discussion

Solid cancers typically consist of complex configurations of multiple, diverse, interacting cancer and normal cells. While xenografts of cultured cancer cell lines can be useful as experimental models for studying basic aspects of the metastatic process, they lack the molecular and biological complexity of the malignancies (e.g., tumor heterogeneity) and hence are not adequate for discovering novel metastasis-driving genes which drives the development of metastatic prostate cancer in patients [102-104]. The LTL-313H and LTL-313B prostate cancer models used in the present study are based on grafting of patient-derived cancer tissue in the sub-renal capsule graft site and hence more closely resemble the original malignancy [109, 114]. As such, they should be more suitable for identification of prostate cancer metastasis-driving genes.

While altered expressions of TIMELESS and DLX1 have been reported for a variety of cancers [132, 134, 135, 144, 145], they have not been associated with prostate cancer metastasis.
The elevated expression of *TIMELESS* found in the metastatic xenografts (relative to its non-metastatic counterparts; Table 2.2), as well as in publicly available gene expression data of metastatic prostate cancer patients’ samples (MSKCC cohort; Fig. 2.2A), suggests that this gene has a significant role in prostate cancer metastasis. Supporting evidence for this suggestion is given by the strong correlation found between elevated *TIMELESS* expression and poor clinical patients’ outcome (increased seminal vesicle invasion and lymph node involvement; Fig. 2.2B). Disappointingly, silencing of *TIMELESS* in PC3M cells only led to limited inhibition of cell migration (Fig. 2.3D), and had no effect on metastatic properties, such as cell migration and tissue invasiveness, of PC3 and C4-2 cells (data not shown). Similarly, while elevated expression of *DLX1* was also found in clinical samples (Fig. 2.2C & D), silencing or overexpression of *DLX1* did not show any effect on *in vitro* metastatic properties of the above three prostate cancer cell lines (Fig. 2.4A-H and text). These *in vitro* data suggest that *TIMELESS* and *DLX1* may not be metastasis-driving, master regulatory genes whose changes in expression can trigger amplification of cascades of downstream metastasis-mediating genes (Fig.2.1). However, the single cell type-based assays used to assess the effects of the genes on metastatic cell properties may not be adequate [146], as it has been reported that metastasis can involve interaction of a variety of cell types, including immune cells, not present in the assay systems [147]. Moreover, the prostate cancer cells used in *in vitro* experiments are AR-independent cells in which have different genetic context compared to the AR-dependent LTL-313B and H prostate cancer tissue xenografts used in the identification step. In view of the above, further studies, particularly *in vivo*, are needed to establish whether or not *TIMELESS* or *DLX1* are prostate cancer metastasis-driving genes.
The above identification of *TIMELESS* and *DLX1* as potential prostate cancer metastasis-driving genes was solely based on changes in their expression in metastatic versus non-metastatic patient-derived prostate cancer xenografts. In the next chapter, an alternative approach for identifying metastasis-driving genes is presented.
Table 2.1: qPCR primers sequences used

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLX1 F</td>
<td>5’- TCCAGCCCCCTACATCAGTTCCGCT-3’</td>
</tr>
<tr>
<td>DLX1 R</td>
<td>5’- TTCCCTTGCCATTGAAGCCGCAC-3’</td>
</tr>
<tr>
<td>DLX1_cloningF</td>
<td>AACTCGAGATGACCATGACCACCCATGACCACCACC</td>
</tr>
<tr>
<td>DLX1_cloningR</td>
<td>AAGGATCCCTACCCTCCAGAGCCGCCCC</td>
</tr>
<tr>
<td>GAPDH F</td>
<td>5’-CACCAGGGCTGCTTTAACTC-3’</td>
</tr>
<tr>
<td>GAPDH R</td>
<td>5’-GACAAGCTTCCCCGTTCTCAG-3’</td>
</tr>
<tr>
<td>HPRT1 F</td>
<td>5’-GGTCAGGCAGTATAATCCAAAG-3’</td>
</tr>
<tr>
<td>HPRT1 R</td>
<td>5’- GGTCAGGCAGTATAATCCAAAG-3’</td>
</tr>
<tr>
<td>TIMELESS F</td>
<td>5’- AAGACTGAGGCCTCCAGCGGAG-3’</td>
</tr>
<tr>
<td>TIMELESS R</td>
<td>5’- GGCCCTGGCGACGTGTTAGGCA-3’</td>
</tr>
<tr>
<td>TUBA1B F</td>
<td>5’-GAGGTTGTGTGGATTCTCTGT-3’</td>
</tr>
<tr>
<td>TUBA1B R</td>
<td>5’-AGCTGAAATTCTGGAGCAT-3’</td>
</tr>
</tbody>
</table>

Table 2.2: Gene expression changes of TIMELESS and DLX1 in prostate cancer tissue xenografts. Elevated gene expression of TIMELESS and DLX1 in metastatic LTL-313H prostate cancer tissue xenografts compared to non-metastatic LTL-313B counterparts. Fold changes were calculated using normalized microarray reads. The data were validated by q-RT-PCR.
**Figure 2.1: Master regulatory gene in the development of metastasis.** Metastasis is thought to result from increases in the expression of specific master regulatory, metastasis-driving genes that lead to activation of cascades of downstream genes that sets the metastatic process in motion.
Figure 2.2: Elevated expression of TIMELESS and DLX1 associated with metastatic prostate cancer (PCa). A, elevated TIMELESS expression in metastatic PCa (n=19) compared to primary PCa (n=131) and normal prostate tissues (n=28; microarray gene expression data from the MSKCC Prostate Oncogenome Project). B, elevated expression of TIMELESS associated with poor patient prognosis. C, elevated DLX1 expression in both primary and metastatic tumors compared to normal prostate samples (MSKCC cohort). D, elevated DLX1 expression in 5 prostate tumors compared to 6 normal prostate samples. Note that 3 out of 6 normal samples have undetectable DLX1 levels. Statistical significance was established using the Student’s t-test; *, p < 0.05.
**A**

α- TIMELESS
α- Vinculin

**B**

Absorbance at A590

Days after transfection

siTIMELESS
siControl
Non-transfected

**C**

% of cells

Phases of cell cycle

G1
S
G2

siTIM
sicontrol
Non-transfected

**D**

% of recovered wound areas

siTIMELESS
siControl

* 

**E**

α- TIMELESS
α- Vinculin

**F**

Absorbance at A590

Days after transfection

pcDNA
pcDNA TIMELESS
Non-transfected
Figure 2.3: Analysis of **TIMELESS** function in PC3M prostate cancer cells. Treatment of PC3M cells with siTIMELESS leads to **A**, a marked decrease in TIMELESS protein levels as indicated by Western blotting. **B**, growth rate of the indicated PC3M cells was measured using MTS assay. **C**, phases of the cell cycle of the indicated PC3M cells were measured using flow cytometry. **D**, a monolayer of PC3M cells was scratched to examine the rate of cell migration into the wounded area. The bar graph represents the percentage of cell-recovered wound areas after 18 hours of incubation (*, p<0.01). **E**, Over-expression of **TIMELESS** in PC3M cells led to a marked increase in TIMELESS protein levels as indicated by western blotting. **F**, growth rate of the indicated PC3M cells was measured using MTS assay. **G**, phases of the cell cycle of the indicated PC3M cells were measured using flow cytometry. **H**, a monolayer of PC3M cells was scratched to examine the rate of cell migration into the wounded area. The bar graph represents the percentage of cell-recovered wound areas after 18 hours of incubation. Results (A-H) shown are representative of three individual experiments with error bars representing standard deviation based on triplicates. Statistical significance was established using the Student’s t-test.
A. Expression level of DLX1 (%)

B. Absorbance at A590 vs. Days after transfection

C. % of cells in different phases of cell cycle

D. % of recovered wound areas

E. Normalized DLX1 mRNA level (log10)

F. Absorbance at A590 vs. Days after transfection
**Figure 2.4: Analysis of DLX1 function in PC3M prostate cancer cells.** Treatment of PC3M cells with siDLX1 leads to A, a marked decrease in DLX1 gene expression as indicated by qRT-PCR. B, growth rate of the indicated PC3M cells was measured using MTS assay. C, phases of the cell cycle of the indicated PC3M cells were measured using flow cytometry. D, a monolayer of PC3M cells was scratched to examine the rate of cell migration into the wounded area. The bar graph represents the percentage of cell-recovered wound areas after 18 hours of incubation. E, Over-expression of DLX1 in PC3M cells led to a marked increase in DLX1 gene expression as indicated by qRT-PCR. F, growth rate of the indicated PC3M cells was measured using MTS assay. G, phases of the cell cycle of the indicated PC3M cells were measured using flow cytometry. H, a monolayer of PC3M cells was scratched to examine the rate of cell migration into the wounded area. The bar graph represents the percentage of cell-recovered wound areas after 18 hours of incubation. Results (A-H) shown are representative of three individual experiments with error bars representing standard deviation based on triplicates. Statistical significance was established using the Student’s t-test.
Chapter 3: Identification of GATA2 as a Potential Metastasis-driving Gene in Prostate Cancer

3.1 Introduction

As mentioned in Chapter 2, metastasis is thought to result from increases in the expression of specific metastasis-driving genes (master regulatory genes) leading to activation of interacting downstream genes that mediate metastasis [121, 122] (Fig. 2.1). It has long been known that master regulatory gene-induced networks often are in the form of amplification cascades, in which one upstream gene leads to activation of many downstream genes [148, 149]. In such a case, the highest differentially expressed genes in metastatic tissue (compared to non-metastatic tissue) would represent downstream genes and not upstream, metastasis-driving genes, since the latter would show smaller changes in gene expression. Recently, a promising new approach has become available for identification of master regulatory genes in amplification cascades. Thus such genes may be predicted through integrative, software-based analysis of differential gene expression profiles coupled to an increasing body of knowledge of upstream regulatory genes obtained in molecular studies, followed by experimental validation studies. Various software programs are available, such as the Upstream Regulator tool from Ingenuity Pathway Analysis [150-154] used in the following studies.

By comparative analysis of gene expression, I first identified the highest differentially expressed genes in metastatic LTL-313H xenografts, compared to their non-metastatic LTL-313B counterparts [109, 114], and then used the IPA’s Upstream Regulator Analysis tool to try to identify upstream regulators accountable for the differential expression. Six potential regulatory genes were identified, including the GATA2 gene.
**GATA2** is one of six members of the GATA transcription factor gene family characterized by their ability to bind to the DNA consensus sequence of WGATAR (in which W indicates A/T and R indicates A/G) [155]. The GATA factors constitute a family of evolutionarily conserved proteins that play key roles in the regulation of cellular differentiation [156]. The family members share two highly conserved zinc fingers of the C2H2 type that mediate not only DNA binding but also the protein-protein interaction. There is relatively little conservation outside the zinc finger domains between members of this family.

GATA2 is known as the master regulator in the development of the hematopoietic system [157, 158]. Recently the function of **GATA2** as an oncogene was investigated in several types of human cancers. For example, elevated expression of **GATA2** was found in breast cancer in which it promotes tumor growth by blocking **PTEN** expression [159]. Expression of **GATA2** was required in cancer progression of Kras-driven non-small cell lung cancer [160]. **GATA2** protein has been reported as the predominant GATA factor expressed in normal human and mouse prostates [161]. Also, it was known as a pioneer factor in the regulation of AR target gene expression [162, 163]. Bohm et.al. showed that **GATA2** may contribute to the progression to the aggressive prostate cancer through **AR** and androgen regulated genes [164]. However, the role of **GATA2** in the development of metastasis in prostate cancer has not been elucidated in detail. The finding was validated in the following studies.

### 3.2 Materials and Methods

#### 3.2.1 Materials

Chemicals, solvents and solutions were obtained from Sigma-Aldrich, Oakville, ON, Canada, unless otherwise indicated.
3.2.2 Cell Culture

LNCaP and C4-2 cells were used in the GATA2 silencing experiments due to their high endogenous expression. LNCaP cells instead of C4-2 cells were used for the experiments testing cell migration and matrigel invasion, as C4-2 cells did not migrate and invade in the in vitro experiment setting used.

LNCaP prostate cancer cells were obtained from the American Type Culture Collection (ATCC). C4-2 cells were a gift from Dr. L.W.K. Chung (Cedars-Sinai Medical Center, Los Angeles, CA). Monolayer cultures were maintained in RPMI-1640 (Gibco BRL, Gaithersburg, MD) supplemented with 10% fetal bovine serum (FBS) as previously reported [165].

3.2.3 Identification of Upstream Regulatory Genes

The gene expression microarray dataset of xenograft lines LTL-313H vs LTL-313B [114] was normalized with Z-score transformation [166]. Genes showing changes in expression (with a z-ratio > 0.5 or < -0.5) were analyzed for identification of upstream regulatory genes using the Ingenuity Upstream Regulator Analysis tool (IPA; Ingenuity Systems Inc., Redwood City, CA). The gene expression data are accessible through GEO: GSE41193 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41193).

3.2.4 Clinical Relevance Analysis

Gene expression profiles and clinical information of MSKCC prostate adenocarcinomas [41] were downloaded from the CBio Cancer Genomics Portal website [136], and correlations were sought between poor prognostic factors of the patients and the relative expression levels of GATA2 in their prostate cancer tissues. Statistical significance was established using the Student’s t-test.
3.2.5 Tissue Microarray (TMA) Construction and Immunohistochemistry

A total of 359 specimens [60 benign prostate tumors, 137 primary tumors with no lymph node metastasis, 30 primary tumors with lymph node metastasis, 65 neo-adjuvant-treated primary tumors, 67 castration-resistant prostate cancers (CRPC)] were obtained from the Vancouver Prostate Centre Tissue Bank with written informed patients’ consent and institutional study approval. All samples had been obtained through radical prostatectomy except CRPC samples that had been obtained through transurethral resection of prostate (TURP). The TMA construction has previously been described [167]. Immunohistochemical staining with polyclonal rabbit antibody against GATA2 (Cat No NBP1-82581, Novus Biological, Littleton, CO) was conducted using a Ventana autostainer (model Discover XT; Ventana Medical System, Tucson, AZ) with an enzyme-labelled biotin-streptavidin system and a solvent-resistant DAB Map kit (Ventana). Values on a four-point scale were manually assigned to each immunostaining by a pathologist. Descriptively, 0 represents no staining by any tumor cells, 1 represents a faint or focal, questionably present stain, 2 represents a stain of convincing intensity in a minority of cells and 3 a stain of convincing intensity in a majority of cells.

3.2.6 siRNA Transfection

Small interfering RNA (siRNA) targeting GATA2 (siGATA2) and negative control (scrambled) siRNAs were purchased from Dharmacon (Cat No’s J009024-07-0005 and D001810-10-05). To silence GATA2 expression in vitro, cells were transfected with siGATA2 (30 nM; 48 or 72 hours) in oligofectamin reagent (Invitrogen) following the manufacturer’s instructions.

3.2.7 Western Blotting

All cell pellets were obtained from cell cultures that supplemented with 10% FBS. Cell pellets were then lysed using RIPA lysis buffer (50 mM Tris-Cl pH 7.4, 150 mM NaCl, 1% Igepal, 0.5%
Na-deoxycholate, 0.1% SDS) supplemented with a protease inhibitor cocktail (Roche). Whole-cell lysates (20 µg), whose protein concentrations were determined using the BCA protein assay (Thermo Fisher Scientific), were run on 8% SDS polyacrylamide gel for Western blotting. The following antibodies were used: anti-GATA2 (Novus Biological), anti-c-Myc (Cat No Sc-40, Santa Cruz Biotechnology, Santa Cruz, CA), anti-UHRF1 (Cat No MABE308, Millipore, Billerica, MA), anti-BIRC5 (Cat No 2808, Cell Signaling Technology, Danvers, MA) and anti-β-actin (Cat No L002, Epitope Biotech).

### 3.2.8 MTS Cell Proliferation Assay

Cells were seeded onto 96-well culture plates (3000/well, around 30% of cell confluence) and MTS (Promega) was used to determine the cell proliferation rate following the manufacturer’s instructions. The absorbance of formazan (reduced MTS) at 490 nm was measured daily until 100% of cell confluency of control cells were reached. Statistical significance was established using the Student’s t-test.

### 3.2.9 Wound Healing Cell Migration Assay

Cells (8x10^5) that had been transfected with siGATA2 or siControl in maintenance medium were seeded onto poly-L-Lysine-coated 24-well culture plates and incubated at 37°C in a 5% CO₂ incubator; following cell attachment, the medium was changed to serum-free medium. The next day, a wound was created in the middle of a confluent cell monolayer using a pipette tip. Cell debris was removed by washing with 1xPBS (2-3 times) and the cells further incubated in RPMI medium supplemented with FBS (10%). Photographic images were taken using a Zeiss Axiovert 200M microscope (Carl Zeiss Inc.) immediately after generating the wound, and after 8 and 24 hours of further incubation [168]. The cell-recovered areas at 8 hours were measured to estimate
the extent of cell migration using Adobe Photoshop (Adobe). Statistical significance was established using the Student’s t-test.

3.2.10 Modified Boyden Chamber Assays

Migration and matrigel invasion of cells, treated with siGATA2 or siControl, were determined using modified Boyden Chambers (BD Bioscience) following the manufacturer’s instructions. After a 20-hr incubation at 37°C in a 5% CO₂ incubator, both upper and lower chambers were washed twice with 1xPBS. Dissociation buffer (300 µl; Trevigen, Gaithersburg, MD) containing calcein AM (12.5 mM; Trevigen) was added to lower chambers for a further 1-hr incubation. Calcein AM is a non-fluorescent, hydrophilic compound that easily permeates intact, live cells. The hydrolysis of Calcein AM by intracellular esterases produces calcein, a hydrophilic, strongly fluorescent compound that is well-retained in the cell cytoplasm and can be easily quantified. Fluorescence (485 nm excitation, 520 nm emission) of cell suspensions (100 ul) was determined using 96-well plates and an Infinite F500 fluorometer (Tecan, Männedorf, Switzerland). The number of cells migrated/invaded to the bottom chambers was derived from the fluorescence reading using a standard curve. Statistical significance was established using the Student’s t-test.

3.2.11 Focal Adhesion Disassembly Assay and Immunofluorescence Staining

Serum-starved (overnight) cells on cover slips were incubated with nocodazole (10 µM; Sigma) for 4 hours [169]. The cells were then washed with serum-free medium (3x) to remove the drug and the cover slips collected at various time intervals. Cells were fixed with 4% paraformaldehyde in PBS for 10 min and then permeabilized with 0.5% Triton-X100 in PBS for 10 min. For immunofluorescence staining, cells were stained with anti-vinculin (Cat No V4505, Sigma), and anti-GATA2 (Novus Biological); secondary antibodies were obtained from Jackson Immuno Research (West Grove, PA). Slides were mounted using DAPI mounting solution
(Vector Laboratories, Burlingame, CA) and viewed using an LSM 780 Confocal Microscope (Carl Zeiss Inc.).

### 3.2.12 Total RNA Isolation and Quantitative Real-Time PCR (qRT-PCR)

Total RNA was isolated from cultured cells using the RNeasy mini kit (Qiagen) following the manufacturer’s instructions. Total RNA (1 µg) was used to synthesize cDNAs using a QuantiTect Reverse Transcription Kit (Qiagen). qRT-PCR reactions using KAPA SYBR Fast Universal (Kapa Biosystems) were performed in a ViiA 7 Real-Time PCR system (Applied Biosystems). The primer sequences used can be found in the Table 3.1.

### 3.2.13 Gene Expression Data Profiling

The quality of the RNA samples was checked with the Agilent 2100 Bioanalyzer and NanoDrop ND-2000 UV-VIS spectrophotometer. Only samples with RNA Integrity Number (RIN) ≥8.0, A260/280 OD values between 1.8 and 2.0 and an A260/A230 OD value of 2.0 were used for one-color labelling using Agilent’s One-Colour Microarray-Based Gene Expression Analysis Low Input Quick Amp Labelling v6.0 (Agilent Technologies, Santa Clara, CA). Total RNA (100 ng) was used to generate cyanine-3-labelled cRNA. Four replicates from each sample group (siGATA2- or siControl-treated cells) were hybridized on Agilent SurePrint G3 Human GE 8x60K Microarray v2 (Design ID 039494). Arrays were scanned with an Agilent DNA Microarray Scanner at a 3 µm scan resolution and data were processed with Agilent Feature Extraction 11.0.1.1. Processed signals were quantile normalized with Agilent GeneSpring 12.0. The data have been deposited in NCBI’s Gene Expression Omnibus [170] and are accessible through GEO: GSE49342 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49342).
3.2.14 Gene Expression Data Analysis

Microarray gene expression data were filtered for improved quality prior to downstream analysis. Specifically, probes without corresponding gene annotations and probes without detectable expression levels (less than 3 in log2 scale) were removed. Significantly differentially expressed genes after siGATA2 treatment were selected based on the Student’s t-test with multiple test correction (FDR < 0.05) and a fold difference in mean probe expression ≥ 2.0 in the siGATA2-treated samples relative to the control samples.

3.2.15 Weighted Gene Co-expression Network Analysis (WGCNA)

In the WGCNA [171, 172] used, a gene network was first constructed by treating each gene as a node and assigning a weighted edge between each pair of nodes based on the strength of their co-expression across the MSKCC cohort of 132 primary and 18 metastatic samples as calculated by Pearson’s correlation. Correlations found were used to calculate topological overlap measure (TOM). Pairs of genes with high topological overlap were filtered. Highly inter-connected gene clusters, known as modules, were identified using unsupervised hierarchical clustering on the 1-TOM distance values with a dynamic tree-cutting process. The significance of resulting highly inter-connected gene modules was investigated in two ways. First, the module eigengene value was checked for association with clinical outcome. Second, the component genes of each module were used in gene enrichment analysis using IPA software to determine biological relevance. Statistical over-representation of functions was calculated using the Fischer's exact test and Benjamini-Hochberg (BH) multiple-test correction method, where functions with a BH-adjusted p-value <0.05 were considered significant.
3.3 Results

3.3.1 GATA2 as a Potential Upstream Metastasis-driving Gene

Using previously obtained gene expression data (GSE41193) from paired metastatic LTL-313H and non-metastatic LTL-313B prostate cancer xenografts [114], approximately 700 differentially expressed genes (with a z ratio > 0.5 or < - 0.5) were identified. Analysis of these genes using the Ingenuity Upstream Regulator Analysis tool pinpointed 18 potential upstream master regulatory genes, as shown in Table 3.2. Genes that were not expressed in prostate tissue, or genes that showed down-regulated expression in metastatic prostate cancer patients’ specimens, were excluded [41, 136]. As presented in Figure 3.1, the remaining potential prostate cancer metastasis-driving genes were: GATA2, TRIM24, MTPN, HIF1A, WT1 and EZH2. The GATA2 transcription factor gene was of particular interest since, as a pioneer factor in prostate cancer, it has a potential role in cellular reprogramming and, as such, in the development of metastasis [162, 163, 173].

3.3.2 Elevated Expression of GATA2 Correlates with Poor Prostate Cancer Prognosis

Analysis of a large scale, integrated cancer genomic dataset of the MSKCC Prostate Oncogenome Project [41] indicated that GATA2 gene expression was significantly elevated in metastatic prostate cancer samples (Fig. 3.2A). Elevated GATA2 gene expression also correlated with increased lymph node involvement, increased Gleason score and elevated PSA levels at diagnosis (p<0.05; Fig. 3.2B). As shown in Figure 3.2C, a similar correlation was found between elevated GATA2 protein levels and malignant progression of prostate cancer, as shown in clinical prostate cancer samples obtained from the Vancouver Prostate Centre with (i) increased lymph node involvement, (ii) following neo-adjuvant treatment and (iii) development of castration resistance.
3.3.3 Expression of GATA2 in Prostate Cancer Cell Lines and Tissues

Determination of GATA2 protein levels in a panel of 6 prostate cancer cell lines (LNCaP, C4-2, DU145, 22Rv1, PC3 and PC3M) and 2 immortalized prostate cell lines (BPH1 and PNT1B) showed that GATA2 expression was higher in LNCaP, C4-2 and 22Rv1 cells in comparison with BPH1 and PNT1B cells (Fig. 3.3A). However, significantly lower levels or loss of GATA2 expression was found in three other cancer cell lines, i.e. PC3, PC3M and DU145, known to lack AR expression, when compared with LNCaP, C4-2 and 22Rv1 cell lines which express mutated AR. Next, gene expression data of prostate cancer tissues (from the MSKCC Prostate Oncogenome Project) were analyzed for a correlation between AR and GATA2 gene expression. As shown in Figure 3.3B, no correlation was found.

3.3.4 GATA2 Gene Silencing Reduces in vitro Migration, Matrigel Invasion and Proliferation of Prostate Cancer Cells

siRNA-induced silencing of GATA2 gene expression in LNCaP cells led to a very marked reduction in GATA2 protein levels (Fig. 3.4A), and greatly inhibited cell proliferation (Fig. 3.4B). Similar results were found with C4-2 cells (Fig. 3.4C & D). As shown in Figure 3.5A, GATA2 silencing also substantially reduced LNCaP cell motility as revealed by a wound healing assay of 8 hours. The reduced cell motility does not appear to be a consequence of reduced cell proliferation since the doubling time of LNCaP cell cultures is about 48 hours. In addition, Boyden chamber assays showed that GATA2 silencing markedly reduced both migration (Fig. 3.5B) and tissue invasion (Fig. 3.5C) of the cells.

3.3.5 A Role for the GATA2 Gene in Focal Adhesion (FA) Disassembly

Silencing of the GATA2 gene in LNCaP cells induced a number of distinct morphological changes. The normally smooth edged, spindle-like LNCaP cells became flat and developed focal
contacts at the cells’ edges (Fig. 3.6A). Immunofluorescence-staining for vinculin, an FA protein [174], showed FA complexes in > 80% of the GATA2-silenced cells, whereas < 5% of control cells showed such complexes (Fig. 3.6B). In order to study this phenomenon, the effect of GATA2 silencing on FA disassembly was examined. Cells were treated with nocodazole, which stimulates FA formation through interfering with microtubule polymerization, followed by nocodazole washout which activates FA disassembly through re-initiation of microtubule polymerization [169]. As shown in Figure 3.6C, treatment of control cells with nocodazole induced FA formation; as early as 30 minutes after washing the cells, FA complexes were disassembled and were not further observed. In contrast, GATA2 gene-silenced cells showed persistent FA complexes even after 120 minutes of nocodazole washout. The data indicate that the GATA2 gene plays a significant role in promoting FA disassembly, a process important in cell migration and metastasis [96, 100, 101, 175].

3.3.6 Changes in Gene Expression Induced by GATA2 Silencing

Control and GATA2-silenced cells were gene expression profiled using microarray technology (GSE49342). The GATA2-silencing led to major changes in the gene expression pattern of LNCaP cells, i.e. ~1650 down-regulated genes and ~850 up-regulated genes (> 2 fold change in mRNA expression levels, FDR < 0.05; Table 3.3). As depicted in Table 3.4, the down-regulated genes included genes reported to have well-established roles in cancer: i.e., FOXM1, c-MYC, UHRF1, EZH2, BMP6, AURKA and BIRC5. The down-regulation of some of these genes was validated using qRT-PCR or Western blot analysis (Fig. 3.7A, B).

3.3.7 Functions of GATA2 as Indicated by Biostatistical Analysis

To gain greater insight into functions of the GATA2 gene in prostate cancer metastasis, a set of 970 genes was identified that showed significant changes in expression following GATA2 gene
silencing in LNCaP cells (Student’s t-test; FDR < 0.05 and FC > 1.5) and whose expressions significantly correlated with GATA2 gene expressions in the MSKCC Prostate Oncogenome Project (Pearson’s correlation; correlation coefficient > 0.30, FDR < 0.01). Next, these 970 genes were subjected to weighted gene correlation network analysis (WGCNA) to identify clusters (modules) of highly correlated genes [171, 172]. Four modules of genes with high topological overlap were identified, where each module comprised a cluster of highly inter-connected genes (Fig. 3.8). These modules were color-coded as turquoise, blue, brown, and yellow, containing 569, 245, 95, and 53 genes, respectively. Clinical relevance of these modules in the MSKCC cohort was investigated by calculating the association of each module’s eigengene value (a summary of gene expressions in that module) with clinical prostate cancer status (primary or metastatic). The yellow (Table 3.7) and brown modules (Table 3.6) were found to be highly significantly associated with prostate cancer metastasis. To investigate the biological relevance of these modules, gene function enrichment analysis was performed on the genes in each module, annotated with their expression fold-change established in the GATA2 silencing experiments. The data indicate that the brown module was significantly enriched in genes driving cell migration and tissue invasion, as such the genes of the brown module should be useful for studying metastasis in prostate cancer (Table 3.5).

3.4 Discussion

The search for prostate cancer metastasis-driving genes in this chapter was based on identification of master regulatory genes activating amplification cascades of downstream genes mediating metastasis. The finding that the candidate genes included HIF1A, WT1, and EZH2 genes (Table 3.2), reported to be associated with prostate cancer metastasis [176-178], suggests that this approach has merit (see Figure 3.1).
I focussed on *GATA2* as a potential prostate cancer metastasis-driving gene, since this gene is well known as a master regulatory gene in the hematopoietic system and has a role in tumorigenesis [156, 157]. GATA2 protein has also been identified as a pioneer factor in the expression of a number of genes [40, 162, 173]. Pioneer factors form a special class of transcription factors which through association with compacted chromatin can facilitate the recruitment of additional transcription factors. As such, they are likely to have important roles in the functioning of gene network cascades.

Evidence that *GATA2* has an important role in prostate cancer metastasis is indicated by various observations made in this study: (i) the positive correlation found between elevated expression of *GATA2* in clinical metastatic prostate cancers and a number of poor patient prognostic factors (Fig. 3.2A-C), as also indicated by others [164]; (ii) the marked reduction in cell proliferation, cell migration and matrigel invasion found following silencing of *GATA2* in LNCaP cells (Fig. 3.4B, 3.5A-C), which is supported by the biostatistical study (Table 3.5); and (iii) disruption of focal adhesion disassembly, an essential process in cell migration, in *GATA2*-silenced LNCaP cells (Fig. 3.6C) [96, 175]. The results indicate that *GATA2* is a potential metastasis-driving gene in prostate cancer. As such, it provides a potential target for preventing metastasis development in prostate cancer. Furthermore, the growth-inhibitory effect of *GATA2*-silencing indicates that *GATA2* also provides a potential target for inhibiting the growth of metastatic prostate cancer. The data are consistent with the notion that continuously elevated expression of these metastasis-driving genes is required to maintain the metastatic process.

Further evidence that *GATA2* is an important regulatory gene in prostate cancer is the finding that silencing of the *GATA2* gene in LNCaP cells led to significant changes in the expression of as many as 2400 genes (> 2 fold change, FDR < 0.05; Table 3.3). As indicated by a
study of the effects of systematic repression of individual transcription factor genes on global gene expression, such a high number of gene expression changes induced by altered expression of only one gene is rare [179].

In studying the downstream genes regulated by GATA2 in metastatic prostate cancer, WGCNA is a valuable tool for describing the correlation patterns among genes across microarray samples and identifying groups of highly correlated genes [171, 172]. Interestingly, the brown module identified (Fig. 3. 8) does not only consist of genes enriched in cell migration and tissue invasion, but also of genes whose functions are significantly correlated with metastatic prostate cancer (Table 3.5). Follow-up experiments on the genes in the brown module (Table 3.6) appear to be warranted to get further insight into the role of the GATA2 gene in prostate cancer metastasis. Furthermore, the finding that genes whose expressions were down-regulated by GATA2 silencing included FOXM1, BMP6, c-MYC, EZH2, BIRC5 and UHRF1 (Table 3.4), i.e. genes reported to have a role in prostate cancer progression and metastasis, suggests that they represent downstream genes activated by GATA2 in the development of prostate cancer metastasis.

Although GATA2 has been reported as a co-regulator of AR and its expression was found to be coupled to AR expression levels in prostate cancer cell lines (Fig. 3.3A), a similar correlation could not be found for prostate cancer tissues (MSKCC Prostate Oncogenome Project; Fig. 3.3B). Moreover, the present study did not show evidence that AR-mediated signalling in LNCaP cells was among the top pathways affected by GATA2-silencing, indicating that the AR pathway does not constitute a major pathway of GATA2 in prostate cancer metastasis. Interestingly, this finding is consistent with a recent report in which the group has shown the majority of genome-wide GATA2 binding sites did not overlap with those of AR in the LNCaP
cells [180]. Therefore, our findings reveal an AR-independent, central and complex role of GATA2 in prostate cancer development.

In conclusion, the findings of the present study indicate that the GATA2 gene represents a potential prostate cancer metastasis-driving gene which can be used as a target to inhibit the development of metastatic ability and growth of prostate cancer. Further in vivo experimental proof is needed.
<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>AURKA F</td>
<td>5’- GTCAAGTCCCTGTGCAGGTC-3’</td>
</tr>
<tr>
<td>AURKA R</td>
<td>5’- GAGTGAGACCCCTCTAGCTGT-3’</td>
</tr>
<tr>
<td>BIRC5 F</td>
<td>5’- CCTGGGTCCTCTACTGTCTTA-3’</td>
</tr>
<tr>
<td>BIRC5 R</td>
<td>5’- ATTCTGTCCTCTACATCCACCTG-3’</td>
</tr>
<tr>
<td>BMP6 F</td>
<td>5’- CCCGCTCAACCGCAAGA-3’</td>
</tr>
<tr>
<td>BMP6 R</td>
<td>5’- TCCTTGTCGACTCCACCA-3’</td>
</tr>
<tr>
<td>EZH2 F</td>
<td>5’- AGTGTGACCCCTGACCTGT-3’</td>
</tr>
<tr>
<td>EZH2 R</td>
<td>5’- AGATGGGTCCAGCAATAGAT-3’</td>
</tr>
<tr>
<td>FOXM1 F</td>
<td>5’- ATAGCAAGCGAGTCCGCATT-3’</td>
</tr>
<tr>
<td>FOXM1 R</td>
<td>5’- TCAGCTAGCAGCAGTATAAACA-3’</td>
</tr>
<tr>
<td>GAPDH F</td>
<td>5’- CACCAGGGCTGCTTTTAACTC-3’</td>
</tr>
<tr>
<td>GAPDH R</td>
<td>5’- GACAAGCGTTCCCCCTGCTGT-3’</td>
</tr>
<tr>
<td>HPRT1 F</td>
<td>5’- GGTCAGGCAGTATAATCCCAAAG-3’</td>
</tr>
<tr>
<td>HPRT1 R</td>
<td>5’- GGTCAGGCAGTATAATCCCAAAG-3’</td>
</tr>
<tr>
<td>MYC F</td>
<td>5’- GGGAGGCTATTCTGCCCCATTT-3’</td>
</tr>
<tr>
<td>MYC R</td>
<td>5’- AGGCTGCTGGTTTTCCACCTA-3’</td>
</tr>
<tr>
<td>TUBA1B F</td>
<td>5’- GAGGTTGCTGGTGGATTCTGT-3’</td>
</tr>
<tr>
<td>TUBA1B R</td>
<td>5’- AGCTGAAATTCTGGGAGCAT-3’</td>
</tr>
<tr>
<td>UHRF1 F</td>
<td>5’- GGGGGCTATGAGGATGATGGGACC-3’</td>
</tr>
<tr>
<td>UHRF1 R</td>
<td>5’- AGCAAAAGCAGTTGAGAGCCAGCG-3’</td>
</tr>
</tbody>
</table>

Table 3.1: qPCR primers sequences used
Table 3.2: Potential candidates of prostate cancer metastasis-driving genes. Activation z-score infers the activation states of predicted transcriptional regulatory genes. P value of overlap is calculated using Fisher’s Exact Test, and significance is generally attributed when p-values < 0.01.

<table>
<thead>
<tr>
<th>Transcription factor-encoding gene</th>
<th>Activation z-score</th>
<th>p-value of overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIF1A</td>
<td>3.778</td>
<td>$1.11 \times 10^{-13}$</td>
</tr>
<tr>
<td>TRIM24</td>
<td>3.148</td>
<td>$9.93 \times 10^{-05}$</td>
</tr>
<tr>
<td>EPAS1</td>
<td>2.605</td>
<td>$8.89 \times 10^{-08}$</td>
</tr>
<tr>
<td>IKZF1</td>
<td>2.216</td>
<td>$2.22 \times 10^{-01}$</td>
</tr>
<tr>
<td>MYCN</td>
<td>2.141</td>
<td>$2.26 \times 10^{-02}$</td>
</tr>
<tr>
<td>NFKB2</td>
<td>2.117</td>
<td>$2.21 \times 10^{-02}$</td>
</tr>
<tr>
<td>E2F1</td>
<td>2.116</td>
<td>$1.55 \times 10^{-01}$</td>
</tr>
<tr>
<td>WT1</td>
<td>1.985</td>
<td>$1.39 \times 10^{-02}$</td>
</tr>
<tr>
<td>GATA2</td>
<td>1.976</td>
<td>$6.38 \times 10^{-05}$</td>
</tr>
<tr>
<td>AHR</td>
<td>1.963</td>
<td>$3.21 \times 10^{-06}$</td>
</tr>
<tr>
<td>BCL3</td>
<td>1.954</td>
<td>$2.02 \times 10^{-02}$</td>
</tr>
<tr>
<td>PPARA</td>
<td>1.857</td>
<td>$4.03 \times 10^{-02}$</td>
</tr>
<tr>
<td>ARNT</td>
<td>1.752</td>
<td>$1.73 \times 10^{-02}$</td>
</tr>
<tr>
<td>STAT4</td>
<td>1.714</td>
<td>$1.82 \times 10^{-03}$</td>
</tr>
<tr>
<td>EZH2</td>
<td>1.701</td>
<td>$3.43 \times 10^{-03}$</td>
</tr>
<tr>
<td>PPARGC1B</td>
<td>1.673</td>
<td>$1.40 \times 10^{-03}$</td>
</tr>
<tr>
<td>PDX1</td>
<td>1.580</td>
<td>$1.44 \times 10^{-03}$</td>
</tr>
<tr>
<td>MTPN</td>
<td>1.508</td>
<td>$6.38 \times 10^{-08}$</td>
</tr>
</tbody>
</table>

Table 3.3: Number of genes showing expression changes following GATA2 silencing. FC: fold change; FDR: false discovery rate. FDR was calculated by the Fischer's exact test and Benjamini-Hochberg (BH) multiple-test correction method.
Table 3.4: Genes whose expression changed following GATA2 silencing. Corrected p value was calculated by the Fischer's exact test and Benjamini-Hochberg (BH) multiple-test correction method.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fold change</th>
<th>Corrected p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metastasis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMP6</td>
<td>-4.04</td>
<td>0.0078</td>
</tr>
<tr>
<td>CXCL12</td>
<td>-3.69</td>
<td>0.0288</td>
</tr>
<tr>
<td>ITGB1</td>
<td>-3.43</td>
<td>0.0117</td>
</tr>
<tr>
<td>SNAI1</td>
<td>-3.03</td>
<td>0.0159</td>
</tr>
<tr>
<td>FOXM1</td>
<td>-2.84</td>
<td>0.0042</td>
</tr>
<tr>
<td>ITGA6</td>
<td>-2.46</td>
<td>0.0023</td>
</tr>
<tr>
<td>EZH2</td>
<td>-2.24</td>
<td>0.0037</td>
</tr>
<tr>
<td>F2R</td>
<td>-2.01</td>
<td>0.0012</td>
</tr>
<tr>
<td><strong>Cell proliferation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UHRF1</td>
<td>-7.99</td>
<td>0.0002</td>
</tr>
<tr>
<td>MYC</td>
<td>-6.72</td>
<td>0.0003</td>
</tr>
<tr>
<td>BIRC5</td>
<td>-5.29</td>
<td>0.0022</td>
</tr>
<tr>
<td><strong>Cell cycle regulation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AURKB</td>
<td>-4.47</td>
<td>0.0245</td>
</tr>
<tr>
<td>AURKA</td>
<td>-4.00</td>
<td>0.0174</td>
</tr>
<tr>
<td><strong>Angiogenesis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANG</td>
<td>-4.40</td>
<td>0.0004</td>
</tr>
<tr>
<td>VEGFB</td>
<td>-2.16</td>
<td>0.0243</td>
</tr>
</tbody>
</table>
Table 3.5

<table>
<thead>
<tr>
<th>Module</th>
<th>Total gene count</th>
<th>Significance of association with metastatic prostate cancer</th>
<th>Ingenuity Pathway Analysis (IPA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Function annotation</td>
</tr>
<tr>
<td>Brown</td>
<td>95</td>
<td>$1.66 \times 10^{-14}$</td>
<td>Cell migration</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tissue invasion</td>
</tr>
<tr>
<td>Yellow</td>
<td>53</td>
<td>$1.07 \times 10^{-06}$</td>
<td>Cell proliferation</td>
</tr>
<tr>
<td>Turquoise</td>
<td>569</td>
<td>$1.80 \times 10^{-04}$</td>
<td>Activation of DNA endogenous promoter</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Organism survival</td>
</tr>
<tr>
<td>Blue</td>
<td>245</td>
<td>$8.56 \times 10^{-02}$</td>
<td>Cell cycle progression</td>
</tr>
</tbody>
</table>

Table 3.5: Module significance in prostate cancer metastasis and function annotation. P value was calculated using the Fischer's exact test and Benjamini-Hochberg (BH) multiple-test correction method.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Correlation With GATA2</th>
<th>Adjusted p value</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>LONRF3</td>
<td>-0.5132</td>
<td>0.01281</td>
<td>-4.2966</td>
</tr>
<tr>
<td>SPG20</td>
<td>-0.4943</td>
<td>0.00218</td>
<td>-4.126</td>
</tr>
<tr>
<td>LAYN</td>
<td>-0.4907</td>
<td>0.04176</td>
<td>-2.3916</td>
</tr>
<tr>
<td>MITF</td>
<td>-0.4832</td>
<td>0.01018</td>
<td>-1.9606</td>
</tr>
<tr>
<td>MYLK</td>
<td>-0.4678</td>
<td>0.0405</td>
<td>-1.5401</td>
</tr>
<tr>
<td>ACTG2</td>
<td>-0.465</td>
<td>0.02765</td>
<td>-4.1037</td>
</tr>
<tr>
<td>SYNM</td>
<td>-0.4525</td>
<td>0.01178</td>
<td>1.93265</td>
</tr>
<tr>
<td>MYL9</td>
<td>-0.4512</td>
<td>0.0194</td>
<td>-3.2335</td>
</tr>
<tr>
<td>SLC16A7</td>
<td>-0.4507</td>
<td>0.04813</td>
<td>-3.3797</td>
</tr>
<tr>
<td>MEIS1</td>
<td>-0.4465</td>
<td>0.03042</td>
<td>-1.5308</td>
</tr>
<tr>
<td>ARHGAP20</td>
<td>-0.4389</td>
<td>0.01164</td>
<td>-2.2</td>
</tr>
<tr>
<td>DENND2C</td>
<td>-0.4354</td>
<td>0.02071</td>
<td>1.55485</td>
</tr>
<tr>
<td>DNAJB4</td>
<td>-0.4285</td>
<td>0.01566</td>
<td>3.9015</td>
</tr>
<tr>
<td>C1S</td>
<td>-0.4267</td>
<td>0.03475</td>
<td>3.81342</td>
</tr>
<tr>
<td>CHRD1</td>
<td>-0.4218</td>
<td>0.00072</td>
<td>-5.0858</td>
</tr>
<tr>
<td>MET</td>
<td>-0.4134</td>
<td>0.00728</td>
<td>-5.7148</td>
</tr>
<tr>
<td>MLKL</td>
<td>-0.4119</td>
<td>0.00287</td>
<td>-3.3838</td>
</tr>
<tr>
<td>ACTA2</td>
<td>-0.4118</td>
<td>0.00468</td>
<td>-10.678</td>
</tr>
<tr>
<td>ARHGAP10</td>
<td>-0.4058</td>
<td>0.01127</td>
<td>-2.1732</td>
</tr>
<tr>
<td>SOD2</td>
<td>-0.4021</td>
<td>3.15E-05</td>
<td>-4.0496</td>
</tr>
<tr>
<td>ANP32E</td>
<td>-0.4005</td>
<td>0.04324</td>
<td>-1.7786</td>
</tr>
<tr>
<td>RND3</td>
<td>-0.3989</td>
<td>0.02146</td>
<td>-3.0221</td>
</tr>
<tr>
<td>PLA2G4A</td>
<td>-0.3988</td>
<td>0.00091</td>
<td>-2.8294</td>
</tr>
<tr>
<td>PDE4D</td>
<td>-0.3932</td>
<td>0.0006</td>
<td>-4.4232</td>
</tr>
<tr>
<td>PRDM1</td>
<td>-0.3905</td>
<td>0.00344</td>
<td>-6.9993</td>
</tr>
<tr>
<td>CRYAB</td>
<td>-0.39</td>
<td>0.02028</td>
<td>6.83396</td>
</tr>
<tr>
<td>LGALS3</td>
<td>-0.3856</td>
<td>0.00364</td>
<td>-1.9673</td>
</tr>
<tr>
<td>HIPK3</td>
<td>-0.3851</td>
<td>0.03821</td>
<td>2.58703</td>
</tr>
<tr>
<td>VCL</td>
<td>-0.3848</td>
<td>0.00281</td>
<td>2.13184</td>
</tr>
<tr>
<td>SAMD8</td>
<td>-0.3832</td>
<td>0.00423</td>
<td>2.68912</td>
</tr>
<tr>
<td>C3orf58</td>
<td>-0.3814</td>
<td>0.02718</td>
<td>-1.8246</td>
</tr>
<tr>
<td>ITGB1</td>
<td>-0.3793</td>
<td>0.00272</td>
<td>-2.9816</td>
</tr>
<tr>
<td>QSER1</td>
<td>-0.3792</td>
<td>0.00983</td>
<td>-1.9749</td>
</tr>
<tr>
<td>TGFB1R3</td>
<td>-0.3787</td>
<td>0.03328</td>
<td>-2.3847</td>
</tr>
<tr>
<td>SACS</td>
<td>-0.3785</td>
<td>0.00912</td>
<td>-2.0236</td>
</tr>
<tr>
<td>TAGLN</td>
<td>-0.3741</td>
<td>0.03545</td>
<td>3.1914</td>
</tr>
<tr>
<td>CD59</td>
<td>-0.373</td>
<td>0.04684</td>
<td>1.77779</td>
</tr>
<tr>
<td>FLJ42709</td>
<td>-0.3691</td>
<td>0.00586</td>
<td>-4.312</td>
</tr>
<tr>
<td>Gene</td>
<td>Correlation With GATA2</td>
<td>Adjusted p value</td>
<td>Fold change</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------</td>
<td>------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>CCDC88A</td>
<td>-0.3668</td>
<td>0.0076</td>
<td>-1.5277</td>
</tr>
<tr>
<td>HLF</td>
<td>-0.3666</td>
<td>0.00071</td>
<td>-3.493</td>
</tr>
<tr>
<td>CYP3A5</td>
<td>-0.3644</td>
<td>0.00244</td>
<td>6.05425</td>
</tr>
<tr>
<td>SPRY1</td>
<td>-0.3641</td>
<td>0.0102</td>
<td>-2.8494</td>
</tr>
<tr>
<td>KIT</td>
<td>-0.3611</td>
<td>0.04604</td>
<td>-1.7262</td>
</tr>
<tr>
<td>NOSTRIN</td>
<td>-0.3578</td>
<td>0.02176</td>
<td>-4.2551</td>
</tr>
<tr>
<td>NAV2</td>
<td>-0.3564</td>
<td>0.01268</td>
<td>-2.4007</td>
</tr>
<tr>
<td>KCNMA1</td>
<td>-0.3553</td>
<td>0.00656</td>
<td>-2.7596</td>
</tr>
<tr>
<td>STK33</td>
<td>-0.3443</td>
<td>0.00556</td>
<td>-2.0406</td>
</tr>
<tr>
<td>MAP7D3</td>
<td>-0.3437</td>
<td>0.00095</td>
<td>-3.2362</td>
</tr>
<tr>
<td>CYLD</td>
<td>-0.343</td>
<td>0.00347</td>
<td>2.00727</td>
</tr>
<tr>
<td>GPM6B</td>
<td>-0.3428</td>
<td>0.03126</td>
<td>-2.654</td>
</tr>
<tr>
<td>HEPH</td>
<td>-0.3423</td>
<td>0.04492</td>
<td>-1.7529</td>
</tr>
<tr>
<td>CLIC4</td>
<td>-0.3406</td>
<td>0.0059</td>
<td>-2.5422</td>
</tr>
<tr>
<td>PRNP</td>
<td>-0.3399</td>
<td>0.00038</td>
<td>-3.6836</td>
</tr>
<tr>
<td>STAC</td>
<td>-0.3388</td>
<td>0.0021</td>
<td>-11.025</td>
</tr>
<tr>
<td>TNS1</td>
<td>-0.3379</td>
<td>0.00873</td>
<td>-5.0551</td>
</tr>
<tr>
<td>CACNA1C</td>
<td>-0.3368</td>
<td>0.00037</td>
<td>9.43653</td>
</tr>
<tr>
<td>DST</td>
<td>-0.3352</td>
<td>0.01423</td>
<td>2.27821</td>
</tr>
<tr>
<td>ANO6</td>
<td>-0.3341</td>
<td>0.00429</td>
<td>-2.3702</td>
</tr>
<tr>
<td>STK38L</td>
<td>-0.3306</td>
<td>0.01121</td>
<td>-2.6554</td>
</tr>
<tr>
<td>GBP1</td>
<td>-0.3291</td>
<td>0.02816</td>
<td>-5.5096</td>
</tr>
<tr>
<td>MPZL2</td>
<td>-0.3284</td>
<td>0.00209</td>
<td>-5.4494</td>
</tr>
<tr>
<td>CD55</td>
<td>-0.3265</td>
<td>0.00271</td>
<td>1.87859</td>
</tr>
<tr>
<td>DPYSL3</td>
<td>-0.3256</td>
<td>0.00013</td>
<td>-6.0925</td>
</tr>
<tr>
<td>ACESS3</td>
<td>-0.3229</td>
<td>0.00296</td>
<td>-3.5487</td>
</tr>
<tr>
<td>GNA14</td>
<td>-0.32</td>
<td>0.00167</td>
<td>-4.8301</td>
</tr>
<tr>
<td>SLC2A9</td>
<td>-0.3194</td>
<td>0.00549</td>
<td>-2.4106</td>
</tr>
<tr>
<td>CYTIP</td>
<td>-0.3182</td>
<td>0.01603</td>
<td>-2.2473</td>
</tr>
<tr>
<td>RBMS1</td>
<td>-0.3162</td>
<td>0.04486</td>
<td>1.82284</td>
</tr>
<tr>
<td>TCEAL6</td>
<td>-0.316</td>
<td>0.04252</td>
<td>1.69499</td>
</tr>
<tr>
<td>MEF2A</td>
<td>-0.3146</td>
<td>0.02013</td>
<td>2.07927</td>
</tr>
<tr>
<td>RASGRP1</td>
<td>-0.3142</td>
<td>0.00214</td>
<td>-3.861</td>
</tr>
<tr>
<td>PTPRC</td>
<td>-0.314</td>
<td>0.00267</td>
<td>1.9423</td>
</tr>
<tr>
<td>GLT8D2</td>
<td>-0.313</td>
<td>0.00688</td>
<td>2.53407</td>
</tr>
<tr>
<td>ANXA1</td>
<td>-0.3126</td>
<td>0.00566</td>
<td>-6.4983</td>
</tr>
<tr>
<td>PHACTR2</td>
<td>-0.3118</td>
<td>0.01428</td>
<td>1.83833</td>
</tr>
<tr>
<td>CRIM1</td>
<td>-0.3113</td>
<td>0.03578</td>
<td>2.11996</td>
</tr>
<tr>
<td>HBEGF</td>
<td>-0.3102</td>
<td>0.00154</td>
<td>5.27119</td>
</tr>
<tr>
<td>Gene</td>
<td>Correlation With GATA2</td>
<td>Adjusted p value</td>
<td>Fold change</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------</td>
<td>------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>AHNAK2</td>
<td>-0.3096</td>
<td>0.0484</td>
<td>1.74005</td>
</tr>
<tr>
<td>DIO2</td>
<td>-0.3071</td>
<td>0.03029</td>
<td>1.93226</td>
</tr>
<tr>
<td>IL6ST</td>
<td>-0.3061</td>
<td>0.02396</td>
<td>1.60508</td>
</tr>
<tr>
<td>RSU1</td>
<td>-0.3056</td>
<td>0.02197</td>
<td>-2.8243</td>
</tr>
<tr>
<td>TNC</td>
<td>-0.3049</td>
<td>0.04855</td>
<td>-2.4948</td>
</tr>
<tr>
<td>ACOT9</td>
<td>-0.3016</td>
<td>0.03211</td>
<td>2.02258</td>
</tr>
<tr>
<td>ADAMTS1</td>
<td>-0.3007</td>
<td>0.00137</td>
<td>-3.948</td>
</tr>
<tr>
<td>C18orf45</td>
<td>0.32164</td>
<td>0.00095</td>
<td>-3.1684</td>
</tr>
<tr>
<td>C19orf46</td>
<td>0.32393</td>
<td>0.00032</td>
<td>1.54976</td>
</tr>
<tr>
<td>ZNF606</td>
<td>0.32755</td>
<td>0.01562</td>
<td>1.7606</td>
</tr>
<tr>
<td>DDX31</td>
<td>0.34373</td>
<td>0.04926</td>
<td>-1.5112</td>
</tr>
<tr>
<td>MTL5</td>
<td>0.35123</td>
<td>0.00015</td>
<td>3.05416</td>
</tr>
<tr>
<td>EPB41L4B</td>
<td>0.3513</td>
<td>0.0142</td>
<td>9.90451</td>
</tr>
<tr>
<td>C8orf45</td>
<td>0.35703</td>
<td>0.02044</td>
<td>-1.6138</td>
</tr>
<tr>
<td>POLE3</td>
<td>0.36401</td>
<td>0.00693</td>
<td>-2.2607</td>
</tr>
<tr>
<td>TFAP4</td>
<td>0.50794</td>
<td>0.01514</td>
<td>-2.4704</td>
</tr>
<tr>
<td>XYLB</td>
<td>0.51579</td>
<td>0.00381</td>
<td>-1.8185</td>
</tr>
<tr>
<td>SDCCAG3</td>
<td>0.53606</td>
<td>0.00257</td>
<td>-2.285</td>
</tr>
</tbody>
</table>

Table 3.6: Genes in brown module. Correlation with GATA2 was calculated using Pearson’s correlation; adjusted p value was calculated using the Fischer's exact test and Benjamini-Hochberg (BH) multiple-test correction method.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Correlation With GATA2</th>
<th>Adjusted p value</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAK1IP1</td>
<td>-0.3552</td>
<td>0.01781</td>
<td>-2.31255</td>
</tr>
<tr>
<td>SOCS2</td>
<td>-0.32147</td>
<td>0.006641</td>
<td>-1.75464</td>
</tr>
<tr>
<td>RAB27A</td>
<td>-0.31955</td>
<td>0.000489</td>
<td>-6.63441</td>
</tr>
<tr>
<td>SLC22A3</td>
<td>-0.30623</td>
<td>0.008764</td>
<td>-5.11507</td>
</tr>
<tr>
<td>ACADL</td>
<td>-0.30595</td>
<td>0.000951</td>
<td>-9.07752</td>
</tr>
<tr>
<td>HMMR</td>
<td>0.305637</td>
<td>0.001211</td>
<td>-4.07241</td>
</tr>
<tr>
<td>LAS1L</td>
<td>0.30626</td>
<td>0.008641</td>
<td>-2.53949</td>
</tr>
<tr>
<td>THOC3</td>
<td>0.307727</td>
<td>0.023492</td>
<td>-2.22384</td>
</tr>
<tr>
<td>THOC3</td>
<td>0.307727</td>
<td>0.004174</td>
<td>-2.13884</td>
</tr>
<tr>
<td>FOXM1</td>
<td>0.308731</td>
<td>0.001595</td>
<td>-3.05127</td>
</tr>
<tr>
<td>CENPF</td>
<td>0.309605</td>
<td>0.007453</td>
<td>-3.34477</td>
</tr>
<tr>
<td>CENPF</td>
<td>0.309605</td>
<td>0.02914</td>
<td>-3.08195</td>
</tr>
<tr>
<td>CCNA2</td>
<td>0.313929</td>
<td>0.000392</td>
<td>-5.7118</td>
</tr>
<tr>
<td>CENPE</td>
<td>0.314665</td>
<td>0.001397</td>
<td>-5.30428</td>
</tr>
<tr>
<td>TYMS</td>
<td>0.315522</td>
<td>0.001225</td>
<td>-6.26035</td>
</tr>
<tr>
<td>FAM111B</td>
<td>0.318428</td>
<td>0.016061</td>
<td>-4.05898</td>
</tr>
<tr>
<td>PRR11</td>
<td>0.323755</td>
<td>0.001669</td>
<td>-4.28779</td>
</tr>
<tr>
<td>DNMT1</td>
<td>0.329273</td>
<td>0.009147</td>
<td>-2.19833</td>
</tr>
<tr>
<td>KIF11</td>
<td>0.331679</td>
<td>0.000131</td>
<td>-3.84153</td>
</tr>
<tr>
<td>HIST1H4C</td>
<td>0.335718</td>
<td>0.004335</td>
<td>-1.79351</td>
</tr>
<tr>
<td>KIAA0101</td>
<td>0.336343</td>
<td>0.000676</td>
<td>-3.44118</td>
</tr>
<tr>
<td>NCAPG2</td>
<td>0.337699</td>
<td>0.000676</td>
<td>-2.30179</td>
</tr>
<tr>
<td>FANCC</td>
<td>0.339344</td>
<td>0.01212</td>
<td>-1.67444</td>
</tr>
<tr>
<td>FANCC</td>
<td>0.339344</td>
<td>0.001624</td>
<td>-2.01168</td>
</tr>
<tr>
<td>CCNB1</td>
<td>0.339526</td>
<td>0.000493</td>
<td>-2.50249</td>
</tr>
<tr>
<td>CCNB1</td>
<td>0.339526</td>
<td>0.000529</td>
<td>-3.82772</td>
</tr>
<tr>
<td>PRC1</td>
<td>0.348935</td>
<td>0.007308</td>
<td>-3.10049</td>
</tr>
<tr>
<td>TRIP13</td>
<td>0.34931</td>
<td>0.003692</td>
<td>-4.05536</td>
</tr>
<tr>
<td>TRIP13</td>
<td>0.34931</td>
<td>0.000949</td>
<td>-4.38246</td>
</tr>
<tr>
<td>RFC5</td>
<td>0.349452</td>
<td>0.002278</td>
<td>-2.4736</td>
</tr>
<tr>
<td>TPX2</td>
<td>0.349754</td>
<td>0.001619</td>
<td>-4.09663</td>
</tr>
<tr>
<td>NUSAP1</td>
<td>0.3507</td>
<td>0.000422</td>
<td>-4.0992</td>
</tr>
<tr>
<td>CKS2</td>
<td>0.35129</td>
<td>0.001251</td>
<td>-3.27055</td>
</tr>
<tr>
<td>MEST</td>
<td>0.35339</td>
<td>0.001029</td>
<td>-3.25122</td>
</tr>
<tr>
<td>CENPO</td>
<td>0.361833</td>
<td>0.000218</td>
<td>-2.69905</td>
</tr>
<tr>
<td>CENPO</td>
<td>0.361833</td>
<td>0.003485</td>
<td>-3.2309</td>
</tr>
<tr>
<td>MCM4</td>
<td>0.363412</td>
<td>0.000951</td>
<td>-3.98323</td>
</tr>
<tr>
<td>Gene</td>
<td>Correlation With GATA2</td>
<td>Adjusted p value</td>
<td>Fold change</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------</td>
<td>------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>ANLN</td>
<td>0.364401</td>
<td>0.002055</td>
<td>-4.1656</td>
</tr>
<tr>
<td>MCM3</td>
<td>0.370106</td>
<td>0.00212</td>
<td>-4.21467</td>
</tr>
<tr>
<td>SHMT1</td>
<td>0.370137</td>
<td>0.002221</td>
<td>-2.01711</td>
</tr>
<tr>
<td>SHMT1</td>
<td>0.370137</td>
<td>0.003229</td>
<td>-1.84422</td>
</tr>
<tr>
<td>NCAPD2</td>
<td>0.374036</td>
<td>0.008222</td>
<td>-3.11189</td>
</tr>
<tr>
<td>NCAPD2</td>
<td>0.374036</td>
<td>0.009301</td>
<td>-3.2239</td>
</tr>
<tr>
<td>HIST1H4L</td>
<td>0.375162</td>
<td>0.004476</td>
<td>-2.1419</td>
</tr>
<tr>
<td>TTK</td>
<td>0.376024</td>
<td>0.000661</td>
<td>-5.52477</td>
</tr>
<tr>
<td>PBK</td>
<td>0.377209</td>
<td>0.000305</td>
<td>-4.7381</td>
</tr>
<tr>
<td>WHSC1</td>
<td>0.379735</td>
<td>0.013815</td>
<td>-2.01341</td>
</tr>
<tr>
<td>WHSC1</td>
<td>0.379735</td>
<td>0.048132</td>
<td>-1.78158</td>
</tr>
<tr>
<td>EZH2</td>
<td>0.381262</td>
<td>0.003046</td>
<td>-2.09993</td>
</tr>
<tr>
<td>POLA2</td>
<td>0.386344</td>
<td>0.000263</td>
<td>-3.05461</td>
</tr>
<tr>
<td>HIST1H2AL</td>
<td>0.39701</td>
<td>0.003103</td>
<td>-4.53626</td>
</tr>
<tr>
<td>MLF1IP</td>
<td>0.400321</td>
<td>0.003485</td>
<td>-3.3317</td>
</tr>
<tr>
<td>SKP2</td>
<td>0.414971</td>
<td>0.008291</td>
<td>-3.37504</td>
</tr>
<tr>
<td>DNAJC9</td>
<td>0.427035</td>
<td>0.000279</td>
<td>-1.79793</td>
</tr>
<tr>
<td>STMN1</td>
<td>0.428373</td>
<td>0.000305</td>
<td>-3.84333</td>
</tr>
<tr>
<td>DSN1</td>
<td>0.436839</td>
<td>0.012907</td>
<td>-2.48592</td>
</tr>
<tr>
<td>ZHX3</td>
<td>0.44412</td>
<td>0.01167</td>
<td>-1.54987</td>
</tr>
<tr>
<td>HIST1H3B</td>
<td>0.45548</td>
<td>0.037235</td>
<td>-2.78946</td>
</tr>
<tr>
<td>HIST1H1B</td>
<td>0.464027</td>
<td>0.000956</td>
<td>-6.92635</td>
</tr>
<tr>
<td>UNG</td>
<td>0.479554</td>
<td>0.000564</td>
<td>-2.60195</td>
</tr>
<tr>
<td>CM5</td>
<td>0.525975</td>
<td>0.000457</td>
<td>-4.27254</td>
</tr>
<tr>
<td>MCM5</td>
<td>0.525975</td>
<td>0.00179</td>
<td>-4.58211</td>
</tr>
<tr>
<td>HIST2H3A</td>
<td>0.595562</td>
<td>0.015027</td>
<td>-5.83131</td>
</tr>
</tbody>
</table>

**Table 3.7: Genes in yellow module.** Correlation with GATA2 was calculated using Pearson’s correlation; adjusted p value was calculated using the Fischer's exact test and Benjamini-Hochberg (BH) multiple-test correction method.
Figure 3.1: A two-step strategy used in identifying potential metastasis-driving genes in prostate cancer. Differentially expressed genes with $z$ ratio of $>0.5$ or $<-0.5$ were identified by comparing gene expression profiles from paired metastatic LTL-313H and non-metastatic LTL-313B prostate cancer tissue xenografts. The differential gene expression profile was then analyzed using IPA’s Upstream Regulator Analysis tool, in combination with reported, relevant molecular data, to predict potential metastasis-driving genes. A number of potential prostate cancer metastasis-driving genes were identified including $GATA2$. 
B

<table>
<thead>
<tr>
<th>Poor prognostic factor</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metastasis</td>
<td>1.5x10^{-08}</td>
</tr>
<tr>
<td>High Gleason Score</td>
<td>0.0001</td>
</tr>
<tr>
<td>High PSA level at diagnosis</td>
<td>0.0007</td>
</tr>
<tr>
<td>Increased lymph node involvement</td>
<td>0.0010</td>
</tr>
</tbody>
</table>

C

D

Primary tumor with no LN metastasis
NHT-treated
CRPC
Figure 3.2: Elevated expression of GATA2 associated with poor prostate cancer (PCa) patient prognosis. A, GATA2 is highly expressed in 19 metastatic PCa tissues compared to 131 primary PCa samples (gene expression data from the MSKCC Prostate Oncogenome Project; *, p < 0.001). B, elevated expression of GATA2 is associated with poor patient prognosis with the indicated p value. The p values were calculated using the Student’s t-test. C, increased GATA2 immunostaining intensity was observed in advanced prostate cancers. 359 Specimens with scores 0-3 (0 represents no staining by any tumor cells; 1 represents a faint or focal, questionably present stain; 2 represents a stain of convincing intensity in a minority of cells; 3 represents a stain of convincing intensity in a majority of cells) were manually assigned to each immunostaining by a pathologist. The bar graph represents the percentage of immunostained specimens with indicated score. D, from left to right, representative images from a primary tumor with no lymph node (LN) metastasis, a neoadjuvant hormonal therapy (NHT)-treated primary tumor (1-12 months), and a castration-resistant prostate cancer (CRPC).
Figure 3.3: Expression of GATA2 in prostate cancer cell lines and tissues. A, GATA2 protein level in a panel of prostate cancer cell lines (PC3, PC3M, DU145, LNCaP, C4-2, 22Rv1) and 2 immortalized prostate cell lines (BPH1, PNT1B) as indicated in Western blotting. All cells were cultured in 10% FBS supplemented medium B, no correlation found between GATA2 and AR expression in gene expression data of 131 primary and 19 metastatic prostate cancer tissues from MSKCC Prostate Oncogenome Project (with a R-squared value of 0.085).
**Figure 3.4: GATA2 silencing decreases proliferation of prostate cancer cells.** Treatment of LNCaP and C4-2 cells with siGATA2 leads to a marked reduction in GATA2 protein levels (A and C respectively) and a marked decrease in cell proliferation (B and D respectively).
A

% of recovered wound areas

<table>
<thead>
<tr>
<th>siGATA2</th>
<th>siControl</th>
<th>Non-transfected</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

* * *

B

% of migrated cells

<table>
<thead>
<tr>
<th>siGATA2</th>
<th>siControl</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>25</td>
</tr>
</tbody>
</table>

* *

C

% of invaded cells

<table>
<thead>
<tr>
<th>siGATA2</th>
<th>siControl</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>

* *
Figure 3.5: GATA2 silencing decreases cell migration and matrigel invasion of prostate cancer cells. A, a monolayer of LNCaP cells was scratched to examine the rate of cell migration into the wounded area. The bar graph represents the percentage of cell-recovered wound areas after 8 hours of incubation (*, p<0.01). Representative images of the wound captured at different time points are shown (at bottom). B & C, cell migration and matrigel invasion assays show a marked decrease in cell motility and tissue invasiveness of siGATA2-treated LNCaP cells. Calcein AM was used to stain migrated/invaded cells following a 20-hr incubation in Boyden chambers. The number of migrated/invaded cells was derived from the fluorescence reading using a standard curve. Statistical significance was established using the Student’s t-test.
A

siGATA2

siControl

Non-transfected

B

siGATA2

siControl

Non-transfected

C

<table>
<thead>
<tr>
<th>Treatment</th>
<th>0 minute washout</th>
<th>30 minute washout</th>
<th>60 minute washout</th>
<th>120 minute washout</th>
</tr>
</thead>
<tbody>
<tr>
<td>siGATA2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>siControl</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-transfected</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.6: A role for the GATA2 gene in focal adhesion disassembly. A, following treatment with siGATA2 for 48 hours, the smooth edge rounded spindle-like LNCaP cells became flat and developed visible focal contacts (arrow in red). B, siGATA2-treated LNCaP cells showed a clear focal adhesion immunostaining pattern (arrow in white) after 48 hours of treatment. LNCaP Cells were immunostained with anti-vinculin-TRITC (red), anti-GATA2-FITC (green), and DAPI (blue). C, treatment with siGATA2 led to a failure in the disassembly of focal adhesions. Serum-starved siControl and siGATA2-treated LNCaP cells were incubated with 10 μM nocodazole for 4 hours. Cells at the indicated times after nocodazole washout were fixed and immunostained with vinculin antibody. Images were taken at 63x magnification.
Figure 3.7: Gene expression changes induced by GATA2 silencing in LNCaP cells. A number of genes were validated for gene expression changes by A, qRT-PCR and/or B, Western blot analysis after 48 hours of siGATA2 treatment.
Figure 3.8: Weighted gene co-expression network analysis (WGCNA) was used to identify genes that are potentially regulated by GATA2. Correlation expressions of genes with GATA2 were first calculated by Pearson’s correlation across the MSKCC cohort of 132 primary and 18 metastatic samples. The correlation found were next used to calculate topological overlap measure (TOM). Four highly inter-connected gene clusters (modules color-coded as turquoise, blue, brown, and yellow) were identified using unsupervised hierarchical clustering on the 1-TOM distance values with a dynamic tree-cutting process.
Chapter 4: A GATA2-based Gene Signature as a Prognostic Factor for Metastatic Progression of Early Stage Prostate Cancers

4.1 Introduction

Since the PSA blood test has become a routine procedure for prostate cancer screening of men over the age of 50, an increasing number of patients has been diagnosed with early stage (low grade, organ-confined) prostate cancer [27-29]. Most of these men are not at risk of developing metastatic prostate cancer and should be spared from unnecessary interventions leading to morbidity [181]. Conversely, men that are at risk of developing metastatic disease should be treated aggressively at this early stage of the disease to improve patients’ survival, as indicated by recent clinical trials [30, 31]. Unfortunately, there are at present no methods for reliably predicting whether a given early stage prostate cancer will become metastatic. As Gleason scores, tumor stages and serum PSA levels do not have adequate predicting ability [182-184], gene signatures, i.e. gene expression patterns, that can predict metastasis will be useful [185]. Such metastatic biomarkers may be identified by examining the molecular changes underlying the development of metastatic prostate cancer.

As described in previous studies (chapters 2 and 3), metastasis is a multi-step process of complex, interrelated events [50], whose development is generally thought to result from changes in the expression of specific, master regulatory (metastasis-driving) genes that lead to activation of cascades of downstream genes that mediate the metastatic process [121, 122]. In view of this, changes in the expressions of metastasis-driving genes and/or their downstream genes could provide potential metastatic biomarkers. In chapter 3, I identified GATA2 as a potential prostate cancer metastasis-driving gene. Accordingly, changes in the expression of GATA2 and GATA2-regulated genes could provide a gene signature for predicting metastatic
potential of early stage prostate cancers. Indeed, the expression patterns of the GATA2-regulated genes, in contrast to that of the GATA2 gene (data not shown), show potential for use as a prognostic factor.

The present study was aimed at (i) identifying GATA2-regulated genes and (ii) determining whether their expression patterns could be used as a metastatic gene signature. To this end, I analyzed gene expression data obtained from (i) GATA2-silenced LNCaP cells and (ii) the MSKCC prostate cancer cohort. This led to identification of 593 potential GATA2-regulated genes. Further analysis led to 10 genes showing potential prognostic ability. A gene signature for risk prediction of metastatic progression of early stage prostate cancers was developed based on these 10 genes and their expression in a cohort of 545 prostate cancers from the Mayo Clinic. This metastatic gene signature was validated using two independent cohorts. Its prognostic value was confirmed. In addition, it was shown to be a prognostic factor for metastasis development, independent of the widely used D’Amico prognostic classification system based on Gleason scores, tumor stages and serum PSA levels.

4.2 Materials and Methods

4.2.1 Data Sets Used

The microarray gene expression data of GATA2-silenced and control LNCaP cells have been deposited in NCBI’s Gene Expression Omnibus [170] and are accessible through GEO: GSE49342.

For evaluation of the prognostic potential of GATA2-regulated genes, their gene expressions in the MSKCC cohort (131 primary prostate cancers; endpoint: biochemical recurrence) were subjected to log-rank test analysis [41, 136].
For training of the gene signature, gene expression data of 545 prostate cancers (training set) from the Mayo Clinic (endpoint: development of regional or distant metastases) were used [186].

For validation of the gene signature, gene expression data of 235 prostate cancers (validation set) from the Mayo Clinic, and 182 prostate cancers from the Cleveland Clinic (endpoint: development of regional or distant metastases) were used. Clinical and pathological features of the human prostate cancer data sets used are presented in Table 4.

4.2.2 Pearson’s Correlation Analysis
A correlation was established by Pearson’s correlation analysis between the expressions of all individual genes (~20,000) and GATA2 in 131 primary and 19 metastatic tumor samples of the MSKCC Prostate Cancer cohort.

4.2.3 Oncomine Analysis
Oncomine, a data-mining platform based on cancer microarrays [187], was used to determine if there was a correlation between the 593 potential GATA2-regulated genes and the clinical outcomes of the prostate cancer patients. Two groups of genes were uploaded separately into the Oncomine database: (i) GATA2-positively-regulated genes, i.e. genes that showed down-regulated expression in GATA2-silenced LNCaP cells and whose expression in the MSKCC Prostate Cancer Cohort was positively correlated with GATA2 gene expression, and (ii) GATA2-negatively-regulated genes, i.e. genes that showed up-regulated expression in GATA2-silenced LNCaP cells and whose expression in the MSKCC Prostate Cancer Cohort was negatively correlated with GATA2 gene expression. Expressions of genes found associated with clinical
outcomes of prostate cancer patients (thresholds: \( p < 10^{-4} \), odds ratio > 2) were subjected to the Log-Rank test.

### 4.2.4 IPA Analysis

IPA (Ingenuity Systems Inc.) was used to identify GATA2-regulated genes that play an important role in gene networks associated with a variety of cancer and disease-related processes. The 593 GATA2-regulated genes were analyzed using IPA by uploading their changes in expression established in GATA2 silencing experiments. Expressions of genes in the top 10 networks were subjected to the Log-Rank test.

### 4.2.5 Development of a GATA2-based Gene Signature

As previously described [186], a random forest machine learning algorithm was used to assemble the selected 10 genes into a gene signature [188]. As a training set, gene expression data of 545 prostate cancers specimens from the Mayo Clinic were used (endpoint: development of regional or distant metastases).

### 4.2.6 Statistical Analysis

Log-Rank tests were conducted with GraphPad Prism 5 software (GraphPad Software, San Diego, CA). Other statistical analyses were performed using the R Project for Statistical Computing v2.14.1 package, and all tests were two-sided using a 5% significance level. The prognostic ability of the GATA2-based gene signature was determined in several cohorts using the area under the ROC curve (AUC) and univariable (UVA) logistic regression. Independent prognostic ability of the gene signature relative to D’Amico classification was compared using multivariable (MVA) logistic regression. Patients were grouped into three categories: low,
intermediate, and high risk, according to parameters used in D’Amico classification. The low risk group was used as the reference group.

Based on a majority rule criterion, the cancers with GATA2-based gene signature scores greater than 0.5 were classified as high risk, whereas those with a score lower than or equal to 0.5 were classified as low risk. Kaplan Meier survival curves were generated for the metastatic development endpoints.

4.3 Results

4.3.1 Identification of Potential GATA2-regulated Genes

GATA2-regulated genes were identified following the protocol presented in Figure 4.1, Step 1. GATA2 silencing of LNCaP cells led to marked changes in gene expression of 4602 genes, showing $\geq 1.5$ fold difference in mean probe expression and a false discovery rate (FDR) $< 0.05$ (dataset 1). These genes likely consist of genes directly and indirectly regulated by GATA2. Examination of gene expression profiles of the MSKCC prostate cancer cohort (131 primary and 18 metastatic tumor samples) revealed that the expressions of 4134 genes correlated with GATA2 expression as determined by Pearson’s correlation analysis (Pearson’s correlation coefficient $\geq 0.3$ or $\leq -0.3$; FDR $< 0.01$; dataset 2) [41]. Overlapping of these two data sets led to identification of 974 common genes. To identify potential GATA2-regulated genes in this group, the 974 genes were examined for the following correlations: (i) genes that showed down-regulated expression in GATA2-silenced LNCaP cells and whose expression in the MSKCC Prostate Cancer Cohort was positively correlated with GATA2 gene expression, and (ii) genes that showed up-regulated expression in GATA2-silenced LNCaP cells and whose expression in the MSKCC Prostate Cancer Cohort was negatively correlated with GATA2 gene expression. This led to identification of 593 genes, i.e. (i) 552 GATA2-positively-regulated genes and (ii) 41
GATA2-negatively-regulated genes. These genes likely represent a core set of GATA2-regulated genes as they showed matched expressions with GATA2 in both experimental and clinical data sets.

### 4.3.2 Clinical Outcome-driven Analysis of the GATA2-regulated Genes

The two groups of 552 GATA2-positively- and 41 GATA2-negatively-regulated genes identified above were separately uploaded to the Oncomine platform for analysis. This led to identification of 40 and 23 genes, respectively, that were associated with poor clinical outcomes of prostate cancer patients in Oncomine prostate cancer studies [(i) death at 3 and 5 years, (ii) recurrence at 5 years; see Table 4.1]. Expressions of these 63 genes were then subjected to the Log-Rank test, the top five genes with highest prognostic potential (p < 0.05) being TPX2, PRC1, NUSAP1, CRYAB and DST (Fig. 4.2).

### 4.3.3 Pathway-driven Analysis of the GATA2-regulated Genes

I next subjected the 593 GATA2-regulated genes to IPA analysis. Expressions of the 239 genes that were found in the top 10 gene networks associated with a variety of cancer-related processes (Table 4.2) were then subjected to the Log-Rank test, the top five genes with highest prognostic potential (p < 0.05) being CYP3A5, SRCAP, STMN1, XPO6 and RFC5 (Fig. 4.3).

### 4.3.4 Development of a GATA2-based Gene Signature

A GATA2-based gene signature was developed, based on a combination of five top genes (TPX2, PRC1, NUSAP1, CRYAB and DST) identified in the clinical outcome-driven analysis and five top genes (CYP3A5, SRCAP, STMN1, XPO6 and RFC5) identified in the pathway-driven analysis of the 593 GATA2-regulated genes (see Table 4.3). Gene expression data of 545 prostate cancer samples from the Mayo Clinic (Table 4) were used as a training set for the gene
signature development. Every gene in the gene signature is weighted differently to reflect the prognostic power of the corresponding gene. The final GATA2-based gene signature provides a continuous variable score ranging between 0 and 1; a high score is indicative that a cancer has a high probability for developing metastasis.

4.3.5 Performance of the GATA2-based Gene Signature in Training and Validation Data Sets
As shown in Table 4.5 and Figure 4.4, univariable logistic regression analysis indicates that the GATA2-based gene signature has prognostic power comparable to that of common clinicopathologic variables, e.g., Gleason score, PSA level and lymph node involvement. Furthermore, the area under the ROC curve (AUC) values for the GATA2-based gene signature in the training (Mayo Clinic Discovery) and validation (Mayo Clinic Validation and Klein-Cleveland Clinic) sets were 0.70, 0.67 and 0.71, respectively, indicative of a fair prognostic power of the gene signature for predicting metastatic development (Fig. 4.4).

4.3.6 The GATA2-based Gene Signature is an Independent Prognostic Variable
To determine whether the GATA2-based gene signature is an independent prognostic factor, the gene expression validation sets were subjected to multivariable analysis, adjusted by the low risk reference group (as defined by D’Amico classification). The p values obtained were < 0.05 indicating that the gene signature is an independent prognostic factor of metastatic development (Table 4.6). The odds ratio for every 0.1 increase in the score of the GATA2-based gene signature in the data sets of Mayo Clinic Validation and Klein-Cleveland Clinic was 1.71 and 3.48, respectively. Taken together, the data indicate that the GATA2-based gene signature is a prognostic factor for metastasis development, independent of the widely used D’Amico prognostic classification system.
4.3.7 High GATA2-based Gene Signature Scores Correlate with Increased Probability of Metastasis Development

The probabilities of metastasis-free survival outcomes of patients with low (≤ 0.5) and high gene signature scores (> 0.5) were compared using Kaplan-Meier analysis. Cancers with lower gene signature scores had a higher probability of metastasis-free survival (p value <0.05) in both the training and validation sets (Figure 4.5).

4.4 Discussion

Lack of reliable methods for risk assessment of metastasis development in early stage prostate cancer is a major hurdle in the management of the disease. Recently, methods based on gene expression patterns, i.e. metastatic gene signatures, have been shown to be potentially useful for predicting whether a given cancer will develop metastasis [189, 190]. However the gene signatures so far reported have not seen wide implementation in the clinic.

The approach so far used for establishing gene signatures is based on the differential expression of genes associated with the development of metastasis in primary tumors over extended periods. In the present study, I used a novel approach based on establishing whether expression patterns of a subset of specific, metastasis-related genes could be used as a metastatic gene signature for prostate cancer. As GATA2 was identified as a potential metastasis-driving gene (Chapter 3), I focused on GATA2-regulated genes, i.e. genes whose expression correlated with that of GATA2 in both experimental and clinical data sets. The ten genes (TPX2, PRC1, NUSAP1, CRYAB, DST, CYP3A5, SRCAP, STMN1, XPO6 and RFC5) used to develop the gene signature were selected from two groups of genes: (i) GATA2-regulated genes associated with poor patient outcome, and (ii) GATA2-regulated genes with a key role in cancer and disease-related pathways. As shown in Table 4.3, many of the ten genes (i.e. PRC1, NUSAP1,
STMN1, TPX2, DST) play a key role in structural organization of cytoskeleton in cells, and thus might promote cell migration. As such, they could not only serve as biomarkers, but also potentially serve as therapeutic targets for metastatic prostate cancers.

In the development and validation of the GATA2-based metastatic gene signature, use was made of large cohorts with lengthy follow-ups of prostate cancer patients. Moreover, the evaluation involved use of a more clinically meaningful endpoint, i.e. metastasis development, as distinct from surrogate endpoints such as biochemical recurrence, commonly used in other biomarker studies [8]. As such, the GATA2-based signature appears to have a good chance to provide a reliable prognostic factor for risk assessment of metastasis development in early stage prostate cancers. However, comparison of the performance of GATA2-based gene signature to currently used prognostic tools (e.g. the UCSF CAPRA, Kattan Nomograms) and previously reported gene signatures [191, 192] are needed to better illustrate the significance of our finding. Moreover, a thorough validation in larger cohorts is critical and, if successful, the GATA2-based gene signature could lead to a paradigm shift in the management of early prostate cancer.

As GATA2 appears to provide a therapeutic target for metastatic prostate cancer (see Chapter 3), the GATA2-based gene signature has potential use as a predictive biomarker for GATA2-targeting therapy.
<table>
<thead>
<tr>
<th>Identified GATA2-regulated genes</th>
<th>Oncomine concepts (gene sets) related to GATA2-regulated genes</th>
<th>Oncomine datasets</th>
<th>Odds ratio§</th>
<th>P value</th>
<th>Q value</th>
<th>Sample size</th>
<th>Number of overlapping genes between GATA2-regulated genes and Oncomine gene sets</th>
<th>Number of GATA2-regulated genes associated with poor patient outcome (overlapping genes in the same dataset)</th>
</tr>
</thead>
<tbody>
<tr>
<td>552 GATA2-positively-regulated</td>
<td>Top 10% of up-regulated genes in patients who died at 3 years</td>
<td>*Setlur prostate</td>
<td>3.5</td>
<td>2.31x10^{-13}</td>
<td>2.12x10^{-11}</td>
<td>358</td>
<td>62</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Top 10% of up-regulated genes in patients who died at 5 years</td>
<td>*Setlur prostate</td>
<td>3.8</td>
<td>1.27x10^{-15}</td>
<td>1.58x10^{-13}</td>
<td>363</td>
<td>66</td>
<td>40</td>
</tr>
<tr>
<td>41 GATA2-negatively-regulated</td>
<td>Top 10% of down-regulated genes in patients whose cancers recurred at 5 years</td>
<td>‡Taylor prostate 3</td>
<td>12.3</td>
<td>1.46x10^{-13}</td>
<td>2.87x10^{-9}</td>
<td>61</td>
<td>23</td>
<td>23</td>
</tr>
</tbody>
</table>

Table 4.1: Number of GATA2-regulated genes in prostate cancers associated with poor patient outcome. GATA2-regulated genes were uploaded onto the Oncomine platform (thresholds: p<10^{-4}, odds ratio >2); the related Oncomine concepts (gene sets) are shown in the second column. Overlapping genes between the GATA2-regulated genes and Oncomine concepts were subjected to the Log-Rank test. *Setlur prostate data set: Journal of the National Cancer Institute. 2008; 100(11):815-825. ‡Taylor prostate 3 data set: Cancer cell. 2010; 18(1):11-22. §The odds that cancers carrying the changes in expression of the listed genes have poor clinical outcome.
Table 4.2

<table>
<thead>
<tr>
<th>Cancer and disease-related processes</th>
<th>Number of GATA2-regulated genes</th>
<th>Network Score*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellular Development, Amino Acid Metabolism, Small Molecule Biochemistry</td>
<td>33</td>
<td>48</td>
</tr>
<tr>
<td>Cancer, Organismal Injury and Abnormalities, Reproductive System Disease</td>
<td>30</td>
<td>41</td>
</tr>
<tr>
<td>Cell Cycle, Reproductive System Development and Function, Organ Morphology</td>
<td>27</td>
<td>35</td>
</tr>
<tr>
<td>Cancer, Dermatological Diseases and Conditions, Hereditary Disorders</td>
<td>26</td>
<td>33</td>
</tr>
<tr>
<td>DNA Replication, Recombination, and Repair, Nucleic Acid Metabolism, Small Molecule Biochemistry</td>
<td>26</td>
<td>33</td>
</tr>
<tr>
<td>Tissue Morphology, Cellular Growth and Proliferation, Neurological Disease</td>
<td>25</td>
<td>31</td>
</tr>
<tr>
<td>Cellular Growth and Proliferation, Cell Cycle, Cancer</td>
<td>24</td>
<td>29</td>
</tr>
<tr>
<td>Cancer, Cell Death and Survival, Tumor Morphology</td>
<td>23</td>
<td>27</td>
</tr>
<tr>
<td>Free Radical Scavenging, Developmental Disorder, Immunological Disease</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Tissue Morphology, Cardiovascular Disease, Congenital Heart Anomaly</td>
<td>17</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 4.2: GATA2-regulated genes associated with cancer and disease-related processes.

239 GATA2-regulated genes were found associated with gene networks involved in a variety of cancer and disease-related processes. *The network score is the negative exponent of the right-tailed Fisher's exact test result, which calculates the likelihood that genes are identified in a network solely by random chance.
<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Entrez gene ID for human</th>
<th>Function</th>
<th>Protein type</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRYAB</td>
<td>Crystalin alpha b</td>
<td>1410</td>
<td>chaperone protein</td>
<td>other</td>
</tr>
<tr>
<td>CYP3A5</td>
<td>Cytochrome P450 3A5</td>
<td>1577</td>
<td>metabolizes dihydrotestosterone</td>
<td>enzyme</td>
</tr>
<tr>
<td>DST</td>
<td>Dystonin</td>
<td>667</td>
<td>cytoskeletal linker protein; acts as an integrator of intermediate filaments, actin and microtubules</td>
<td>other</td>
</tr>
<tr>
<td>NUSAP1</td>
<td>Nucleolar And Spindle Associated Protein 1</td>
<td>51203</td>
<td>plays a role in spindle microtubule organization</td>
<td>other</td>
</tr>
<tr>
<td>PRC1</td>
<td>Protein regulator of cytokinesis 1</td>
<td>9055</td>
<td>involved in cytokinesis</td>
<td>other</td>
</tr>
<tr>
<td>RFC5</td>
<td>Replication factor c</td>
<td>5985</td>
<td>required in the elongation of primed DNA templates by DNA polymerase delta and epsilon</td>
<td>enzyme</td>
</tr>
<tr>
<td>SRCAP</td>
<td>Snf2-related CREBBP activator protein</td>
<td>10847</td>
<td>ATPase necessary for the incorporation of the histone variant H2A.Z into nucleosomes</td>
<td>transcription regulator</td>
</tr>
<tr>
<td>STMN1</td>
<td>Stathmin</td>
<td>3925</td>
<td>prevents assembly and promotes disassembly of microtubules</td>
<td>other</td>
</tr>
<tr>
<td>TPX2</td>
<td>Targeting protein for Xklp2</td>
<td>22974</td>
<td>microtubule related; spindle assembly factor</td>
<td>other</td>
</tr>
<tr>
<td>XPO6</td>
<td>Exportin 6</td>
<td>23214</td>
<td>nuclear membrane transporter</td>
<td>transporter</td>
</tr>
</tbody>
</table>

Table 4.3: Description of the 10 genes of which the GATA2-based gene signature is composed.
Table 4.4: Clinical and pathological features of the human prostate cancer cohorts used. N/A, information was either not available or not applicable.
Table 4.5: Performance of the GATA2-based gene signature in training and validation sets.

The Wilcoxon ROC and univariable (UVA) analysis obtained for GATA2-based gene signature in the training and validation sets are shown.

<table>
<thead>
<tr>
<th>Set</th>
<th>Cohort</th>
<th>Wilcoxon (ROC) Analysis</th>
<th>UVA Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AUC</td>
<td>p value</td>
</tr>
<tr>
<td>Training</td>
<td>Mayo Clinic Discovery</td>
<td>0.70</td>
<td>1.9x10^{-15}</td>
</tr>
<tr>
<td>Validation</td>
<td>Mayo Clinic Validation</td>
<td>0.67</td>
<td>1.7x10^{-05}</td>
</tr>
<tr>
<td></td>
<td>Klein-Cleveland Clinic</td>
<td>0.71</td>
<td>1.7x10^{-05}</td>
</tr>
</tbody>
</table>

Table 4.6: The GATA2-based gene signature is an independent prognostic variable.

Multivariable (MVA) odds ratio, as well as 95% confidence interval (CI), and p value for the GATA2-based gene signature in each validation data sets are shown.

<table>
<thead>
<tr>
<th></th>
<th>MVA (adjusted by D’Amico classification)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio (95% CI)</td>
</tr>
<tr>
<td>Mayo Clinic Validation</td>
<td>1.71 (1.13-3.04)</td>
</tr>
<tr>
<td>Klein-Cleveland Clinic</td>
<td>3.48 (1.66-7.89)</td>
</tr>
</tbody>
</table>
**Figure 4.1: Study outline.**

**Step 1: Identification of potential GATA2-regulated genes**
- 4602 differentially expressed genes in GATA2-silenced LNCaP cells (FC ≥ 1.5, FDR < 0.05)
- 4134 GATA2-correlated genes in the MSKCC prostate cancer cohort (Pearson’s correlation coefficient ≥0.3 or ≤-0.3, FDR < 0.01)

**Overlap**
- 974 common genes

**Step 2: Identification of GATA2-regulated genes with prognostic potential**
- GATA2-regulated genes

**Clinical outcome-driven analysis using Oncomine Platform**
- TPX2, PRC1, NUSAP1, CRYAB, DST

**Pathway-driven analysis using Ingenuity Pathway Analysis (IPA)**
- STMN1, XPO6, RFC5, CYP3A5, SRCAP

**Step 3: Development of a gene signature**
- Based on ten genes and their expression in a cohort of 545 high risk prostate cancers from the Mayo Clinic Discovery data set

**Step 4: Validation of the GATA2-based gene signature**
- Validated using 2 independent gene expression data sets from the Mayo Clinic Validation and Klein-Cleveland Clinic

---

**85**
Figure 4.2: Five GATA2-regulated genes with potential for predicting the risk of biochemical recurrence in early stage prostate cancer. *CRYAB, DST, NUSAP1, TPX2* and *PRC1* were the top five genes identified via clinical outcome-driven analysis of GATA2-regulated genes using the Oncomine platform. Log-Rank p values are shown. Low: patients with expression lower than median expression; high: patients with expression higher than the median expression of the indicated genes.
Figure 4.3: Five GATA2-regulated genes with potential for predicting the risk of biochemical recurrence in early stage prostate cancer. *CYP3A5, RFC5, XPO6, SRCAP* and *STMN1* were the top five genes identified from pathway-driven analysis of GATA2-regulated genes using IPA. Log-Rank p values are shown. Low: patients with expression lower than median expression; high: patients with expression higher than the median expression of the indicated genes.
Figure 4.4: Forest plots of univariable analysis showing the prognostic power of GATA2-based gene signature for the development of metastasis as endpoints. GS7: Gleason score = 7; GS8+: Gleason score ≥ 8; LNI: lymph node involvement; SM: surgical margin; ECE: extracapsular extension; SVI: seminal vesicle invasion; PSA10TO20: PSA level of 10-20 ng/mL; PSA20+: PSA level of ≥ 20ng/mL.
Figure 4.5: GATA2-based gene signature used in Kaplan Meier analysis with endpoint as development of metastasis. Patients were separated into high (≥ 0.5) or low (< 0.5) risk according to their score obtained with the GATA2-based gene signature. Log-Rank p values are shown. Time to metastasis development and PCSM is measured from biochemical recurrence (BCR) in months.
Chapter 5: Conclusions

5.1 Summary of Findings

The overall objective of this doctoral study was to identify prostate cancer metastasis-driving genes that can be used (i) as potential therapeutic targets for inhibiting metastasis development and growth of prostate cancer and (ii) for the development of a gene signature to reliably predict whether a given early stage prostate cancer will become metastatic.

In chapter 2, I obtained evidence that the TIMELESS and DLX1 genes were potential prostate cancer metastasis-driving genes. Thus these genes showed elevated expression in metastatic versus non-metastatic patient-derived prostate cancer tissues and this differential expression was also found in clinical samples. However, commonly used in vitro assays for detection of metastatic ability of cells (e.g., the wound healing migration assay) showed that both silencing and overexpression of TIMELESS and DLX1 in various prostate cancer cell lines had limited or no significant effect on metastatic ability, indicating that these genes were not metastasis-driving genes. A different approach for identifying prostate cancer metastasis-driving genes was therefore needed (see below).

In chapter 3, I used a different strategy for identifying metastasis-driving genes. It is based on the notion that metastasis-driving genes may activate genes in an amplification cascade fashion, in which one upstream gene leads to activation of many metastasis-mediating downstream genes [148, 149]. In such a case, the highest differentially expressed genes in metastatic tissue (compared to non-metastatic tissue) would represent downstream genes and not upstream metastasis-driving genes, and our previous approach would not be effective. Instead, I used the integrative IPA’s Upstream Regulator Analysis tool to analyze the differential gene
expression profile of the metastatic and non-metastatic tissues and to predict the upstream master regulatory (metastasis-driving) genes accountable for the differential expression. This led to identification of the GATA2 gene as a potential prostate cancer metastasis-driving gene. This finding was supported by a correlation established between elevated GATA2 expression in metastatic prostate cancer tissues and poor patient prognosis. As well, GATA2 gene silencing led to a dramatic change in the transcriptional activity of human prostate cancer LNCaP cells which is in agreement with GATA2 acting as a master regulatory gene. Furthermore, GATA2 gene silencing led to a marked reduction in LNCaP cell migration and tissue invasion, consistent with an important role for GATA2 in the development of metastasis. As GATA2-silencing also led to marked inhibition of cell proliferation, GATA2 may not only be used as a target to inhibit metastasis-development, but also to inhibit the growth of prostate cancer.

In chapter 4, I identified a set of GATA2-regulated genes, i.e. genes that were significantly differentially expressed in GATA2-silenced LNCaP cells (Student’s t-test; FDR < 0.05 and FC > 1.5) and whose expressions in the MSKCC Prostate Cancer Cohort correlated with GATA2 gene expression (Pearson’s correlation; correlation coefficient > 0.30 or < -0.30, FDR < 0.01). I then investigated whether expression patterns of GATA2 and GATA2-regulated genes could be used as potential metastatic biomarkers (gene signatures) for prostate cancer. Indeed, the expression patterns of 10 of the GATA2-regulated genes, in contrast to that of the GATA2 gene, showed potential for use as a prognostic factor. This led to the development of a GATA2-based gene signature for risk prediction of metastatic progression of early stage prostate cancers based on these 10 genes and their expression patterns in a cohort of 545 prostate cancers from the Mayo Clinic. The metastatic gene signature was validated using two independent prostate cancer patient cohorts. Accordingly, the prognostic ability of the GATA2-based gene
signature was confirmed and was shown to be independent of the widely used D’Amico prognostic classification system. The finding that the GATA2-based gene signature has merit as a metastatic biomarker indicates that the GATA2 gene indeed is a prostate cancer metastasis-driving gene. However, in vivo studies are needed to confirm this.

5.2 Significance of the Research Findings

This doctoral study has led to findings that are potentially useful for the identification of novel therapeutic targets and metastatic biomarkers required for better management of metastatic prostate cancer. First, the identification of GATA2 as a potential metastasis-driving gene indicates that such genes can act as upstream master regulatory genes in cascade amplification networks. In such a case, these genes may be identified by analyzing the differential expression profile of metastatic versus non-metastatic patient-derived prostate cancer tissues using integrative software coupled to information on upstream regulatory genes, e.g., the IPA upstream regulator analysis tool. Second, evidence was obtained indicating that the GATA2 gene may be used as a novel target for therapy of prostate cancer; its silencing would inhibit both the development of metastasis and growth of metastatic prostate cancer. Third, the expression patterns of 10 GATA2-regulated genes led to the development of a GATA2-based gene signature which can be potentially used for risk assessment of metastasis development in early stage prostate cancers – a biomarker that is critically needed in the clinic. In conclusion, the above findings appear to be potentially useful for improved management of metastatic prostate cancer.

5.3 Ongoing Research: Silencing of GATA2 as a Potential Therapeutic Approach for Metastatic Prostate Cancer

Several GATA2-targeting therapies are being developed in our laboratory. They include (i) peptide-directed lysosomal degration of GATA2 protein [193] using a virus-based delivery
system and (ii) siRNA-mediated GATA2 silencing using a nanoparticle delivery system. If these GATA2-targeting therapies are successful in inhibiting growth and metastatic ability of prostate cancer cell lines, their effects on growth and development of metastasis in prostate cancer xenografts will be investigated. The primary limitation to targeting GATA2 is potential off-target effects, especially to the blood cells where GATA2 is highly expressed and plays an important role in hematopoietic progenic cells. One of the possible solutions is to develop drugs that can be specifically delivered to prostate cancer cells (e.g. organ specific distribution of nanoparticles).
References


