Molecular Stratification and Characterization of Clear Cell Renal Cell Carcinoma

Angela Rose Brannon

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum in Genetics and Molecular Biology.

Chapel Hill
2010

Approved by:

W. Kimryn Rathmell, MD, PhD
Adrienne Cox, PhD
William Kim, MD
Charles Perou, PhD
Kristy Richards, MD, PhD

ABSTRACT

ANGELA ROSE BRANNON: Molecular Stratification and Characterization of Clear Cell Renal Cell Carcinoma
(Under the direction of W. Kimryn Rathmell, MD, PhD)

It is estimated that there will be 58,240 new diagnoses of kidney cancer in 2010. Most cases will be clear cell renal cell carcinoma (ccRCC) and have little information as to how their disease will progress. This diversity of disease natural history is especially noteworthy in a disease so well characterized by the inactivation of the von Hippel Lindau (VHL) tumor suppressor and resulting stabilization of Hypoxia Inducible Factors (HIF). Previous studies had suggested the presence of two or more clusters in ccRCC. Based on the nonuniformity within the disease's natural progression and previous research, we hypothesized that distinct inherent molecular subclasses of ccRCC must exist and, therefore, sought to define and characterize them. In fact, two robust subtypes of ccRCC were identified, designated ccA and ccB. These subtypes are associated with survival by multivariate analysis, conferring a median survival of 8.6 years versus 2 years, respectively.

We postulated that the underlying molecular pathways within the data would explain the survival difference. ccA tumors overexpress angiogenesis, hypoxia, and metabolism pathways, common pathways characterizing ccRCC tumors. In contrast, ccB tumors overexpress more aggressive genes related to epithelial to mesenchymal transition, cell cycle, and Wnt targets. *VHL* analysis and HIF immunohistochemistry suggests that neither appear to be driving subtype differences.

To understand what is causing the differences, underlying genetic changes were analyzed. Both subtypes show deletion of chromosome 3p, location of *VHL*, in greater than 75% of tumors, corresponding with previous research and suggesting a common initiating tumorigenic event. Overall, copy number patterns look very similar between the subtypes; however, more ccB tumors show deletion of chromosomes 9 and 14, which previous studies have shown to correlate with decreased survival. Additionally, ccA tumors have mutations in a number of histone modification genes, suggesting that epigenetic modification may play a role in subtype differences.

Finally, a biomarker panel of 120 probes was defined to distinguish ccA and ccB tumors. This panel is the basis of an assay using FFPE tissue for clinical use. This assay will classify tumors into the inherent subtypes identified by this study, with prognostic impact and potentially predictive import.

This dissertation is dedicated to my father, Philip Alan Brannon, who pushed me to succeed and, more importantly, to thrive.

Acknowledgements

Thank you to the very many people that helped me to complete this dissertation project. I couldn't have done it without you.

Special notes of thanks go out to:

My advisor – Dr. W. Kimryn Rathmell, who encouraged, pushed, and shook her head, at me over the years. She also tried to teach me tact, which only partially took. Plus, she brought us cake.

My committee – Drs. Adrienne Cox, William Kim, Charles Perou, and Kristy Richards, who have given me guidance and hard questions over the years.

My coworkers, past and present - Alex Arreola, Dr. Shufen Chen, Dr. Lance Cowey, Michelle DeSimone, Kate Hacker, Leslie Kennedy, Dr. Caroline Martz Lee, Courtney McGuire, Neal Rasmussen, Dr. Christie Sanford, and Dr. Tricia Wright. These people provided scientific commentary and, more importantly, laughter over the years.

Collaborators, especially those at Rutgers - Dr. Gyan Bhanot, Dr. Anupama Reddy, Michael Seiler, and Erhan Bilal. Without these people, this dissertation would have taken far longer. Anupama, in particular, was a great pleasure to work with and helped me better understand how to use a variety of different computer programs.

People at UNC who directly helped me in my projects – Dr. Yan Shi in the Genomics and Bioinformatics core, Dr. Dominic Moore in Biostatistics, Dr. Mei Huang in the Tissue Procurement Facility, Dr. Katie Hoadley for teaching me microarray basics, Grace Silva for helping with copy number analysis, Jeremy Simon for helping with Cygwin.

Department, faculty and staff – IBMS was my first home here on campus. Thanks to Dr. Sharon Milgram for bringing me in and continuing to guide me since then. GMB then became my official curriculum. While TIBBS is not a department, they met a lot of my needs. Dr. Pat Phelps guided me to grow professionally and personally, and she made me laugh on countless occasions. Drs. Christy Ahn and Patrick Brandt were also helpful in a variety of situations. The Diversity Education Team, especially Drs. Cookie Newsom, Donna Bickford and Terri Phoenix, provided an outlet for my desire to grow in understanding diversity and sharing that with the campus. Finally, universities cannot run without administrative assistants. Special thanks to Kathy Allen, Becky Muller, Sausyty Hermreck, Pat White, and Dean Staley.

Friends and mentors at the NIH and in Bethesda – Thank you to Dr. Alison McBride for taking me into her lab while I was still extremely green and continuing to be supportive as I progress further up the ladder. Thank you also to Kerri Penrose, who taught me some tricks to finding balance and helped me remember that laughter is very important in science – important tips that have continued to serve me.

Friends – I don't even know who to begin with, especially as many have already been listed above. Dr. Tamara Moyo, who I knew before I even came here and was my first graduate mentor, also helped me through a lot of sticky situations and is an amazing cook. My IBMS class, especially Pamela Hesker, who made graduate school a supported place. My friends at home, many of whom didn't understand, but were supportive anyway. Rachel Faber Machacha, who could not provide indoor picnics or cheese trays, but did fly down more than once when I needed her support. My church families, both at Newman and Jubilee, and my Bible study group who provided support in my faith and cheering in my studies.

My soon to be wife, Jennifer D. Polley, who inadvertently almost derailed my graduate career, but then cheered me as I worked to finish it.

Family – I would not be here (on this planet or at UNC) without my parents, Phil and Rose Brannon. Thank you for your support over all of these years. Honey and Dziadzia, my mom's parents, also deserve credit for helping to raise me and instilling a determination to succeed. My aunt, Terry Zajda, and my cousin, Barry Wolcyk, round out my immediate family and took me out to shoot clay pigeons when I was home to blow off some of that excess stress that comes from research.

While I am sure that I am forgetting to list someone here, please know that you helped me through my time here at UNC, and I greatly appreciate it.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

4EBP	Eukaryotic translation initiation factor 4E-binding protein
AMPK	5' adenosine monophosphate-activated protein kinase
AJCC	American Joint Committee on Cancer
AKT	v-akt murine thymoma viral oncogene homolog 1
Ang2	Angiopoietin 2
ARNT	Aryl hydrocarbon receptor nuclear translocator or HIF-1β
BHD	Birt-Hogg-Dube
BNIP3	BCL2/adenovirus E1B 19 kDa protein-interacting protein 3
CAIX	Carbonic Anhydrase IX
cAMP	cyclic Adenosine Monophosphate
CBP/p300	Creb-binding protein/ E1A binding protein p300
ccA	clear cell Renal Cell Carcinoma, subtype A
ccB	clear cell Renal Cell Carcinoma, subtype B
CCND1	cyclin D1
CGH	comparative genomic hybridization
ccRCC	clear cell Renal Cell Carcinoma
CREB	cAMP response element binding
cT	Cycle Threshold
CXCR4	Chemokine (C-X-C motif) receptor 4
Cyclin D1	cyclin family member D1
DAPI	4',6-diamidino-2-phenylindole
DAVID	Database for Annotation Visualization and Integrated Discovery
DWD	Distance Weighted Discrimination

ECM	Extracellular matrix
EDNRB	Endothelin receptor type B
EGFR	Epithelial growth factor receptor
Egln3	Prolyl hydroxylase family member - egl nine homolog 3
EMT	Epithelial-to-Mesenchymal Transition
EPAS1	Endothelial PAS domain protein 1 or HIF-2α
ERBB2	erythroblastic leukemia viral oncogene homolog 2
FDR	False discovery rate
FDG-PET	fluorodeoxyglucose positron emission tomography
FLT1	fms-related tyrosine kinase 1
FFPE	formaldehyde-fixed, paraffin-embedded
Glut1	Glucose transporter 1
H&E	Hematoxylin and eosin
H1H2	expressing HIF1α and HIF2α
H2	expressing HIF2α only
HIF	Hypoxia Inducible Factor
HIF-1α	Hypoxia inducible factor 1 alpha
HIF-1β	Hypoxia inducible factor 2 beta or ARNT
HIF-2α	Hypoxia inducible factor 2 alpha or EPAS1
HR	Hazards ratio
HRE	Hypoxia response element
IHC	Immunohistochemistry
JNK	c-Jun N-terminal kinases
LAD	Llogical Analysis of Data
LDH	lactate dehydrogenase
LDHA	lactate dehydrogenase A

LKB1	serine/threonine kinase 11
LOO	Leave One Out
Lox	Lysl oxidase
MES	2-(N-morpholino)ethanesulfonic acid
MET	MNNG (N-Methyl-N'-nitro-N-nitroso-guanidine) HOS Tranforming gene
MMP2	Matrix metalloproteinase 2
MSKCC	Memorial Sloan-Kettering Cancer Center
mTOR	Mammalian target of rapamycin
Myc	v-myc myelocytomatosis viral oncogene homolog (avian)
Oct-4	Octamer-4 or POU5F1 (POU class 5 homeobox 1)
PAI-1	Plasminogen activator inhibitor
PCA	Principle Component Analysis
PDGF	Platelet-derived growth factor
PDGFR	Platelet-derived growth factor receptor
PDK	Pyruvate dehydrogenase kinase
PHD	Prolyl hydroxylase
PI3K	Phosphatidylinositol 3-kinases
PIP2	Phosphatidylinositol (3,4)-bisphosphate (PI(3,4)P2)
PIP3	Phosphatidylinositol (3,4,5)-trisphosphate (PI(3,4,5)P3
PTEN	Phosphatase and tensin homolog
pVHL	von Hippel-Lindau tumor suppressor protein
qPCR	quantitative real time polymerase chain reaction
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
Rbx1	Ring box protein 1
RCC	renal cell carcinoma
RGS5	Regulator of G-protein signaling 5

REDD	DNA-damage-inducible transcript 4
Rheb	Ras homolog enriched in brain
RT-PCR	Reverse transcription polymerase chain reaction
SAM	Significance Analysis of Microarrays
SAM-GSA	SAM gene set analysis
SNP	single nucleotide polymorphism
SSIGN	Stage, Size, Grade, and Necrosis
TAD	Transcriptional activation domain
TCE	Trichloroethylene
TGF	Transforming Growth Factor
Tie2	TEK tyrosine kinase, endothelial
TMA	tissue microarrays
TNM	tumor node metastasis
TSC	Tuberous sclerosis complex family members
Twist1	Twist homolog 1
UISS	UCLA integrated scoring system
VEGF	Vascular endothelial growth factor
VEGFR	Vascular endothelial growth factor receptor
VCAM1	vascular cell adhesion molecule-1
VHL	von Hippel-Lindau tumor suppressor
WT	Wildtype

Chapter One:

Introduction

This work is modified from Brannon and Rathmell, Current Oncology Reports, 2010[1].

Renal Cell Carcinoma

One in 67 people will develop kidney cancer during their lifetime[3], being the seventh leading cause of cancer in men and eighth in women in the United States[4]. In the United States alone, 2009 is estimated to bring about 57,760 new cases of kidney cancer, and this disease will cause the death of approximately 12,980 people. In the US, the average age for diagnosis is 64 and for death is 71. Men will bear this burden roughly 2 times more than women, for reasons that remain unclear. In 2008, worldwide incidence was estimated at 271,348 new cases (Figure 1.1) and 116,309 deaths[2]. Incidence rates are higher in industrialized countries, possibly due to increased life spans, better access to diagnostic equipment, and increased obesity (see risk factors below). Additionally, incidence has increased 2.9% per year from 1997-2007[3].

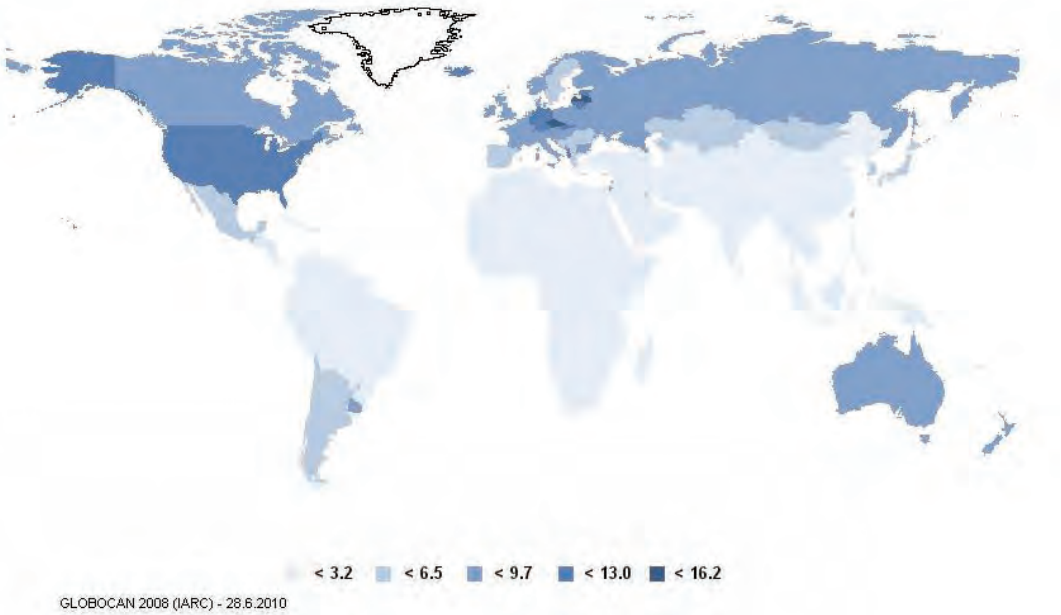

Figure 1.1 Worldwide incidence of kidney cancer in 2008.
Data calculated per country as estimated age-standardized incidence rate per 100,000 people; both sexes, all ages. Data are unavailable for Greenland. (Figure rights reserved by [2])

Fortunately, US mortality from kidney cancer has decreased 0.5% annually over that same time period[3]. This decrease is predominantly caused by cancers being detected at an earlier stage due to increased imaging capabilities.

Certain risk factors are associated with a predisposition to kidney cancer. The strongest risk factor is a family history of von Hippel Lindau (VHL) disease. Additionally, other kidney syndromes, such as cystic disease or chronic end stage renal disease increases risk. As with the majority of other cancers, cigarette smoking is a major risk factor for the development of kidney cancer, doubling the lifetime risk for heavy smokers. As is emerging for many cancers, obesity is also associated with increased incidence of kidney cancer, but decreased mortality from localized disease[5]. Being of African American descent increases risk by 2% compared to caucausians and Native Americans, while those of Hispanic origin are 2% less likely. Asian or Pacific Islander confers almost half the average risk[3]. Finally, certain occupational exposures, particularly to the organic solvent trichloroethylene (TCE), which is widely used in carpet cleaning, paint removing, and metal degreasing, can augment the probability of developing kidney cancer. Interestingly, TCE was originally used to extract vegetable oil in the 1920s, and from the 1930s-1970s, TCE was used as a general anesthetic in much of North America and Europe. In spite of these and other known risk factors most tumors arise in scenarios where an inciting factor cannot be identified.

For those patients who are diagnosed with kidney cancer, approximately 20% of them present with synchronous metastatic disease. This stage of disease confers a 10.6% five-year survival rate, with only about 25% survival at 2 years. Surgical resection of the tumor (and often the entire kidney) for those patients with organ-defined disease is the only opportunity for cure. Nevertheless, 30% of these patients go on to recur with metastatic disease after an apparently successful surgery. These tumors are also universally resistant to radiation and traditional forms of chemotherapy, and as a result,

chemotherapy is only implemented for palliation. Molecularly targeted therapies have become the common form of treatment, but while they increase progression free survival, they have not been shown conclusively to increase overall survival of patients.

Kidney cancer can be subdivided based on histological examination to grant some further information about diagnosis, progression and response. Renal cell carcinomas (RCCs) make up approximately 90% of all kidney cancers[6], but in itself encompasses a heterogeneous group of cancers. Clear cell RCC (ccRCC) is the largest histological subcategory, including 60% to 80% of cases, and will be the focus of this dissertation. Papillary and chromophobe histologies cover the majority of the other common subtypes. These stratifications represent highly dissimilar diseases and not strictly variants of RCC. Recently, an increased appreciation of the distinct biology of these subtypes has led to considerations of histology when managing these patients; however, even this major subdivision provides little immediate guidance regarding disease prognosis and management. Given this uncertainty, there is great need for both prognostic and predictive biomarkers.

Tremendous efforts have been expended in the search for reliable indicators of the underlying biology of renal carcinomas. With advancing technological opportunities to probe the genetic and molecular underpinnings of this cancer, many critical discoveries have led to major innovations in RCC, including a panel of molecularly targeted therapies which grew directly from these discoveries. Our appreciation of the genetic steps contributing to renal cancer development has been broadened, although some of the results have been surprising. However, RCC stands apart as a notoriously chemotherapy-resistant cancer that has been coaxed into submission using molecularly targeted agents that inhibit a target far removed from the inciting genetic lesion. The investigations leading to these advances are reviewed here and form a roadmap for future cancer therapeutic developments.

In the setting of these advancements, modern treatment decisions and the future of RCC drug development will benefit greatly from increased understanding of the underlying tumor biology. Tremendous gains in the treatment of this cancer remain to be made. The state-of-the-art science of RCC is a continuously evolving topic, but one that promises to provide us with valuable tools for defining the unique biology of an individual's tumor to inform predictions about recurrence or response to therapy for patient-driven clinical decisions, and to aid in the discovery of new strategies to effectively target this cancer.

Biomarkers

Before going further, it is important to first define biomarker terminology and the main categories of biomarkers that will be discussed herein. In general, a biomarker is a measurable characteristic that can be used to indicate certain physiological processes or responses.

1. Diagnostic biomarkers are used to determine whether a patient might have the disease in question. For example, a high prostate-specific antigen (PSA) measure is an indicator that a man might have prostate cancer.
2. Prognostic biomarkers provide a means to forecast the natural progression of the disease, i.e., whether a patient has a tumor associated with good survival outcome or poor survival. Clinical measures such as performance status or stage meet these criteria. Molecular measures such as Oncotype DX or MammaPrint have been approved for clinical use to predict survival outcome for breast cancer patients.

3. Predictive biomarkers assist in foretelling whether a tumor will respond to a particular treatment. For example, when a breast tumor overexpresses ERBB2, it is more likely to respond well to treatment with Herceptin.
4. There are additional biomarkers that we are unlikely to discuss fully. Risk assessment biomarkers are measures of the likelihood that a person will develop a particular disease and are generally broken down into the categories of exposure, susceptibility, and effect. Pharmacodynamic biomarkers assess the effectiveness that a drug is metabolized or hits its target to help clinicians determine which dose will be most effective for a patient, as well as what dosage might prove toxic. Pharmacogenomic biomarkers are very similar, except the biomarker tends to be expression of a gene or a particular single nucleotide polymorphism.

Prognostic Nomograms

In the absence of available molecular biomarkers, clinical measures have fulfilled the need to provide prognostic information. In fact, there are a number of prognostic scoring systems to assign risk for death to ccRCC patients already in common use based on clinical variables. An understanding of these patient stratification schemes is necessary as the field moves toward the routine incorporation of molecular biomarkers into strategies for patient stratification. For initial prognostication of risk for recurrence or death following a definitive surgical procedure, the American Joint Committee on Cancer Tumor Node Metastasis (TNM)[7], the UCLA Integrated Scoring System (UISS)[8], and the Memorial Sloan-Kettering Cancer Center (MSKCC)[9] nomograms all use clinical information including radiographic size and clinical performance status, and add in histologic information to the noninvasive clinical measures. The Mayo Clinic's Stage,

Size, Grade, and Necrosis (SSIGN) algorithm further includes tumor necrosis[10], and another nomogram from MSKCC uses all of the above and vascular invasion[11]. Thus, eventual transitions to inclusion of molecular information to the clinical scenario will be relatively straightforward once the most relevant molecular biomarkers emerge.

For prognosticating survival in the metastatic setting, a metastatic disease MSKCC score is one of the most commonly used algorithms, incorporating blood measurements of hemoglobin, serum calcium, and lactate dehydrogenase, as well as clinical evaluation of performance status and nephrectomy status[12]. A similar nomogram was identified by the Cleveland Clinic Foundation based on an independent multivariate analysis[13]. The Mayo Clinic devised a nomogram for metastatic clear cell tumors only that scored patients based on symptoms at nephrectomy, bone/liver metastases, multiple metastases, resection of all metastases, time to progression, tumor thrombus, primary tumor grade, and coagulative tumor necrosis[14]. A recent outstanding review by Isbarn and Karakiewicz[15] provides a complete overview of these nomograms, which are widely used by clinicians to provide a crude assessment of the expected survival of an individual patient.

The pVHL/HIF axis

The biology of the von Hippel-Lindau (*VHL*) gene product, pVHL, and its regulation of the hypoxia-inducible factor (HIF) family of dynamically regulated transcription factors, is indelibly linked to ccRCC biology. The discovery of the *VHL* gene, and its association with the VHL syndrome of central nervous system hemangioblastomas, pheochromocytoma, and ccRCC, in 1993[16] led almost immediately to the discovery that *VHL* mutation is tightly associated with sporadic ccRCC as well[17,18].

The *VHL* protein (pVHL) is part of an E3 ubiquitin complex, which also contains Elongin B, Elongin C, Cullin 2, and Rbx1[19]. Under physiologic conditions, pVHL recruits the hypoxia inducible factors (HIF-1α, HIF-2α, and HIF-3α variants 1-3) to the E3 ubiquitin ligase complex leading to proteasomal degradation[20-22] (Figure 1.2A). This recruitment requires the HIF-α subunits to be hydroxylated by prolyl hydroxylases (PHDs or EGLNs) on specific prolyl residues (Pro402 or Pro564) located within HIF-α's oxygen dependent domain[23-25]. In addition to oxygen, the PHDs need iron, 2-oxoglutarate, and ascorbic acid to catalyze the reaction. However, in hypoxic conditions (less than 3% oxygen) or when *VHL* is mutated, the PHDs are unable to hydroxylate HIF thereby inhibiting pVHL interaction. Xenograft studies have confirmed that restoration of pVHL expression or suppression of deregulated HIF impairs the growth of *VHL* deficient renal cell carcinoma models, verifying that *VHL* loss mediates renal cell carcinoma development via HIF deregulation[26,27].

Because of the essential role of *VHL* in RCC, the presence and type of *VHL* mutations in tumors have been consistently considered as possible biomarkers. Cowey et al.[28] recently thoroughly reviewed its potential in prognosis and prediction. Further research is still required to establish *VHL*'s efficacy as a biomarker, but given the frequency of its inactivation, more hope may lie in looking downstream.

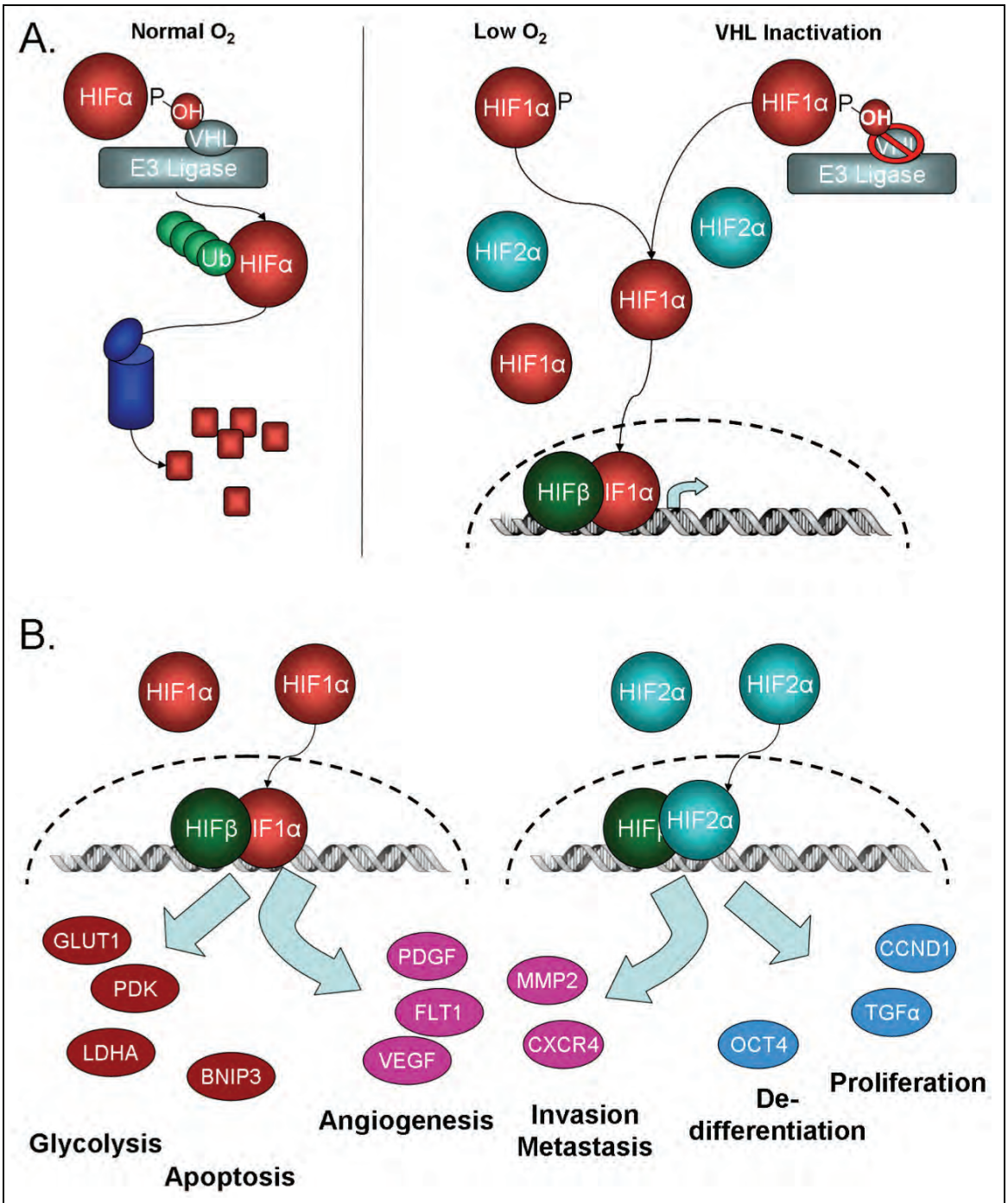

Figure 1.2 *VHL*/HIF Pathway
Diagram depicting regulation of HIF by VHL (A) and subsequent transcriptional regulation of target genes (B). Figure altered from one created by Neal Rasmussen.

When *VHL* is inactivated and HIF expression thereby stabilized, HIF-1α and HIF-2α are then available to bind HIF-1β/ARNT (aryl hydrocarbon receptor nuclear translocator) and transcriptionally activate a variety of genes (Figure 1.2B) by binding to hypoxic response elements located within the gene's promoter or enhancer[29]. Maximal activation is achieved by the additional binding of CREB (cAMP response element binding) protein (CBP) and p300[30]. HIF-3α splice variants 1-3 may also transcriptionally activate specific genes, but their targets have yet to be determined. A different splice variant, HIF-3α4, acts to dominantly negatively regulate HIF by interacting with HIF-1α, HIF-2α, and HIF-1β, and has been found to be downregulated in ccRCC[31,32].

Which of these transcriptionally activated factors or combination of factors participates in forming and maintaining the malignant phenotype of these tumors remains an open question. Certainly many hypoxia-responsive genes are outstanding candidates. One HIF target, the vascular endothelial growth factor (VEGF), has been found to be vastly upregulated in kidney tumors compared to its elevated expression in many other cancers [33,34]. This growth factor contributes to the highly vascular nature of this tumor, acting as a mitogen for tumor endothelial cells. Multiple therapeutic strategies have been developed to target VEGF, neutralizing its activity as a soluble growth factor or inhibiting the activated VEGF receptor tyrosine kinase. Remarkably, these strategies have consistently demonstrated an effect of inhibiting tumor progression and have produced therapeutic responses [35]. These breakthroughs demonstrate how much ccRCC remains dependent on key elements of HIF pathway activation, and that even if we can only target a fraction of the perturbed system, there can be tremendous clinical benefits.

In addition to VEGF, both HIF-1 and HIF-2 regulate expression of other angiogenesis genes, such as PDGF, Ang2, Flt1, and Tie2, and invasion/metastasis

genes, such as CXCR4, MMP2, Lox, and PAI-1. HIF-1 alone controls glycolytic genes such as Glut1, PGK, and LDHA, and the apoptosis gene BNIP3. In contrast, HIF-2 preferentially regulates proliferation genes, such as cyclin D1 and transforming growth factor α (TGFα), and the de-differentiation gene Oct4 (reviewed in [36]).

We have also learned that in spite of the tremendous correlation of ccRCC with loss or inactivation of *VHL*, the effect on HIF deregulation is not uniform. Variant mutations in *VHL* may contribute to imbalances of HIF1α and HIF2α deregulation leading to distinct effects on cell growth [37,38]. Renal tumors can in fact be characterized as H1H2 (expressing HIF1α and HIF2α) or H2 (expressing only HIF2α), with dramatically differing effects on tumor cell metabolism and C-myc regulation [39]. Recent evidence suggests that the H2 tumors may lose HIF1α expression as a result of nonsense or missense mutations in a subset of tumors [40], suggesting a potential selective pressure to lose the HIF1α gene during tumor progression. These insights to potentially narrow the key tumorigenic events within the *VHL*/HIF axis will undoubtedly lead to novel strategies for prognostic and therapeutic maneuvers.

Other means of regulating HIF

Interestingly, HIF can be regulated independently of pVHL[41]: In chromophobe RCC, patients with mutations in the Birt-Hogg-Dube gene overexpressed HIF. Germline MET mutations in type 1 papillary patients also overexpressed HIF proteins. Type 2 patients carry a mutation in fumarate hydratase, whereupon fumarate accumulates and binds PHDs, preventing the binding of 2-oxo-glutarate[42]. This again allows HIF to be upregulated. Accumulation of succinate due to inactivation of succinate dehydrogenase similarly prevents PHD from being able to attach a hydroxyl group to the HIF prolyl sites[43]. Additionally, HIF-1α can be phosphorylated and activated by p42/p44 MAPK[44].

Therefore, as Rathmell et al. [41] point out, even though HIF overexpression had been considered the distinguishing factor for clear cell, HIF upregulation may be a key factor in all RCC subtypes. Variance of HIF target profiles may then be important, as might previously unstudied genes and pathways, even within the bounds of clear cell RCC.

HIF levels can also increase due to increased translation caused by alterations of the mTOR (mammalian target of rapamycin) pathway. As reviewed by Dowling et al.[45], one of the ways this pathway can be activated is by the binding of insulin or a growth factor to its receptor, e.g. PDGF to PDGFR. The p85 subunit of PI3K (phosphatidylinositol-3-kinase) is phosphorylated by the kinase and inhibition of the p110 subunit is released. PI3K phosphorylates PIP2 to PIP3, which recruits Akt and PDK1 to the cell membrane. PDK1 then phosphorylates and activates Akt, which can then directly activate mTORC1 (mTOR with Raptor). Akt also inhibits the tuberous sclerosis complex (TSC), made up of TSC1 and TSC2, that normally inhibits mTOR's activating protein Rheb. The activated mTOR phosphorylates p70 S6 kinase (p70 S6K) and eIF4E binding proteins 1, 2 and 3 (4EBP). p70 S6K phosphorylates the ribosomal S6 kinase, which increases translation of mRNAs with terminal oligopyrimidine tracts, sequences that are contained within HIF1 and HIF2. The phosphorylated 4EBP releases initiation factor eIF4E, increasing translation of CAP-dependent mRNAs, including cyclins and c-Myc.

Both hypoxia and energy deprivation can downregulate this pathway through REDD1 and LKB1's activation of AMPK, respectively, which both phosphorylate TSC2. Limitations in amino acid levels are sensed by Rag GTPases, which seem to bring mTORC1 into contact with Rheb. Rapamycin, or rapamycin derivatives, can inhibit mTORC1, and therefore have become a second major category of treatments for RCC.

Cytogenetic Studies

Beyond *VHL* loss and HIF activation lies the great morass of genetic events that supplement these common molecular features to give the "teeth" to RCC. Major efforts have yet to identify a simple linear progression of genetic lesions accounting for the gains in aggressiveness in RCC. Rather, it appears that many events, most surprisingly dissimilar to other epithelial cancers, participate in this progression, discovered via both new strategies to examine the cancer genome and conventional cytogenetic studies. These studies have enhanced our understanding of the cancer genome in RCC.

Cytogenetic studies have been performed on kidney tumors since 1966[46]. Thirty years later, comparative genomic hybridization[47] (CGH) and microsatellite analysis[48] of clear cell tumors showed that that the majority (56% and 98%, respectively) of tumors had deletion within 3p, the chromosomal area where *VHL* sits. Additionally, both showed amplification of chromosome 5q (17% and 70%), and CGH identified amplification of chromosome 7. Microsatellite analysis identified other common regions of deletion as 6q, 8p, 9, and 14q, where the latter three correlated with advanced stage disease. CGH identified 9p and 13q as the most common after 3p. The CGH study also showed that increased number of chromosomal losses correlated with decreased survival, and that loss of 9p was associated with tumor recurrence. Other studies confirmed many of these regions[49-51] and validated survival association for chromosome 9[7,52-56]. Recent single nucleotide polymorphism (SNP) arrays continue to identify these same regions as important[57-59].

An interesting study performed CGH on primary ccRCC tumors and metastases attempted to understand genetics steps for metastatic progression[60]. None of the metastases were genetically identical to the primary tumors, and 32% of the metastases were completely different. Metastases to different organs also showed genetic

variability. Additionally, some metastases showed fewer genetic changes than the primary tumors. The most common genetic changes in metastatic tumors that were not present in the primary tumors were loss of 8p and 9p, and gain of 17q, 21q, and Xq. Metastases also often lacked deletion of 3p, despite the presence of the deletion in the primary tumor, a result also seen in a previous study[61].

In 2000, a group used CGH data from 116 tumors to attempt to create a disease progression model for ccRCC[62]. They put forth several branching tree models, suggesting that there are at least 2 subgroups of ccRCC. Three other groups suggested likewise: Furge et al. used gene expression data to predict cytogenetic profiles and observed two clusters in the data, associated with survival and predominantly tied to loss of 14q[6]. Arai et al. used CGH and identified two clusters, where one group had more common deletions in 1p, 4, 9, 13q, and 14q and decreased DNA methylation[63]. Most recently, Zhang et al. combined their data with 5 other groups and saw at least 2 subtypes[64]. Additionally, they created their own model for the formation and progression of ccRCC tumors.

Two important large-scale ccRCC cytogenetic studies were published within the past year. One study performed both single nucleotide polymorphism (SNP) analysis and gene expression analysis on 54 cases of sporadic ccRCC and 36 tumors from 12 patients with VHL disease [65]. Importantly, this group confirmed a widely held, but previously unproven assumption about ccRCC and VHL disease: tumors from sporadic and VHL-disease ccRCC tumors have overall similar profiles, but sporadic tumors are more heterogeneous and contain more events per tumor. In fact, unsupervised analysis of gene expression data from these two groups could not distinguish them. While this study did not identify any prognostic or predictive biomarkers, knowing that VHL disease induced ccRCC and sporadic ccRCC tumors are so similar suggest that they may be

able to be targeted with the same treatments, and that VHL disease models may faithfully mimic the more common sporadic disease.

The other study was a prospective study of 282 ccRCC patients with up to 108 months of follow-up using traditional cytogenetic karyotyping techniques[66]. They determined that loss of 3p was significantly associated with increased disease-specific survival, while loss of 4p, 9p, and 14q were significantly associated with decreased disease-specific survival. Only loss of 9p remained significant in multivariable analysis in the presence of standard clinical measures, and was further validated in an expanded study[67]. The specific genes in these regions implicated in causing the poor prognosis remain to be characterized.

In determining these individual genes associated with RCC, we turn to sequencing studies. Although whole-scale sequencing has not yet been performed on large numbers of renal carcinomas, this tumor type is being examined as a priority tumor in the cancer genome atlas and by other international efforts. Large-scale sequencing of cancer genomes is becoming more common as technology becomes better and the cost decreases. In ccRCC, the Futreal group has resequenced 3544 genes in 96 pretreatment tumors, as well as performing SNP and gene expression analyses on these tumors [40]. They then sequenced genes with at least two non-synonymous mutations in another 246 ccRCC tumors. Using a false discovery rate cutoff of 20%, the authors suggest that mutations in SETD2, JARID1C, NF2, UTX, and MLL2 have been selected for a role in cancer development or progression, opening up several interesting themes in tumor progression, particularly pertaining to the role of histone methylation. These studies will likely enhance our understanding of the steps that may permit or promote renal tumorigenesis, ultimately to the benefit of patient-centered therapy.

Gene Expression Studies

Following the overwhelming success of gene expression analyses in breast cancer[68,69], including the resulting US Food and Drug Administration–approved gene panels predictive of risk for breast cancer recurrence[70-72], it was logical to attempt similar studies in ccRCC. Gene expression studies in ccRCC studies have been numerous, but fall into four major categories: comparisons to normal tissue, comparisons between subtypes, clinically driven, and biologically driven. Table 1.1 gives an overview of the microarray gene expression studies performed in RCC.

Table 1.1 Gene expression studies in RCC

Study	Year	Samples	Analytical focus	Results
Comparisons to normal				
Boer et al.[73]	2001	37 RCC 37 normal	General tumor biology	Overexpression of cell adhesion, signal transduction, nucleotide metabolism
Gieseg et al.[74]	2002	9 clear cell 2 chromophobe 2 oncocytoma 8 normal	General tumor biology and histopathogical	355 genes compared to normal
Skubitz et al[75].	2002	8 RCC 11 normal	Diagnostic	50 genes separate RCC from normal and other diseased kidney
Lenburg et al.[76]	2003	9 clear cell 9 normal	Carcinogenesis	Identification of several oncogenes and tumor suppressors
Liou et al.[77]	2004	6 clear cell 6 normal	General tumor biology	Cell adhesion upregulated, transport downregulated
Hirota et al.[78]	2006	15 clear cell Renal cortex	Diagnostic	Laser capture microdissection provided 24 novel genes
Dalgin et al.[79]	2007	Liou, 2004	Diagnostic	158 genes that can distinguish between cc and normal
Comparisons to other histologies				
Young et al.[80]	2001	4 clear cell 1 chromophobe 2 oncocytoma	Histological	ccRCC clusters separately from chromophobe and oncocytoma
Yamazaki et al.[81]	2003	10 clear cell 2 papillary 3 chromophobe 15 normal	Histological	67 genes upregulated per group. KIT is marker for chromophobe.
Takahashi et al.[82]	2003	39 clear cell Mix of others	Histological	Distinguishing gene groups; 2 ccRCC clusters
Higgins et al.[83]	2003	23 clear cell 5 granular 4 papillary 3 chromophobe 2 oncocytoma	Histological	Cluster based on histology Granular ccRCC clusters separately from conventional ccRCC
Furge et al.[6]	2004	Takashi, 2003 33 from SMD	Histological	Transcriptional and cytogenetic classifier to distinguish subtypes; 2 ccRCC clusters
Schuetz et al.[84]	2005	13 clear cell 5 papillary 4 chromophobe 3 oncocytoma 6 angiomyeloma	Histological	Cluster by histology, with associated pathways and genes. 3 clear cell cluster with 1 papillary.
Sultmann et al.[85]	2005	65 clear cell 13 papillary 9 chromophobe 25 normal	Histological	88 genes discriminate between subtypes. cc might be 2 groups. Genes for metastases. Cytogenetic abnormalities
Rogers et al[86].	2009	7 biopsies and full tumor	Histological	Classification possible with core biopsy samples

Clinically driven analyses				
Takahashi et al.[87]	2001	29 clear cell 29 normal	5 year survival	51 probes predict for survival with 96% accuracy
Vasselli et al.[88]	2003	51 clear cell 6 papillary 1 unknown	Survival	45 genes most associated with survival. VCAM-1 alone can stratify patients by survival.
Jones et al.[89]	2005	22 clear cell 10 metastases 37 other 24 normal	Progression and metastases	31 genes that are continuously deregulated in disease progression. 155 genes that predicted metastases with 88.9% accuracy
Kosari et al.[90]	2005	10 aggr. cc 9 nonaggr. cc 9 metastatic cc 12 normal	Tumor aggressiveness	35 genes distinguish between non-aggressive and aggressive tumors. Survivin expression predicts survival by multivariate analysis in 183 patients
Yao et al.[91]	2005	28 clear cell 3 chromophobe 9 normal	Histological and survival	Genes upregulated in ccRCC vs. chromophobe/normal. ADFP correlates to survival
Zhao et al.[92]	2006	177 clear cell	Survival	259 genes associated with survival by univariate and multivariate analysis
Yao et al.[93]	2008	25 clear cell (14 metastatic) 2 metastases	Metastatic vs non-metastatic	3 genes (VCAM-1, EDNRB, RGS5) that by qRT-PCR can predict survival
Wuttig et al.[94]	2009	20 metastases	Early vs late metastasis	55 genes to predict DFI 35 genes predict few vs. many
Biology-driven analyses				
Vasselli et al. [88]	2003	51 clear cell 6 papillary 1 unknown	Unsupervised	2 clusters of metastatic tumors with survival difference
Skubitz et al.[95]	2006	16 clear cell 21 normal	Unsupervised	2 subtypes distinguishable by 546 genes, with possible pathway differences
Zhao et al.[92]	2006	177 clear cell	Unsupervised	2 clusters composed of 5 subclusters with survival difference.
Gordan et al.[39]	2008	21 clear cell	Wild-type *VHL* vs H1H2 vs H2 tumors	3 groups have distinct biological pathways. H2 tumors overexpress c-Myc, leading to increased proliferation
Zhao et al.[96]	2009	177 clear cell	Biology of survival gene set	Good prognosis tumors resemble normal renal cortex or glomerulus. Poor prognosis tumors associated with wound healing and loss of differentiation.
Brannon, et al.[97]	2010	48 clear cell 18 normal	Unsupervised consensus clustering	2 subtypes of clear cell with pathway and survival differences, differentiable by <120 probes

Aggr, aggressive; cc, Clear Cell; DFI, disease free interval; H1H2, HIF-1 and HIF-2 overexpressing; H2, HIF-2 only overexpressing; SMD, Stanford Microarray Database

Comparisons to normal tissue

The earliest gene expression analysis focused primarily on identifying the changes between RCC tumors and normal tissue in an effort to gain a better understanding of RCC tumor biology and the process of RCC carcinogenesis[73,74,76,77]. In general, these groups identified genes involved with cell adhesion and signal transduction, as well as previously identified tumor suppressors and oncogenes. A few other groups worked to identify genes that are diagnostic in nature, to distinguish the difference between clear cell and tumor[75,78,79]. Given that few biopsies are done, how distinct ccRCC is from normal, and that small growths are generally observed or ablated, diagnostic gene sets for ccRCC currently have limited utility.

Comparisons to other histologies

The next group of studies focused on genes that distinguish between the different renal cell carcinoma histologies[6,80-86]. Once a tumor is removed, pathologists have a relatively easy time differentiating ccRCC from other RCC subtypes, although occasional diagnoses of “mixed histology” or “unclassified” are used. However, this may be particularly useful for distinguishing a chromophobe tumor from an oncocytoma. Additionally, a recent study showed that core biopsies and extracted tumors had the same gene expression and that it is possible to classify a tumor based on a core biopsy using molecular markers[86]. As core or fine needle biopsies become more common, molecular markers that can identify the correct histology may become more important.

For the sake of this dissertation, several of these studies are more pertinent. Takahashi et al., Furge et al., Schuetz et al. and Sultmann et al. saw 2 clusters of ccRCC tumors within their data[6,82,84,85]. Interestingly, Schuetz et al. found that one

papillary clustered with 3 of their clear cell tumors, suggesting a vastly different expression pattern for those 3 ccRCC tumors.

Analyses focused on clinical outcomes

Supervised analyses are designed to reveal the differences among tumors based on preselected criteria, often survival, easily deriving biomarkers for the clinical characteristic of interest. In contrast, unsupervised analyses work with the data a priori and, therefore, are more likely to determine the underlying biological differences. While these biological differences may also correspond with survival or other clinical characteristics, these correlations are tangential to the original analyses; thus, these two types of analyses generate very different kinds of results.

One of the earliest studies examined 29 ccRCC tumors and identified 51 genes that could classify tumors based on 5 year disease-specific survival[87]. This study verified the possibility that gene expression profiles could be used to predict outcome, but remains to be examined in a validation study or to be defined by biological parameters which may account for this difference in disease activity. Two years later, another group examining 51 metastatic clear cell tumors identified 45 survival genes, with vascular cell adhesion molecule-1, VCAM-1, being the most predictive[88]. Since then, two retrospective studies have shown that VCAM-1 has prognostic significance[93,98]. Intriguingly, high expression of this molecule predicted for better overall survival for both clear cell and papillary histology, suggesting that VCAM-1 expression may generally indicate tumor cells with lower metastatic potential. The further implications for anti-angiogenic therapy are not yet known.

Another study described a gene signature for RCC progression, including three genes (caveolin 1, lysl oxidase, and annexin A4) that had been previously associated

with RCC aggression and/or survival[89]. A similar study concurrently identified a potential gene panel for aggressive clinical behavior in ccRCC by analysis of gene expression profiles of a set of non-aggressive (low Fuhrman grade), aggressive (mostly high Fuhrman grade), metastatic, and normal kidney samples[90]. One of these genes, Survivin, was shown to independently predict clear cell progression and risk of death[99] and, therefore, was incorporated into a new prognostic algorithm[100].

The largest study included 177 clear cell tumors and identified 340 transcripts (including VCAM-1) that could be used to assign a risk score to a patient, which was significant in multivariate analysis with stage, grade and performance status[92]. When this group later investigated the biology associated with their survival gene set, they found that tumors from patients who survived longer more resembled normal renal cortex or glomerular tissue, while poor survival patients had tumors that exhibited a wound-healing signature[96]. Further delineation and validation of pathways that contribute to tumor progression and an enhanced appreciation of the originating cell of ccRCC would be extremely useful for modeling RCC and identifying pre-cancerous changes earlier.

Biologically driven analyses

While all of the above studies performed supervised analysis, many of them[87-90,92] started with an unsupervised analysis. A common practice in array analysis is to perform unsupervised analyses to get a general understanding of the data, then move on to a supervised analysis to achieve the answers sought. Two of the unsupervised analyses from above bear further examination: The study that identified VCAM-1 as a prognostic biomarker first showed that there seemed to be two subgroups within the stage IV tumors, with possible survival differences[88]. This suggests that molecular features beyond clinical staging could provide informative data in understanding even metastatic

tumor behavior. Zhao, et al. examined their 177 tumors using 3,674 genes and saw 5 different subgroups within two larger groups of ccRCC, with significant survival differences as well as predicted biological pathway distinctions[92]. These studies helped set the stage for further delineation of subgroups within ccRCC.

In a strategy to intersect the supervised analyses with biological rationale directed toward the most studied and understood pathway in RCC, gene expression profiles were linked with von-Hippel Lindau tumor suppressor protein (pVHL) mutation analysis and expression characteristics of the of hypoxia inducible factors (HIF)[39]. In this study, 160 ccRCCs were classified as *VHL* mutant or wild type and according to HIF protein expression. *VHL* mutant, HIF1 and HIF2 expressing tumors (H1H2) overexpressed the Akt/mTOR pathway, while *VHL* mutant tumors expressing solely HIF2 (H2 tumors) replicated more rapidly, marked by overexpression of Ki-67 and activation of c-Myc signaling. While, survival data was not available for this study, other groups have identified Ki-67 as a poor-risk marker[66,101-111]. Further studies on the efficacy of HIF1 profile as a prognostic marker are anticipated.

Finally, one study stands out as being predominantly geared toward identifying the inherent subgroups and underlying biological differences of ccRCC. The Skubitz group[95] looked at 16 ccRCC tumors and saw that there seemed to be two types of clear cell, one that more highly overexpressed metabolic genes and the other extracellular matrix/cell adhesion genes.

A large number of potential biomarkers have emerged from all these gene expression studies. Encouragingly, trends are beginning to emerge between studies. The next important step will be bringing these potential biomarkers and biomarker profiles to the clinical arena, as well as better understanding the underlying biology to guide drug development.

Other Technologies

A number of other technologies have been utilized in attempting to find good prognostic biomarkers for ccRCC. Among them, we will touch briefly on tissue microarrays (TMA), plasma serum protein analysis, and microRNA profiling.

Tissue microarrays (TMA) allow for quantitative and relatively quick immunohistochemical (IHC) analysis of tumor protein expression patterns. 800 organ-confined ccRCC tumors were recently examined for expression of 15 proteins with regards to tumor stage, Fuhrman grade, and survival data[112]. Surprisingly, while pVHL and phospho-mTOR staining correlated inversely with tumor stage and grade, neither protein correlated with survival. However, expression of p27, PAX2, periostin, p-S6, and CAIX did correlate with 5 year survival. Within the intermediate stage tumors (pT2 and pT3), they found that patients with p27 and CAIX positive tumors faired better. This information could be very useful in making clinical decisions for patients in these difficult to predict categories. Many other potential biomarkers have been identified through other TMA studies, as reviewed in [113].

All of the potential biomarkers listed thus far require removal and processing of at least part of the tumor. In contrast, the use of plasma serum proteins would simply require a blood test. Plasma serum proteins have traditionally been studied to find non-invasive diagnostic markers for the presence of ccRCC as compared to normal or benign renal tissue. To date, no measurable proteins have been moved forward for screening or diagnostic evaluation. However, work from Perez-Gracia, et al, identified potential predictive biomarkers for response to sunitinib in metastatic RCC (mCC) patients[114]. Serum from patients with clinical response or progression was screened by cytokine arrays to discover that TNF-alpha and MMP-9 levels remained low in responders. Additionally, high levels of these proteins in the serum correlated with

decreased overall survival. In another study, low levels of sVEGFR-3 and VEGF-C in the serum corresponded with longer progression free survival (PFS) and objective response rate in bevacizumab-refractory mRCC[115]. A third study suggested that large changes in serum VEGF, sVEGFR-2 and sVEGFR-3 levels corresponded with tumor response[116]. All of these potential predictive biomarkers require external validation in larger sample sizes, but suggest that serum may prove to contain cogent markers of survival and response.

MicroRNA, 21-23 nucleotide segments of single-stranded non-coding RNA, have now been implicated in tumorigenesis of many cancers, even being identified as potential prognostic biomarkers in several of these. The aberrant expression of these non-coding RNAs can provide a powerful method of epigenetic tumor regulation, as an individual microRNA can alter the expression of many target genes. In RCC, various studies have identified various individual or panels of microRNAs that are differentially expressed between normal renal tissue and tumor[117-120] or between histologic subtypes[117,121]. The identification of relevant targets of these tumor associated microRNAs are just becoming realized[117,122]. microRNA is so unique compared to proteins and other small molecules, because their stem-loop structure makes them extremely stable. MicroRNAs can be easily extracted from formalin fixation, paraffin embedded tissue[123], the most common means of storing tumor tissue. Additionally, other studies have shown that microRNAs exist in repeatedly thawed and frozen samples, serum, urine, tear, ascetic fluid, and amniotic fluid[124-128]. The ability to easily use non-invasive measures to identify a stable target makes microRNAs a very attractive biomarker for diagnostic, prognostic, and predictive purposes.

Updating Nomograms

Each of the means described earlier of calculating risk for recurrence or death of disease were designed after 1999 and are well used by clinicians, yet have not included any of the large number of possible biomarkers. In 2005, Kim, *et al*, devised a prognostic model to assess patients metastatic disease that added CA9, vimentin, p53, and pTEN IHC quantification to the common measure of tumor stage and patient performance status[129]. This model had a slightly higher concordance index than did the UISS scale using clinically available parameters (0.68 vs 0.62). While not making a substantial stride in influencing prognostic accuracy, this study opened the door to hybrid nomograms which incorporate both clinical and genetic or molecular features. Table 1.2 provides a list of clinical features incorporated into commonly used algorithms that should be considered when designing hybrid nomograms.

More recently, Yao, et al, fashioned a three-gene signature of VCAM-1, EDNRB, and RGS5 to be measured by quantitative real-time PCR[93]. Their outcome prediction score could stratify patients into low, medium and high risk groups, even in metastatic disease cases. However, while the authors calculated a ROC curve to predict the specificity and sensitivity of their predictor alone and with tumor stage and grade, it remains necessary to be validated in direct comparison with a currently used algorithm. Similarly, the BioScore algorithm was formulated in 2009 based on IHC expression of B7-H1, survivin, and Ki-67[100]. The authors found that dichotomizing the expression levels of these proteins provided a c-index of 0.733, suggesting that BioScore may add prognostic value to both the UISS and SSIGN algorithms. The BioScore group presents an algorithmic model that may be beneficial for other groups to mimic: identifying patient groups not prognostically improved by the addition of BioScore data, such that only groups that would benefit were recommended for further testing. This system is

appropriate to avoid undue testing expenses, and inappropriately applying molecular information in scenarios where the additional data is uninformative.

Table 1.2 Clinical features from RCC nomograms predictive for recurrence or survival

Marker	Nomogram	Year	n	Histology	Stages
Patient characteristics					
Bone/liver metastases	Mayo [14]	2005	727	Clear cell	Metastatic
Hemoglobin	MSKCC [12]	1999	670	All	Metastatic
Multiple metastases	Mayo [14]	2005	727	Clear cell	Metastatic
Nephrectomy	MSKCC [12]	1999	670	All	Metastatic
Presence of hepatic/ pulmonary/lymph node metastases	Cleveland Clinic [13]	2005	353	All	Metastatic
Prior radiotherapy	Cleveland Clinic [13]	2005	353	All	Metastatic
Resection of metastases	Mayo [14]	2005	727	Clear cell	Metastatic
Serum calcium	MSKCC [12]	1999	670	All	Metastatic
	Cleveland Clinic [13]	2005	353	All	Metastatic
Serum LDH	MSKCC [12]	1999	670	All	Metastatic
	Cleveland Clinic [13]	2005	353	All	Metastatic
Symptoms/ performance status	MSKCC [12]	1999	670	All	Metastatic
	MSKCC [9]	2001	601	All	Localized
	UISS [8]	2001	661	All	All
	Mayo [14]	2005	727	Clear cell	Metastatic
	MSKCC [11]	2005	701	Clear cell	Localized
Time to progression	Mayo [14]	2005	727	Clear cell	Metastatic
Time to study entry	Cleveland Clinic [13]	2005	353	All	Metastatic
Tumor characteristics					
Grade	UISS [8]	2001	661	All	All
	SSIGN [10]	2002	1801	Clear cell	All
	Mayo [14]	2005	727	Clear cell	Metastatic
	MSKCC [11]	2005	701	Clear cell	Localized
Histology	MSKCC [9]	2001	601	All	Localized
Microvascular invasion	MSKCC [11]	2005	701	Clear cell	Localized
TNM stage	AJCC [130]	2005	1065	All	Localized
	MSKCC [9]	2001	601	All	Localized
	UISS [8]	2001	661	All	All
	SSIGN [10]	2002	1801	Clear cell	All
	MSKCC [11]	2005	701	Clear cell	Localized
Tumor necrosis	SSIGN [10]	2002	1801	Clear cell	All
	Mayo [14]	2005	727	Clear cell	Metastatic
	MSKCC [11]	2005	701	Clear cell	Localized
Tumor size	MSKCC [9]	2001	601	All	Localized
	SSIGN [10]	2002	1801	Clear cell	All
	MSKCC [11]	2005	701	Clear cell	Localized
Tumor thrombus	Mayo [14]	2005	727	Clear cell	Metastatic

AJCC—American Joint Committee on Cancer; LDH—lactate dehydrogenase; MSKCC—Memorial Sloan-Kettering Cancer Center; RCC—renal cell carcinoma; SSIGN—Stage, Size, Grade, and Necrosis; TNM—tumor node metastasis; UISS—UCLA integrated scoring system.

While all of the above biomarker algorithms may enhance prognostic ability, they lack the ability to address underlying tumor biology. The Pantuck group has begun to address tumor biology by developing a nomogram with a c-index of 0.89 which includes TNM staging, Fuhrman grade, and loss of chromosome 9p[66]. The incorporation of biological information into existing nomogram strategies for clinical prognostication or prediction of response to therapy is clearly not trivial. However, neither clinical data, nor biological information, can be treated in isolation. Both are relevant to patient care and patient outcomes. The future success of biomarker programs will take a considered approach to modifying existing algorithms or developing new hybrid algorithms based on large scale multivariate analysis.

Summary

In the last decade, great strides have been made for RCC patients with regard to earlier diagnoses, development of new treatment options, providing better prognostic information, and beginning work on predictive biomarkers. Many challenges remain: most of the new prognostic algorithms still require independent validation, ideally in prospective studies. The large number of biomarkers needs to be culled into a manageable panel of markers for clinical application in prognosis and prediction, made widely available, and covered by health insurance. However, breast cancer has proven to us that these seemingly overwhelming tasks are very possible. RCC is ripe for personalized cancer treatment, which takes into account the underlying biology of an individual's tumor. The state-of-the-art has clearly led this field to the enviable position of having a range of effective molecularly targeted therapies, with further improvements expected on the horizon; mature profiles of protein and nucleic acid biomarkers, which will help us to define the spectrum of tumors that lie under the umbrella of ccRCC; and a

future unmapped territory of genetic mutations to explore that may provide more tools and answers to the questions we ask.

How this body of work builds on previous findings

Some incredible work has been and is continuing to be done in the field of characterizing ccRCC and providing means to predict clinical outcome. As you will see in the coming pages, our work builds on this tremendous foundation.

The literature described above strongly suggested that there must be molecular classes of clear cell renal cell carcinoma that are robustly separable using molecular profiles. All of the previous microarray data and the heterogeneity of the clinical presentation points to it. So, in chapter 2, we describe how we defined two molecular subtypes of ccRCC, which we named ccA and ccB. We also found that these subtypes have a vastly different survival outcome, with the ccB tumors having only 2 years compared to 8.6 years for ccA tumors, making this subclassification system also possible to implement as a prognostic biomarker.

In chapter 3, we examine and validate the underlying molecular pathways that distinguish these two subtypes. ccA tumors have an angiogenic molecular phenotype, while ccB tumors have a more proliferative and aggressive phenotype. These results hint at the prospect that ccA tumors might be more likely to respond to anti-angiogenesis agents.

In chapter 4, we explore the underlying genetic changes of the subtypes. While they are predominantly similar, ccB tumors have additional chromosomal deletions in regions that previous studies show to correlate with decreased survival. These regions may provide important clues to mechanisms of more aggressive disease and what may be driving the differences between ccA and ccB tumors. Additionally, we discovered that

the ccA subtype has mutations in a number of different histone modification genes, suggesting that epigenetic modifications play an important role in ccRCC development, progression, and/or stratification.

Chapter 5 describes the development of an assay to distinguish between ccA and ccB tumors, particularly for use with formaldehyde-fixed, paraffin-embedded (FFPE) tissue. Now that we know that there are these two vastly different subtypes of ccRCC, we need to put it into use. The next challenge will be to validate the prognostic significance of the subtypes and determine whether the underlying biological changes are predictive for response to current treatments. There is great hope for the future of RCC treatment, and it will be exciting to see what new advances this research will spur for the decade to come.

Chapter Two:

Molecular Stratification of Clear Cell Renal Cell Carcinoma by Consensus Clustering Reveals Distinct Subtypes and Survival Patterns

This work is modified from Brannon et al., Genes and Cancer, 2010[97].

Abstract

Clear cell renal cell carcinoma (ccRCC) is the predominant RCC subtype, but even within this classification, the natural history is heterogeneous and difficult to predict. A sophisticated understanding of the molecular features most discriminatory for the underlying tumor heterogeneity should be predicated on identifiable and biologically meaningful patterns of gene expression. Gene expression microarray data were analyzed using software that implements iterative unsupervised consensus clustering algorithms, to identify the optimal molecular subclasses, without clinical or other classifying information. ConsensusCluster analysis identified two distinct subtypes of ccRCC within the training set, designated clear cell type A (ccA) and B (ccB). Based on the core tumors, or most well-defined arrays, in each subtype, Logical Analysis of Data (LAD) defined a small, highly predictive gene set that could then be used to classify additional tumors individually. The subclasses were corroborated in a validation dataset of 177 tumors and analyzed for clinical outcome. Based on individual tumor assignment, tumors designated ccA have markedly improved disease-specific survival compared to ccB (median survival of 8.6 vs. 2.0 years, p=0.002). Analyzed by both univariate and multivariate analysis, the classification schema independently associated with survival. Using patterns of gene expression based on a defined gene set, ccRCC was classified into two robust subclasses based on inherent molecular features that ultimately correspond to marked differences in clinical outcome. This classification schema thus provides a molecular stratification applicable to individual tumors that has implications to influence treatment decisions, define biological mechanisms involved in ccRCC tumor progression, and direct future drug discovery.

Introduction

Clear cell renal cell carcinoma, ccRCC, afflicts upwards of 50,000 patients annually[131]. Most of these patients will present initially with localized disease, managed with surgery, but, unfortunately, nearly a third will develop recurrence and succumb to their disease. ccRCC incidence has increased uniformly over the last 30 years, associated with stage migration toward lower stages, likely due to the increased detection of lesions incidentally. However, there has not been commensurate improvement in survival. ccRCC tumors have variable natural histories, and genetic strategies have been largely unhelpful in identifying patients with higher or lower risk for recurrence due to the overwhelming association of this cancer with von Hippel-Lindau (*VHL*) tumor suppressor gene inactivation[132,133].

The Fuhrman classification system stratifies ccRCC by tumor cell morphology: low grade (grade 1), intermediate grades (grades 2 and 3), and high grade (grade 4) tumors, with corresponding association with RCC-related death[10]. Prognostic scoring systems such as the UCLA Integrated Staging System (UISS) have been developed using these morphologic characteristics, tumor size, and patient performance status as well as the inherent characteristics of stage and nodal status[8,134]. Other algorithms incorporate post-operative clinical information, but have limited discriminative ability for the abundant intermediate grade and intermediate stage tumors, and they fail to account for molecular distinctions in tumors[11]. The molecular basis of this diversity in clinical behavior is unclear and makes ccRCC a ripe target for investigating the nature of these heterogeneities.

Gene expression analyses have provided meaningful insight into the clinical heterogeneity of many solid tumors. Unsupervised clustering of gene expression data with supervised learning methods can provide powerful strategies to identify molecularly

and clinically significant cancer subtypes[68-71]. New unsupervised consensus ensemble clustering strategies have been developed that have successfully identified breast cancer subtypes correlated with significant differences in risk for recurrence[135-138].

In ccRCC, using traditional unsupervised gene expression analysis, we and others have demonstrated that two or more molecular sub-classifications of this tumor type exist[6,92,95,113,139]. Many prior investigations, however, rely on pre-selected molecular features or clinical outcomes as the criteria to identify expression signatures and distinguish gene sets. This type of approach fails to permit the underlying tumor biology, through the molecular endproducts of genetic changes, to inform the formation of tumor subgroups. A robust molecular classification system that connects tumor biology with individual tumor behavior should identify –*a priori*– the inherent patterns of gene expression that classify samples into non-overlapping sets with a high degree of accuracy.

To investigate the molecular features which best define subsets of renal cell carcinoma, we applied unsupervised consensus clustering to the gene expression data of ccRCC tumors, without applying biologic or clinical information. Two robust subtypes (we have designated ccA and ccB) with differentiating biological signatures could be distinguished using a small gene set defined by logical analysis of data (LAD). This gene set allows for assignment of individual tumors within the ccA/ccB classification scheme and is easily translatable to RT-PCR technology. Validation in an independent dataset demonstrated that ccA tumors have a markedly better prognosis than ccB, and that the molecular subtype was significantly associated with survival in both univariate and multivariate analysis. The identification of two robust ccRCC subclasses, which can be assigned by a small but highly significant panel of gene features, will provide a biological resource for future ccRCC investigation, allow better prognostication of ccRCC, and supply a wealth of information for therapeutic decisions.

Results

Identification of ccRCC subtypes. Gene expression data were obtained for 48 ccRCC samples and three independent replicate sample preparations. A flow-diagram depicting the analyses performed is presented in Figure 2.1.

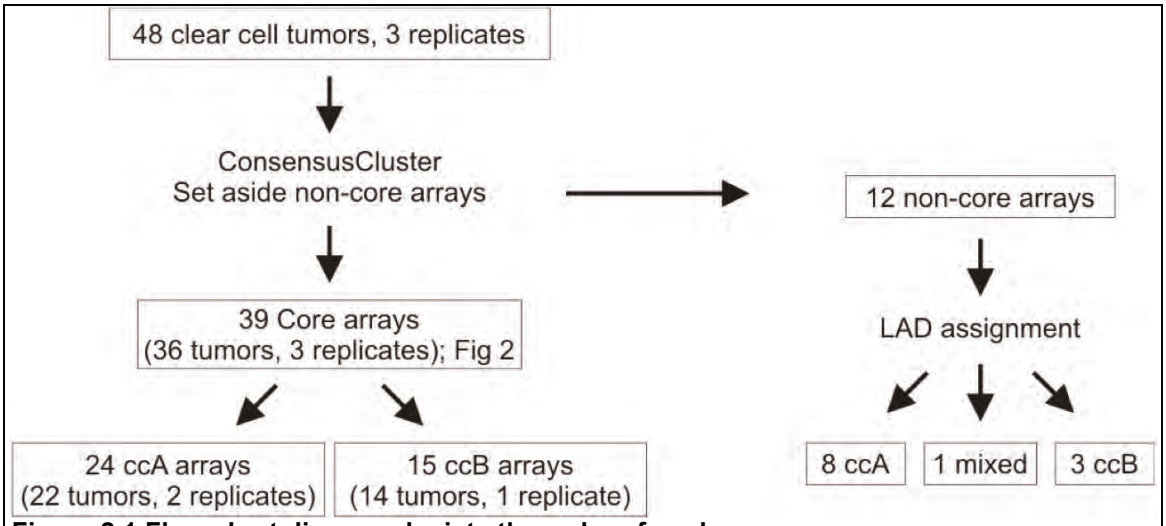

Figure 2.1 Flow chart diagram depicts the order of analyses
Delineation of steps taken to identify ccRCC subtypes.

First, we performed ConsensusCluster, an unsupervised ensemble clustering algorithm, on the ccRCC samples (Table 2.4), yielding two subsets, designated ccA and ccB (Figure 2.2A). Removing the independent replicates produced an identical clustering assignment of tumors (data not shown), further confirming the stability of these clusters. Neither cluster was caused by inclusion of normal tissue in the RNA extraction as normal kidney assorts independently of either cluster (Figure 2.3).

Representative samples within each cluster were used for the development of characteristic gene signatures and the decipherment of biological pathways. Samples whose membership shifted through multiple bootstrapped iterations were set aside for later classification. These “core” clusters included 39 of the original 51 samples, and

permitted tumors with best patterned features to define the cluster. As Figure 2.2B shows, the core cluster samples split into two robust subtypes of ccRCC that are stable when *k* (degrees of freedom) increases to *k*=3 or *k*=4 (Figure 2.2C-D), suggesting that the optimal number of robust clusters in this dataset is two. These analyses demonstrate that ccRCC can be optimally clustered into two distinct subtypes (ccA and ccB), defined purely by molecular characteristics of the tumors.

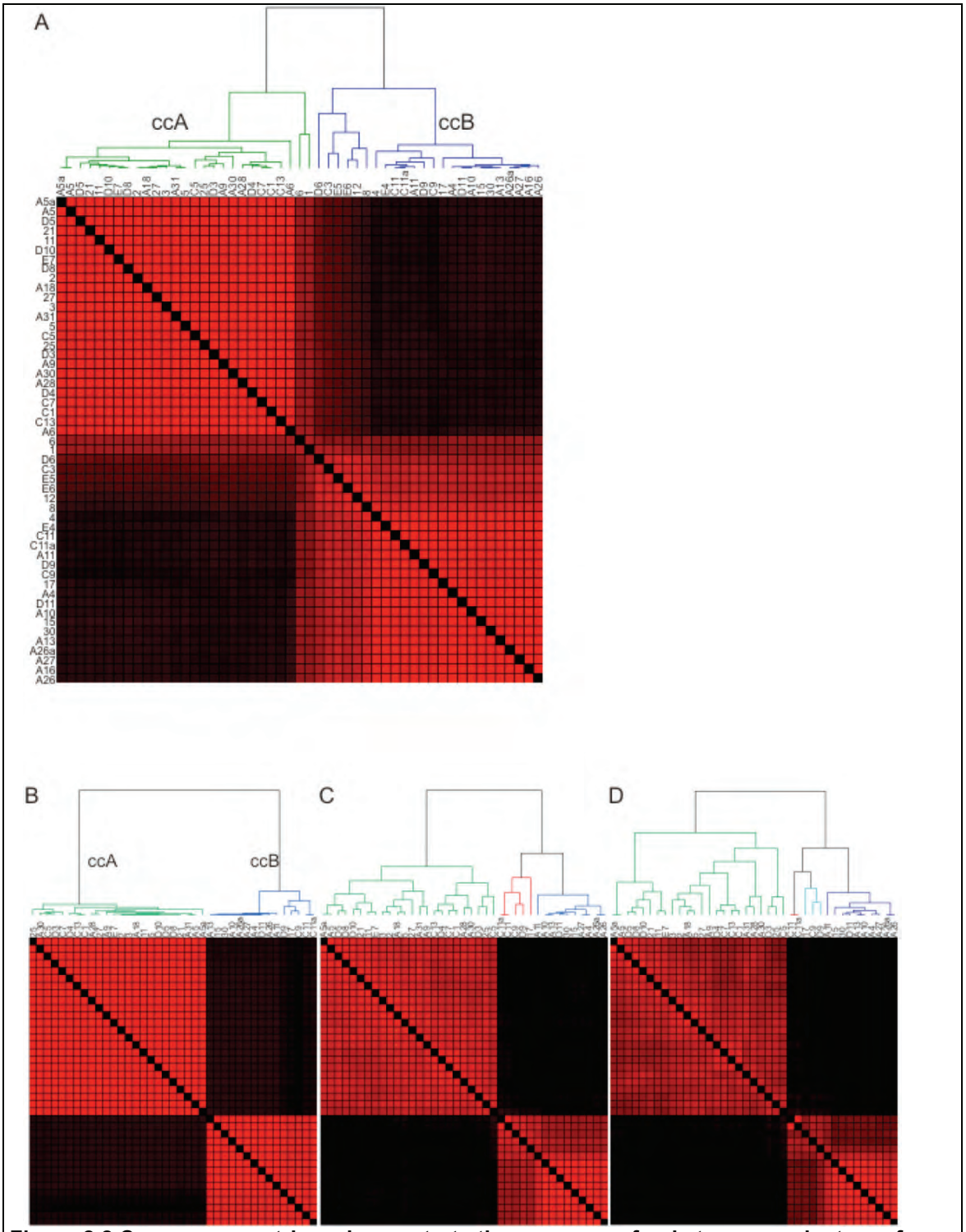

Figure 2.2 Consensus matrixes demonstrate the presence of only two core clusters of intermediate grade ccRCC.

Consensus matrix heatmaps demonstrate the presence of two clusters within all clear cell tumors (A) and invariance of the two ccRCC core clusters using (B) k=2, (C) k=3, and (D) k=4 cluster assignments for each cluster method. Red areas identify the similarity between samples and display samples clustered together across the bootstrap analysis. ccA is color coded in green, ccB in blue.

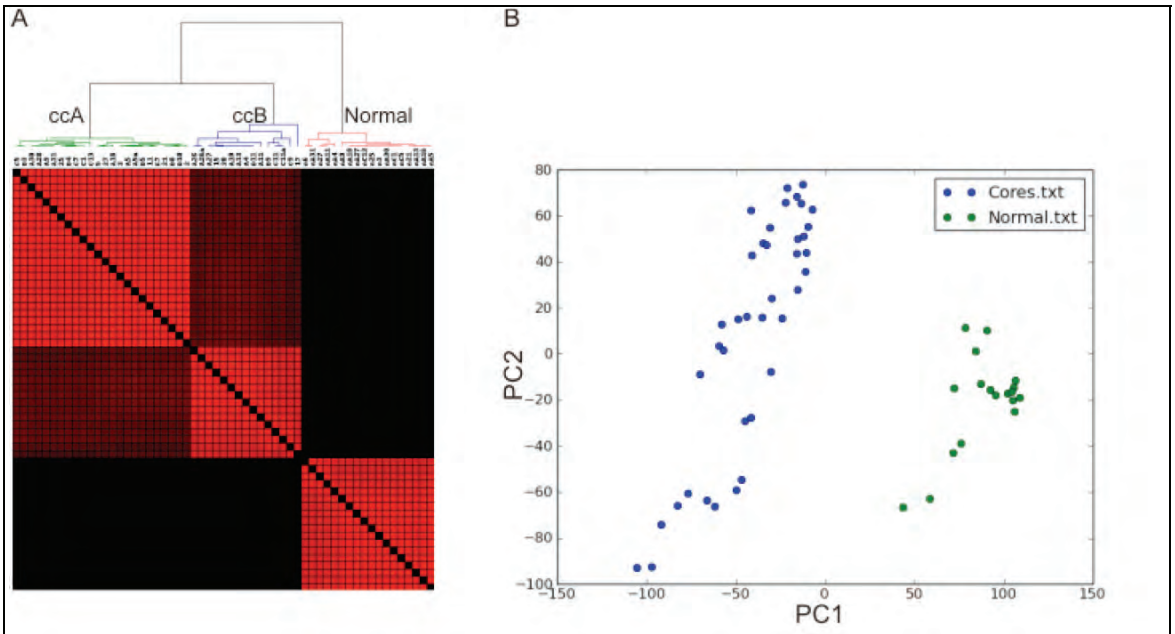

Figure 2.3 Two ccRCC subtypes are distinct from normal kidney tissue.
(A) Both consensus matrix and (B) PCA plot (scatter plot of the top 2 eigenvectors – PC1, PC2) show the complete delineation between the clear cell tumors and corresponding normal kidney tissue removed from ccRCC patients. Red areas identify samples clustered together across the bootstrap analysis. These results verify that the subtypes do not arise from errors in the expression levels due to contamination from normal tissue.

Delineation of a gene set to stratify ccRCC into ccA and ccB. To identify a feature panel that could accurately identify ccA and ccB tumors, we used logical analysis of data (LAD), which uses pattern recognition and supervised learning to identify key discriminating elements and has been successfully implemented in several biomedical studies[136,137,140]. Using the core ccA and ccB tumors, LAD patterns were identified and validated. Using these patterns, we identified 120 probes, consisting of 110 genes, valuable for cluster assignment (Figure 2.4A, Table 2.1). The LAD model was applied to the 12 non-core samples from the original analysis, and predicted cluster membership for 11 samples, 8 ccA and 3 ccB (Table 2.4).

Table 2.1 LAD Probe Set.
Probes identified through logical analysis of data (LAD) to discriminate between ccA and ccB subtypes. All probes were significant at t-test $p < 0.000001$. Fold change was calculated as ccA/ccB.

Type	Agilent Probe ID	Symbol	Fold diff.
ccA	A_23_P89799	ACAA2	4.159
ccA	A_24_P234242	ACADL	2.712
ccA	A_23_P24515	ACAT1	2.795
ccA	A_23_P52127	ACBD6	1.516
ccA	A_23_P134953	ADFP	3.951
ccA	A_23_P135454	AFG3L2	2.247
ccA	A_23_P129896	ALDH3A2	3.327
ccA	A_23_P417974	AQP11	2.899
ccA	A_23_P256084	ARSE	3.24
ccA	A_23_P86900	B3GNT6	2.41
ccA	A_23_P133923	BAT4	1.706
ccA	A_23_P134925	BNIP3L	2.503
ccA	A_23_P150350	C11orf1	2.47
ccA	A_23_P368718	C13orf1	2.483
ccA	A_24_P116233	C13orf1	2.081
ccA	A_23_P60259	C9orf87	4.427
ccA	A_23_P161719	CWF19L2	1.598
ccA	A_23_P147397	DNCH2	2.023
ccA	A_24_P112984	DREV1	2.161
ccA	A_23_P143484	DSCR5	2.553
ccA	A_24_P343621	ECHDC3	3.653
ccA	A_23_P119753	EHBP1	2.003
ccA	A_23_P87964	ESD	1.661
ccA	A_23_P118300	FAHD1	2.671
ccA	A_32_P93852	FAM44B	2.147
ccA	A_32_P213861	FBI4	2.75
ccA	A_32_P116271	FBI4	2.02
ccA	A_23_P41437	FLJ11200	2.149
ccA	A_23_P904	FLJ11588	2.2
ccA	A_23_P5742	FLJ13646	1.997
ccA	A_23_P58676	FLJ14054	9.81
ccA	A_23_P160433	FLJ14146	3.067
ccA	A_23_P165548	FLJ14249	2.159
ccA	A_24_P139943	FLJ14249	1.89
ccA	A_23_P203751	FLJ22104	3.108
ccA	A_24_P181101	FLJ22104	2.885
ccA	A_32_P197942	FLJ23834	2.499
ccA	A_24_P576191	FLT1	3.07
ccA	A_24_P38276	FZD1	3.116
ccA	A_24_P942370	GALNT4	1.804
ccA	A_24_P72064	GHR	3.943
ccA	A_23_P34478	GIPC2	5.447
ccA	A_24_P100301	GIPC2	4.163
ccA	A_23_P147296	HIRIP5	2
ccA	A_23_P253982	HOXA4	3.165
ccA	A_24_P218805	HOXC10	2.467
ccA	A_23_P363936	HSPA4L	2.339
ccA	A_23_P210176	ITGA6	2.15
ccA	A_23_P24948	KCNE3	2.633
ccA	A_24_P944541	KIAA0436	2.394
ccA	A_23_P29185	KIAA1043	1.876
ccA	A_32_P100683	KIAA1648	1.897
ccA	A_23_P215931	LEPROTL1	2.579
ccA	A_24_P252846	LOC119710	2.167
ccA	A_23_P144668	LOC134147	3.346
ccA	A_23_P206899	LOC57146	2.685
ccA	A_23_P337464	LOC90624	2.03
ccA	A_23_P85008	MAOB	3.677
ccA	A_32_P190416	MAP7	3.598
ccA	A_24_P224488	MAPT	4.959
ccA	A_23_P207699	MAPT	3.428
ccA	A_23_P341392	MGC32124	1.938
ccA	A_23_P83976	MGC33887	2.095
ccA	A_23_P115955	MRPL21	1.605
ccA	A_32_P77989	NETO2	4.082
ccA	A_23_P138686	NMT2	2.369
ccA	A_23_P253536	NPR3	7.48
ccA	A_23_P327451	NPR3	7.362
ccA	A_23_P414978	NUDT14	2.408
ccA	A_23_P10442	OSBPL1A	2.354
ccA	A_24_P124349	PDGFD	3.585
ccA	A_23_P115919	PHYH	2.62
ccA	A_23_P211598	PMM1	1.897
ccA	A_23_P52109	PRKAA2	2.832
ccA	A_24_P201404	PTD012	3.632
ccA	A_24_P97785	PURA	2.179
ccA	A_24_P93624	RAB3IP	3.301
ccA	A_23_P96420	RBMX	1.558
ccA	A_23_P203023	RDX	1.988
ccA	A_23_P428738	RNASE4	3.083
ccA	A_23_P144807	SETP8	2.232
ccA	A_23_P216468	SLC1A1	4.695
ccA	A_23_P56810	SLC4A1AP	1.339
ccA	A_32_P358887	SLC4A4	3.022
ccA	A_32_P167791	ST13	1.644
ccA	A_32_P85676	STK32B	3.508

ccA	A_23_P34375	TCEA3	2.726
ccA	A_23_P34376	TCEA3	2.904
ccA	A_24_P327886	TCEA3	2.967
ccA	A_23_P40611	TCN2	2.657
ccA	A_23_P58538	TIGA1	3.288
ccA	A_23_P29922	TLR3	4.409
ccA	A_23_P373819	TUSC1	2.817
ccA	A_32_P133884	TUSC1	2.883
ccA	A_24_P167052	YME1L1	1.46
ccA	A_23_P48705	ZADH1	3.082
ccB	A_24_P73577	ALDH1A2	0.333
ccB	A_23_P160729	AP4B1	0.624
ccB	A_23_P101380	B3GALT7	0.456
ccB	A_23_P50477	BCL2L12	0.609
ccB	A_23_P19182	C5orf19	0.262
ccB	A_23_P49155	CDH3	0.201
ccB	A_23_P2181	CYB5R2	0.408

ccB	A_23_P380266	FLJ23867	0.447
ccB	A_23_P19102	GALNT10	0.356
ccB	A_32_P170206	IMP-2	0.245
ccB	A_24_P262543	KCNK6	0.551
ccB	A_23_P67529	KCNN4	0.35
ccB	A_23_P102622	MATN4	0.317
ccB	A_23_P8649	MGC40405	0.499
ccB	A_32_P104825	NCE2	0.618
ccB	A_23_P52298	NPM3	0.517
ccB	A_23_P87238	SAA4	0.293
ccB	A_23_P91230	SLPI	0.19
ccB	A_23_P46390	SYTL1	0.348
ccB	A_24_P82880	TPM4	0.469
ccB	A_24_P37540	TTLL3	0.415
ccB	A_23_P92860	UNG2	0.283
ccB	A_24_P291598	USP4	0.507
ccB	A_24_P937119	ZNF292	0.303

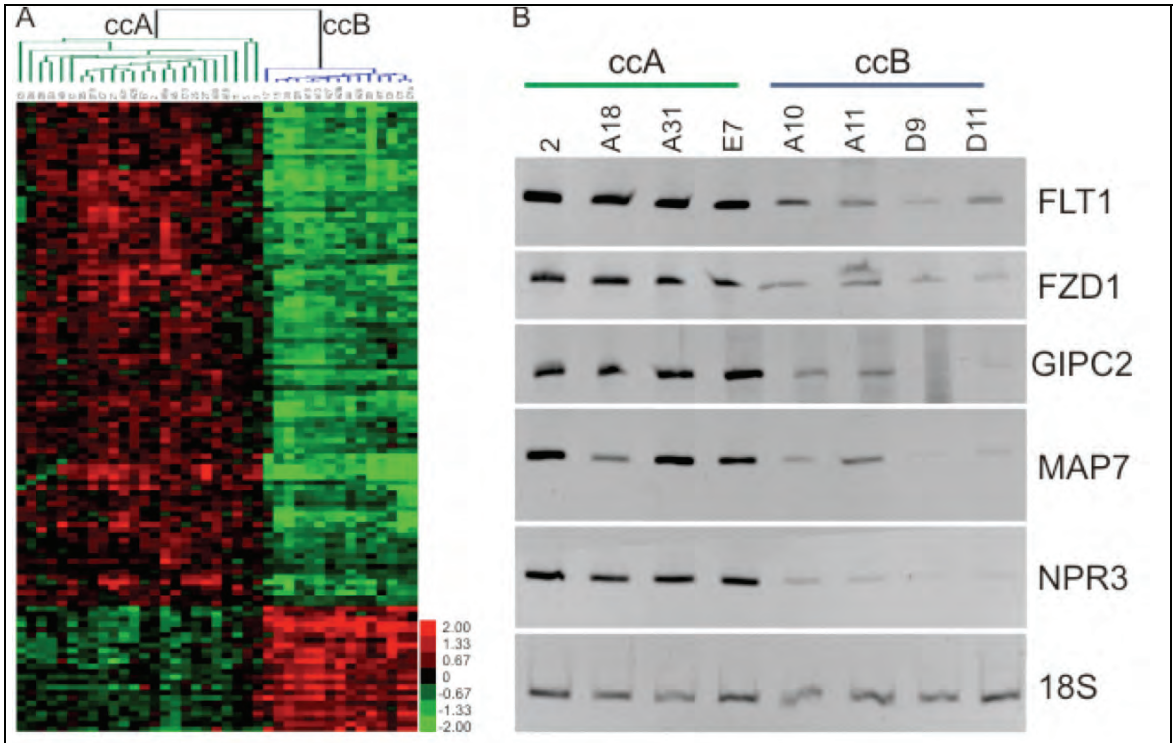

Figure 2.4 LAD probes separate ccA and ccB tumor clusters.
(A) Gene expression data for core arrays and 120 logical analysis of data (LAD) probes. These probes were selected using LAD and leave-one-out analysis from 1075 distinguishing probes with p-value < 0.000001. (B) Semi-quantitative reverse transcription PCR validates the ability of a subset of the LAD probes to clearly distinguish between ccA and ccB tumors.

To confirm that the genes identified by LAD are differentially expressed between ccA and ccB ccRCC subtypes within individual tumors, we tested primers for ccA overexpressed genes FLT1, FZD1, GIPC2, MAP7, and NPR3 on available tumor samples using semi-quantitative RT-PCR. Figure 2.4B demonstrates that each of these products can predict tumor classification for individual tumors. These results collectively indicate the potential for a limited gene set to correctly distinguish between the two ccRCC subtypes using RT-PCR, a platform immediately transferable to formalin-fixed, paraffin embedded tissues.

Validation of ccRCC subtypes. To validate the presence of two ccRCC subtypes in a second, independent dataset, we applied ConsensusCluster and the LAD probe set to 177 ccRCC microarrays generated using a different gene expression profiling technique[92]. Figure 2.5 shows the same two strong clusters in the data, which remained stable when k was increased (data not shown). The clusters were assigned to ccA or ccB by comparison of gene expression patterns to those in the primary dataset.

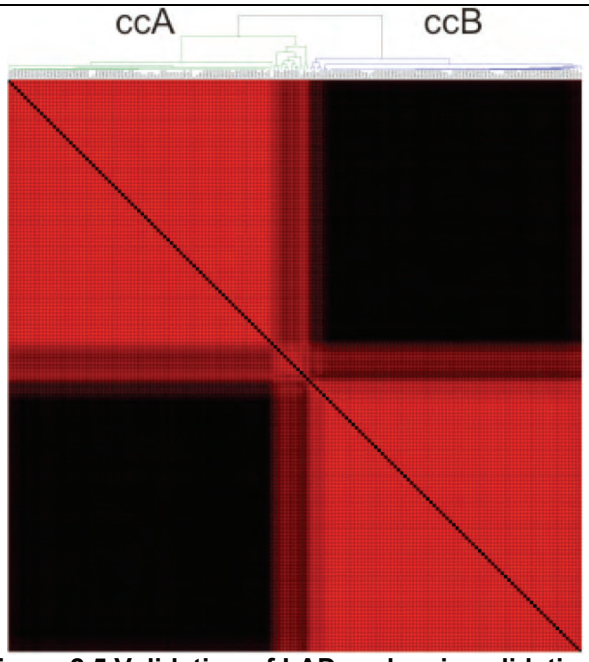

Figure 2.5 Validation of LAD probes in validation dataset show the existence of two ccRCC clusters. Consensus matrix of 177 ccRCC tumors determined by 111 probes corresponding to the 120 LAD probes. Red areas identify samples clustered together across the bootstrap analysis. Two distinct clusters are visible, validating the ability of the LAD probe set to classify ccRCC tumors into ccA or ccB subtypes from other array platforms.

Assignment of individual tumors. Assignment of tumors to a subtype with Cluster3.0 (traditional heatmaps) or ConsensusCluster requires the presence of other tumors. Therefore, we used LAD score to separately assign each individual tumor in the validation dataset to ccA or ccB, without assessing similarity to the rest of the tumors. Assignment was predicted for each sample 100 times with 80% pattern bootstrapping. A tumor was classified only if the assignment occurred in >75% of the prediction runs. Out of the 177 ccRCC tumors, 83 tumors were predicted to be ccA, 60 as ccB, and 34 remained unclassified with these stringent classification rules (online supplementary data). When compared with the cluster assignment predicted by ConsensusCluster, we found a concordance of over 86%, thus validating LAD predicted assignment as a sensitive measure of tumor assignment.

ccA and ccB have different survival outcomes. We then wanted to know whether the underlying differences in tumor biology would show survival differences. Cancer specific survival and overall survival for the ccA and ccB classes from the 177 tumor validation set were plotted using Kaplan-Meier curves (Figure 2.6A-B), calculating 95% confidence intervals (Table 2.2). For cancer specific survival (Figure 2.6A), the ccA subtype was associated with a highly significant survival advantage over ccB patients ($p=0.0002$, median survival of 8.6 vs. 2 years). At five years, cancer specific survival was 56% in ccA patients and only 29% in ccB patients. Figure 2.6B shows the same trend for overall survival, with a significantly greater survival for ccA patients over ccB patients ($p=0.004$, median survival of 4.9 vs. 1.8 years). At five years, survival for ccA patients is 48%, while only 23% for ccB patients.

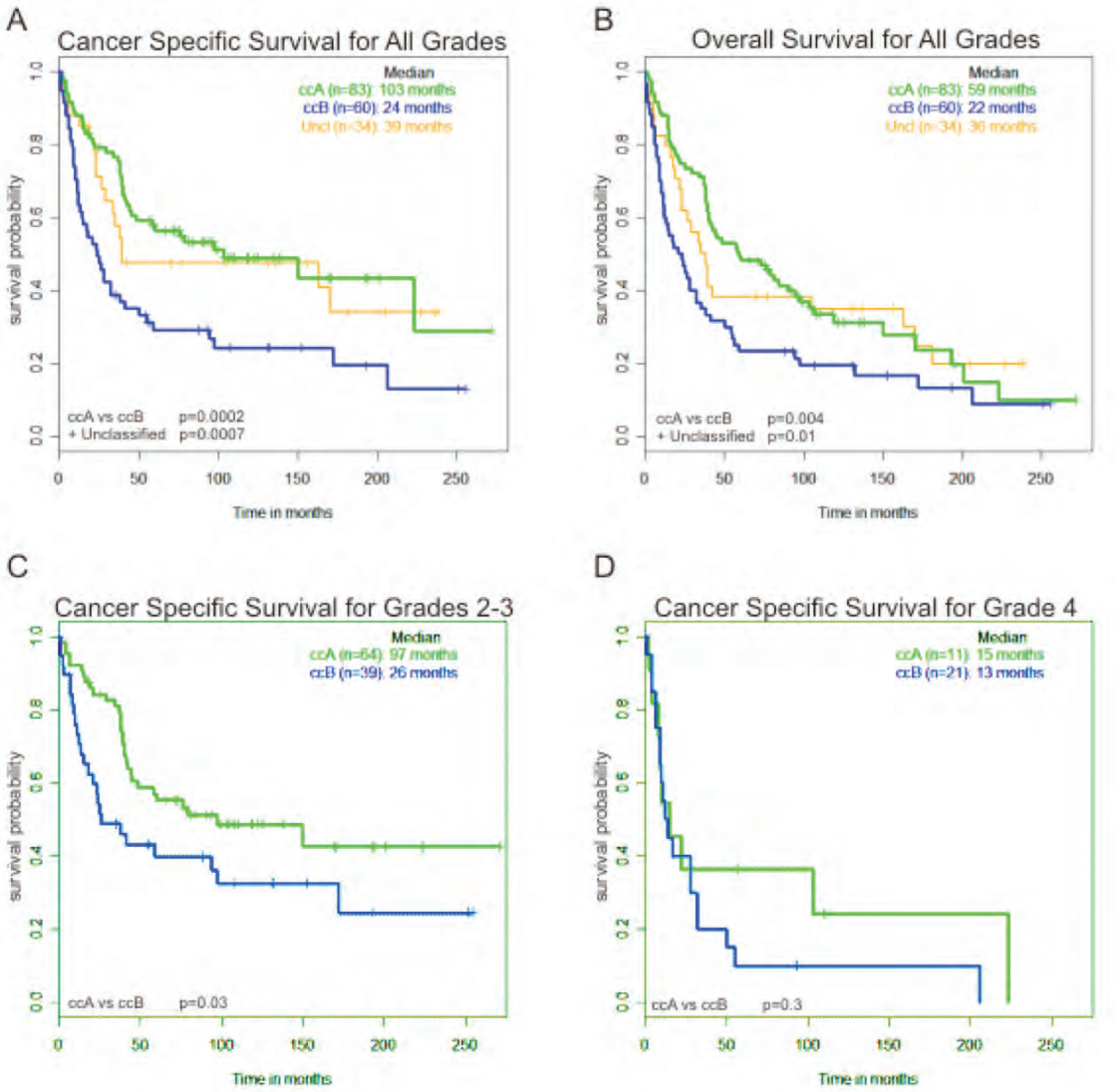

Figure 2.6 Classification of tumors from validation dataset by LAD prediction shows that subtypes have differing survival outcome.
177 ccRCC tumors were individually assigned to ccA (green), ccB (blue), or unclassified (orange) by LAD prediction analysis, and cancer specific (A) or overall survival (B) were calculated via Kaplan-Meier curves. The ccB subtype had a significantly decreased survival outcome compared to ccA, while unclassified tumors had an intermediate survival time (log rank $p<0.01$). (C) Cancer specific survival for intermediate (Fuhrman grade 2-3) tumors shows significant difference between subtypes. (D) Cancer specific survival for high grade (Fuhrman grade 4) shows a trend of better survival for ccA tumors.

Table 2.2 Survival Times with 95% Confidence Intervals.
Calculated median and 5 year survival times with 95% confidence intervals (CI) for ccA and ccB subtypes in disease specific (DSS) and overall survival (OS) analysis.

Survival analysis	Subtype	Median survival (years)	95% CI for median survival (years)	5 Year Survival (%)	95% CI for 5 year survival (%)
DSS	ccA	8.6	3.8 – N/A	56	45 – 67
	ccB	2.0	1.0 – 3.2	29	18 – 41
OS	ccA	4.9	3.3 – 7.8	48	37 – 58
	ccB	1.8	0.9 – 2.6	23	14 – 35

ccA/ccB subtype associates with clinical variables. Fuhrman grade, tumor size (T stage), and performance status, the covariates in the UCLA International Staging System (UISS) for predicting outcome in newly diagnosed patients[8], were evaluated and compared with our molecular classification with regard to survival outcomes. As expected, molecular classification strongly associated with tumor stage (p=0.009) and grade (p=0.0007), but not performance status (p=0.5684). 78% of grade 1 and 69% of stage 1 tumors clustered as ccA, while and 65% of grade 4 and 58% of stage 4 tumors cluster as ccB tumors. As low grade ccRCC tumors tend to have better prognosis, and high grade tumors toward poor prognosis[10], this result was expected. This observation also suggests that the biological characteristics responsible for grade and stage-specific prognosis in ccRCC are encompassed in the classification schema. Figure 2.6C demonstrates that the ccA/ccB subtype still significantly correlates with survival when limiting analysis to intermediate grade (grade 2-3) tumors. As expected, a Kaplan-Meier curve limited to the highly aggressive grade 4 tumors shows a convergence of subtype-specific survival (Figure 2.6D).

Molecular classification is independently associated with survival. To determine how our classification schema compares with current standard clinical parameters as a prognostic factor, univariate Cox regression analyses were performed (Table 2.3). Molecular subtype is strongly associated with survival, with an HR of 2.2 (p=0.0003). Even in the absence of stage 4 (metastatic) tumors, subtype has a strong association with survival (HR=2.143, p=0.0233). Additionally, the use of Schwartz Bayesian Criterion (SBC) suggests[141] that whether the tumor is classified by ccA/ccB/unclassified, ccA/ccB, or LAD score, the measures are strongly associated with survival, with difference in adjusted SBC values of 8, 8.3, and 9 respectively. These

results suggest that defining a tumor as ccA or ccB may be an important prognostic indicator for predicting outcome from patients with ccRCC.

Multivariate analyses were then performed to determine whether our classification schema was still independently associated with survival outcomes in the context of stage, grade, and performance status. The dichotomous classification of ccA/ccB provides a significant association with survival at the 0.1 level (p=0.089), likely influenced by the smaller sample size of the 143 classified tumors. Increasing sample size to 177 by including unclassified tumors, the trichotomous classification increased significance to p=0.0736. Statistical analyses often show that continuous variables provide more statistical discrimination. In fact, LAD score is an independent predictor of survival (p=0.0027) and is more predictive of outcome than Fuhrman grade (p=0.0308). These data intimate that the classification schema presented in this paper may provide independent prognostic information over and above that provided by standard clinical parameters.

Table 2.3 Univariable Cox regression analysis for Disease Specific Survival.
Hazard ratios, with 95% confidence intervals (CI) and p-values, were calculated for the predicted subtype (ccA vs ccB), LAD score, stage, grade and performance status (PS). Analysis of "Subtype ccA/ccB" used only the 143 tumors classified using bootstrap analysis. Analysis of "Subtype all ccA/ccB" included all 177 tumors classified by LAD score without using the 75% confidence cutoff. Analysis of "Subtype ccA/ccB/uncl" included all 177 tumors classified as ccA, ccB, or unclassified by LAD score and bootstrapping. The HR for LAD score is per 0.1 units.

Covariate of Interest	HR	95% CI	p-value
Subtype ccA/ccB	2.2	1.4 – 3.4	0.0003
Subtype all ccA/ccB	1.8	1.2 – 2.7	0.0033
Subtype ccA/ccB/uncl	1.5	1.2 – 1.9	0.0004
LAD score	1.2	1.1 – 1.3	0.0002
Grade	1.9	1.4 – 2.5	<0.0001
Stage	3.4	2.6 – 4.3	<0.0001
Performance Status	1.7	1.4 – 2.1	<0.0001

Discussion

Unsupervised consensus clustering algorithms can identify distinct classifications of histologically similar tumors based on machine learning algorithms. In this analysis, a small gene set distinguishes two inherent molecular subtypes of ccRCC (ccA and ccB), characterized by a highly significant association with survival outcomes. This unique analysis provides a powerful method to discriminate molecular subgroups of tumors that may be informative of tumor biology or influence tumor behavior.

A fundamental problem in gene expression analysis of human tumors is the measurement of genetic noise in pairwise comparisons across thousands of independent and dependent variables. Our combined use of PCA, consensus clustering, and LAD is robust, and, more importantly, identifies stable clusters within patterns of gene expression. This method is highly reproducible and able to classify samples into molecular and clinically meaningful categories. Within these categories, "Core clusters" are sets of non-overlapping samples that are distinguishable from each other with high accuracy. This method of tumor analysis permits a refined assignment into gene expression-defined classifications and yields predictive gene signatures based on a manageable sized number of gene features. These properties permit the identification of limited sets of highly predictive molecular features (ie, genes) useful for the classification of individual samples outside of the primary analysis. The extension of biomarker molecular profiles to small groups of genes, which can assign classification to individual tumors is a major step forward toward the development of a clinically relevant biomarker. Ultimately, such a classification scheme will be applied with such measures as quantitative RT-PCR.

The clinical heterogeneity of ccRCC, coupled with previous gene expression studies[39,95,113,139] suggest that at least two molecular subtypes of ccRCC exist. We

demonstrated that there are likely *only* two primary subtypes of ccRCC stable under bootstrap analysis, although further subclassifications within these subtypes may be identified in much larger datasets, and rare tumors may represent unusual variants**.** Using the LAD predictions in the validation set, a third group of tumors shared pattern features with both ccA and ccB tumors. Such a third group, or other suggested classifications, may represent an intermediate manifestation of tumors undergoing progression from ccA to the ccB subtype, or which simply share common characteristics of both groups.

The subtypes ccA and ccB were associated with a significant difference in survival outcome, with ccA patients having a markedly better prognosis. While the continuous variable of LAD score proved to be an independent predictor of survival, the more immediately clinically useful dichotomous classification of ccA or ccB had a similar effect size and was statistical significant at the p=0.1 level in the multivariable analysis. Future studies on larger numbers of patients are needed to validate the results of the preliminary multivariate analysis reported herein.

Finally, our small, robust panel of genes, whose expression levels can classify individual tumor samples into ccA and ccB subtypes with high accuracy, may provide a valuable resource for clinical decisions for patients following nephrectomy regarding frequency of surveillance or choices for adjuvant therapy in the future. This panel provides the basis for the development and validation by a prospective clinical trial to assign subtypes of ccRCC to individual tumor specimens for implementation in a prognostic algorithm.

Materials and Methods

Samples. 51 specimens from 48 ccRCC patients were collected from by the UNC Tissue Procurement Core Facility consenting patients undergoing nephrectomy for RCC from 1994 – 2008 (Table 2.4), analyzed for quality, flash frozen, and accessed with appropriate IRB approvals. The validation set of 177 cases was described previously[92]. Survival data were updated with median follow-up of 120 months (range 66 to 271).

Table 2.4 Tumor characteristics for 51 clear cell samples.
Tumors suffixed with "a" were independent replicates. Arrays labeled in parentheses were assigned by pattern analysis using the 120 LAD probes. If labeled (unclass), the tumor could not be assigned using LAD pattern analysis. Grade – Fuhrman nuclear grade (1-4). Size – Tumor size (cm). T-stage – Tumor stage according to pathology report. WT – no mutations detected. U – unmethylated. M – methylated. n/a – not available.

Tumor	Core	Grade	Size	T-Stage	VHL mutation	VHL methylation
2	ccA	2	5.2	T1b	n/a	U
3	ccA	2	2.5	T1a	mutated	U
5	ccA	2	6.1	T1b	n/a	U
11	ccA	2	4	T1a	mutated	U
21	ccA	2	4.4	T1b	n/a	U
25	ccA	2	4.7	T1b	mutated	M
27	ccA	2	4.5	T1b	n/a	U
A18	ccA	2	7.5	T2	WT	n/a
A28	ccA	2	8	T2	mutated	U
A30	ccA	2	5.5	T1b	WT	U
A31	ccA	2	2.7	T1a	mutated	U
A5	ccA	3	17	T3a	WT	U
A5a	ccA	3	17	T3a	WT	n/a
A9	ccA	2	8.2	T3b	mutated	U
C1	ccA	3	2.2	T1a	n/a	n/a
C13	ccA	3	4.7	T1b	n/a	n/a
C5	ccA	2	2.7	T1a	n/a	n/a
C7	ccA	3	2.8	T1a	n/a	n/a
D10	ccA	2	3.5	T1a	n/a	n/a
D3	ccA	2	5	T1b	n/a	n/a
D4	ccA	1	5.5	T1b	n/a	n/a
D5	ccA	2	4.1	T1b	n/a	n/a
D8	ccA	2	3.8	T1a	n/a	n/a
E7	ccA	2	5.5	T1b	n/a	n/a

15	ccB	2	5.5	T1b	mutated	U
17	ccB	2	3	T1a	WT	U
30	ccB	3	7	T1b	WT	U
A10	ccB	2	3.2	T1a	WT	U
A11	ccB	3	3	T1a	WT	U
A13	ccB	3	10	T3b	WT	U
A26	ccB	2	3	T1a	WT	M
A26a	ccB	2	3	T1a	n/a	n/a
A27	ccB	2	2	T1a	WT	n/a
A4	ccB	2	3.9	T1a	n/a	U
C11	ccB	2	7.5	T2	n/a	n/a
C11a	ccB	2	7.5	T2	n/a	n/a
C9	ccB	3	8.7	T2	n/a	n/a
D11	ccB	2	2.3	T1a	n/a	n/a
D9	ccB	2	1.8	T1a	n/a	n/a
1	(ccA)	2	7.9	T2	WT	U
6	(ccA)	2	4.3	T1b	mutated	U
12	(ccA)	3	8	T2	mutated	U
A6	(ccA)	2	3.8	T1a	WT	M
C3	(ccA)	2	4.5	T1b	n/a	n/a
D6	(ccA)	3	4.2	T1b	n/a	n/a
E5	(ccA)	2	8	T2	n/a	n/a
E6	(ccA)	3	10.2	T2	n/a	n/a
4	(ccB)	3	5	T3b	n/a	U
A16	(ccB)	1	2.5	T1a	WT	n/a
E4	(ccB)	2	3.5	T1a	n/a	n/a
8	(unclass)	3	4.5	T3a	mutated	M

***Gene Expression Analysis*.** RNA was extracted from fresh frozen tumor specimens (with independent replicates – separate sample preparations – of 3 tumors) and 18 specimens from adjacent normal kidney using the Qiagen RNeasy kit (Valencia, CA). The concentration of the purified RNA was measured on a Nanodrop ND-1000 (Thermo Scientific, Wilmington, DE), and quality was assured using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). The UNC Genomics Core processed RNA samples for amplification, label integration, and hybridization against a modified commercial reference RNA[39] on Agilent Whole Human Genome (4x44k) Oligo

Microarrays (Santa Clara, CA). Microarrays were scanned using the Agilent Scanner model C. Fluorescence ratios were determined by Agilent feature extraction software.

***Data Normalization*:** Expression data from the Agilent Arrays were tabulated in log2 R/G Lowess normalized ratio (median) format, removing probes which had ≤70% good data (Exclude if spot is not found in either channel, spot or spot background is a non-uniform outlier, spot or spot background is a non-uniform outlier for the population, spot is not a positive and significant signal in either channel, or Ch1 and 2 Lowess normalized net (median) <10). Missing data was imputed using k-nearest neighbors method (k=10) using Significance Analysis of Microarrays (SAM, http://www-stat.stanford.edu/~tibs/SAM/). The data for three groups of arrays, which were prepared in separate sample batches, was combined using Distance Weighted Discrimination (DWD, https://cabig.nci.nih.gov/tools/DWD).

Group 1: A4, A5, A6, A9, A10, A11, A13, A16, A18, A26, A26a, A27

Group 2: 2, 5, D3, D4, D5, D6, D8, D9, D10, E5, D11, E4, E6, E7, n6, n21, nC5

Group 3: 1, 3, 4, 6, 8, 11, 12, 15, 17, 21, 25, 27, 30, A28, A30, A31, A5a, A7, C1, C11, C11a, C13, C3, C5, C7, C9, n25, n27, n3, nA11, nA13, nA16, nA18, nA27, nA30, nA31, nA4, nA5, nA9, nC1, nC13

DWD is a tool that performs statistical corrections to reduce systematic biases resulting from different sources of RNA, batches of microarrays etc. It is generally used when combing data from different microarray platforms, but is also valuable to correct for possible biases introduced due to batch handling effects in data generated on the same platform in the same lab. These data are posted on GEO (GSE16449).

The 177 tumor validation set consisted of gene expression data from ccRCC specimens from a previously published paper[92], which is also available on GEO

(GSE3538). It was tabulated and imputed as described above. This data consisted of 10 print runs, which were also combined by DWD as above. Arrays were then standard normalized by subtracting the mean of the array and dividing by the standard deviation.

Principal Component Analysis (PCA). ConsensusCluster[142] (http://code.google.com/p/consensus-cluster/) was used for PCA[143,144] and consensus clustering[135]. Features whose coefficients were in the top |25%| were selected from PCA eigenvectors representing 85% variation in the data, retaining 26 eigenvectors and 347 features.

PCA is a feature selection method which reduces the feature set to those which have significant variation within the sample set. It is essentially a coordinate transformation in feature space which identifies a sorted list of "Principal Components", which are linear combinations of the original features. The starting point of the analysis is the expression matrix E_{ij} where the rows are samples and columns are genes. The analysis proceeds by computing the eigenvalues and eigenvectors of the correlation matrix between feature pairs across samples after E_{ij} is centered and scaled to mean 0 and variance 1 per column. The higher the eigenvalue of the correlation matrix, the greater the variation represented by the direction in feature space defined by its eigenvector. The eigenvalues λi were sorted in decreasing order and the k largest eigenvalues representing a fraction ρ of the variation in the data were identified by solving $[\sum_{i=1}^{k} \lambda_i] = \rho [\sum_{i=1}^{N} \lambda_i]$ where N is the total number of genes. We selected ρ= 0.85; the results are not sensitive to this choice. From an examination of the coefficients of the genes in the eigenvectors for these eigenvalues, we identified the subset of useful genes as those with coefficients in the top 25% in absolute value in these k

eigenvectors. In the 48 tumors plus three replicates dataset, this identified 26 eigenvectors and 347 features which were retained for further analysis.

Unsupervised Consensus Ensemble Clustering. Consensus clustering was applied to PCA features to divide the data successively into *k*=2,3,4… clusters, with 80% bootstrapping of 300 subsamples of genes and/or samples. We applied two clustering techniques, K-Means[145] and Self-Organizing Map[146].

Unsupervised clustering algorithms divide data into groups such that the intra-cluster similarity is maximized and the inter-cluster similarity is minimized. For gene expression data, unsupervised clustering can be performed for genes, for arrays, or for both. Several types of clustering techniques are available to group data into sets. These may be divided into hierarchical, partitioning, probabilistic and grid-based methods. Consensus ensemble clustering [146] is a relatively recent method which uses a weighted combination of these methods to improve the quality and the robustness of the clusters identified by each individual technique. The consensus ensemble approach involves two methods: first, a method that generates a collection of clustering solutions, and second, a method that robustly combines the solutions to produce a single "best" clustering solution for the data. Unlike standard clustering techniques whose solutions divide *all* the data samples into groups, ensemble consensus clustering identifies "core" groups of samples within clusters. These are samples which are consistently clustered into the same group, independent of perturbations of the data and of the choice of clustering methods used. This allows one to identify strong signatures of gene expression within each core cluster which can then be used to classify the remaining samples. It also allows a robust (perturbation independent) characterization of the gene expressions which distinguish the disease classes identified. Often a study of these genes which

have noise independent differential expression between disease classes allows a better understanding of the underlying biological mechanisms driving the subtypes.

We use several techniques to create robust "core" clusters. If the clustering method is stochastic, we reduce the effect of stochastic variation by applying the clustering method repeatedly and taking an appropriate average. To reduce the sensitivity of the results to random variation in the data, we apply each clustering method to multiple sample datasets obtained by bootstrapping both the features (genes/probes) as well as the samples clustered. The core clusters are identified as those groups whose memberships consist of samples consistently classified into the same group over all the bootstrap and clustering experiments. We have developed our own (publicly available) software suite called ConsensusCluster which implements PCA and consensus ensemble clustering. The code is available at http://code.google.com/p/consensus-cluster/.

Consensus ensemble clustering was applied to data limited to the 347 features identified by PCA and the data was split into k=2, 3, 4… clusters, which were made insensitive to data and clustering method bias by bootstrapping over many datasets and averaging over two clustering techniques, K-Means[145] and Self-Organizing Map[146].

The detailed procedure used is described below:

Step 1. 75 datasets were created from the imputed data restricted to the 347 significant features identified by PCA. 75 datasets came from bootstrapping the samples, 75 from bootstrapping genes and 75 by first projecting the data on bootstrapped genes and then by further bootstrapping on samples.

Step 2. k=2,3,4 clusters were created for each dataset using k-means and SOM.

Step 3. For each k and each method, the k resulting clusters were combined into an agreement matrix A_{ij} of size n x n.

Step 4. For each k, the samples were clustered using $d_{ij} = 1\text{-}A_{ij}$ as a distance measure using hierarchical clustering and the hierarchical tree was truncated at the k^{th} level.

Logical Analysis of Data (LAD). Features mapped to genes that discriminate between the two subtypes (t-test p<0.000001) were retained. We then applied LAD[147,148], (http://pit.kamick.free.fr/lemaire/software-lad.html). LAD patterns requiring only one gene for perfect discrimination were generated. LAD was reapplied to identify patterns of degree 1 and degree 2 (homogeneity and prevalence = 0.9). A classifier $C_S=f_P\text{-}f_N$ assigned an unknown sample to a class, where f_N/f_P are the fraction of negative/positive patterns satisfied. If the LAD score (C_S) was negative/positive, the sample was predicted to class ccA/ccB respectively.

Logical analysis of data [147,148], is a method to find patterns distinguishing two classes. For gene expression data, LAD identifies patterns of expression which can stratify labeled data. It has been successfully used in several biomedical studies[136,137,140].In our case, a pattern is a rule based on cutpoints in the expression of genes which can distinguish our two subtypes ccA and ccB. A pattern is characterized by its degree, prevalence, and homogeneity. The *degree* is the number of genes appearing in its defining conditions. The *prevalence* of a pattern is the percent of positive (negative) cases which satisfy the pattern. The *homogeneity* of a pattern is the percentage of positive (negative) cases covered by it. In general, patterns useful for classification have low degree and high prevalence and homogeneity.

To develop patterns to distinguish ccA and ccB, we used the complete set of probes on the chip so as not to bias the analysis in any way. Each sample array was first standard normalized by subtracting the mean of the array and dividing by the standard deviation, in order to create patterns applicable to other datasets. We retained only

those features that could discriminate the subtypes using a *t* test at p-value < 0.000001 and only kept the probes which were mapped to known genes. This reduced the dataset to 1075 probes, which included the set of 347 identified by PCA. We then applied LAD [147,148], using the implementation that is available at (http://pit.kamick.free.fr/lemaire/software-lad.html). LAD patterns requiring only one gene for perfect discrimination were generated in Leave-One-Out experiments (LOO) (see below) to further reduce the gene set to 120. These probes were re-normalized by median centering, and LAD was reapplied to identify patterns of degree 1 and degree 2 (homogeneity and prevalence=0.9) using a single cut-point at expression value 0.

These patterns were used to predict the samples initially set aside as non-core samples. A classifier $C_S=f_P-f_N$ assigns an unknown sample *S* to a class, where f_N/f_P are the fraction of negative/positive patterns satisfied by *S*. If the LAD score (C_S) is negative/positive, the sample is predicted to class ccA/ccB respectively. Confidence levels were computed by running 100 bootstraps of 80% of the patterns from the entire set, and the LAD score was computed for each bootstrapped sample. The final LAD score was the average of 100 runs, and the confidence level was the percent of times the sample was predicted to be in ccA or ccB. Samples with confidence levels < 0.75 were left as unclassified.

Leave-One-Out Analysis (LOO). LOO is a procedure to test the accuracy of a classifier that distinguishes two labeled classes. One sample is left out, then the classifier is created from the remaining samples and is used to predict the class of the sample left out. The procedure is then repeated for all possible selections of "left-out" samples. The prediction accuracy of the classifier is the average fraction of correct classifications across all choices of the "left-out" sample.

Semi-quantitative Reverse Transcription PCR. Where available, RNA was extracted from a second tumor sample from the same patient. Tumors were chosen based on RNA or tumor availability of RNA or tumor with the end goal of equal numbers in each subtype. 500ng of total RNA from training set patient tumor samples was reverse transcribed using Superscript II polymerase (Invitrogen, Carlsbad, CA) using manufacturer recommended standard buffer and temperature conditions. A 1:5 cDNA dilution was amplified by 25 cycles of semi-quantitative PCR with primer sets for FLT1 (ACTTTTACCGAATGCCACC and TGGTTACTCTCAAGTCAATCTTG), FZD1 (CCATCAAGACCATCACCATC and GCCGATAAACAGGTACACGA), GIPC2 (CCTGAGATCAAAAGGTCCTG and CTTCAAACATTGTGGTGGC), MAP7 (GCTACAGATAAGAAAACCAGTGA and GCTTTCCATTTCCCGGA), and NPR3 (TCGGCAGTGACAGGAATT and CCCGATGTTTTCCAAGGT). Primers were designed using IDT (http://www.idtdna.com/). 18S rRNA primers (Applied Biosystems) were used as a control. Equivalent quantities of the semi-quantitative RT-PCR samples were run on a 6% acrylamide gel.

Statistical Methods. All statistical analyses were performed using R v2.4.1 (http://www.r-project.org), SAS (SAS Institute, Inc, Cary, NC), and STATA (Statacorp, College Station, TX). The Kaplan-Meier (or product limit) method was used to estimate the time to event functions of disease specific survival and overall survival. Disease specific survival was defined as the time from the nephrectomy to death due to disease. Overall survival was defined as the time from nephrectomy to death from all causes. The log-rank test was used to test for differences between disease-specific and overall survival Kaplan-Meier curves. Univariable logistic regression was used to evaluate the

relative strength of association of covariates, one at a time, on the outcome probability of being subtype ccA versus ccB. The covariates of interest here were performance status, tumor stage, and grade. Univariable and multivariable Cox regression was used to evaluate the strength of association of individual and multiple covariates on disease specific and overall survival. The covariates of interest in these models were performance status, tumor stage, Fuhrman grade, subtype (ccA/ccB, or ccA/ccB/unclassified), and LAD scores. Model fit was assessed using an approximation to Bayes factors known as the Schwartz Bayesian Criterion (SBC) [141].

Acknowledgements Thanks to Leslie Kennedy and D. Micah Childress for technical assistance; to Perou lab members Katie Hoadley, Aaron Thorner, and Joel Parker for analysis suggestions; and to Tricia Wright for critical reading.

Funding The work of GB was supported in part by the National Science Foundation Grant No. PHY05-51164 and the New Jersey Commission on Cancer Research Grant 09-112-CCR-E0. SG received support from the Sidney Kimmel Foundation and NJCCR. WKR received support from the Lineberger Comprehensive Cancer Center, the Doris Duke Charitable Fund, and the Crawford Fund for kidney cancer research. ARB was supported by the UNC Cancer Cell Biology Training Grant. The UNC Tissue Procurement Facility and Genomics Core are supported by the Lineberger Comprehensive Cancer Center.

Supplementary material for this article can be found at the Genes & Cancer Web site, http://ganc.sagepub.com/supplemental.

Chapter Three:

Molecular pathways of ccRCC subtypes identify an angiogenic/hypoxic vs. a proliferative/aggressive stratification

This work is modified from Brannon et al., Genes and Cancer, 2010[97].

Abstract

Clear cell renal cell carcinoma (ccRCC) is the main histological subtype of kidney cancer, but presents in the population as a clinically heterogeneous disease. Acceptance that ccRCC consists of two main subtypes has been increasing, but these subtypes require further characterization. Therefore, overall gene transcript and pathway differences were examined between the two subtypes and validated in another dataset. ccA tumors, the better prognostic group, overexpresses angiogenesis and hypoxia related genes. In comparison, ccB tumors, which portend a poorer prognosis, overexpress more aggressive sets of genes, including Myc targets, cell cycle and epithelial-to-mesenchymal transition (EMT). ccB tumors also underexpress metabolism related genes in comparison to normal tissue. Despite the angiogenesis and hypoxia signature of ccA tumors, VHL inactivation was identified in both subtypes. Additionally, while the pathway patterns show similarity to differences previously identified in HIF1 and HIF2 vs HIF2 only expressing tumors, HIF protein expression was also confirmed to be relatively equal in both subtypes. Overall, this chapter provides more insight into what causes the split between ccA and ccB tumors, both molecularly and with regards to prognosis.

Introduction

In the last chapter, we identified two molecular subtypes of clear cell renal cell carcinoma, ccA and ccB, with vastly different survival outcomes. While we also identified genes to distinguish between the two subtypes, few of these genes were suggestive of pathways that would create survival stratification. Therefore, we wanted to know what is driving this difference between the more indolent ccA tumors and the aggressive ccB tumors.

Because a nephrectomy can be curative for many patients, the decreased survival outcome in ccB tumors is suggestive of early recurrence by metastasis. Therefore, we hypothesized that ccB tumors overexpress molecular pathways that are indicative of or predictive for metastasis. However, the particular pathways remained to be elucidated.

Only two gene expression studies have been able to provide any insight into biological pathways involved with inherent molecular clusters of ccRCC. Skubitz et al.[95] identified 2 groups in their 16 ccRCC tumors: One group overexpressed angiogenesis genes, while the other overexpressed extracellular matrix and cell adhesion genes. This study, however, included 3 tumors with sarcomatoid features, which are indicative of a far more aggressive disease progression. Therefore, while this information provides grounds from which to work, the results may not be entirely indicative of the general clear cell population.

The other study is by the Brooks group[92] that we used to validate our subtypes and provide survival data in the previous chapter. Of their two main clusters, the better survival group overexpressed 3 genes in each category of angiogenesis, Wnt signaling pathway, cell adhesion, and cellular metabolism. Observing angiogenesis and cell adhesion gene expression in the same ccRCC cluster is opposite the results shown by

the Skubitz group. Additionally, the Brooks group did not perform a full pathway analysis on their unsupervised clustering results. Du to the limited biological information provided by these studies, we wanted to fully explore the molecular differences between our two subtypes, specifically using our previously identified core tumors to obtain the most accurate information about each group.

In addition to general molecular differences, we specifically wanted to examine the VHL/HIF pathway in our subtypes. As discussed in chapter 1, up to 90% of sporadic ccRCC tumors have inactivated the von Hippel-Lindau (VHL) tumor suppressor gene[133], so we would expect to see many HIF regulated pathways such as angiogenesis, glycolysis, and proliferation to be dysregulated compared to normal tissue. However, we also discussed that different *VHL* mutations regulate HIF proteins to varying levels[37,38]. Additionally, Gordan et al. showed that that ccRCC can be subdivided into wild-type VHL (negative for HIF expression), HIF1 and HIF2, and HIF2 only expressing tumors[39]. Using these distinctions, HIF2-only tumors underexpressed glycolysis genes and overexpressed cell cycle and DNA damage genes compared to wild-type and HIF1/HIF2 tumors. Both HIF groups overexpress angiogenesis and oxidative phosphorylation genes. Lastly, the opposing interactions of HIF1 inhibiting and HIF2 promoting C-Myc activity was confirmed in human tumors. This encouragement of C-myc activity enhances the proliferative signature seen in HIF2 only tumors. We wanted to know whether our data would stratify along these lines as well.

Our goal in this chapter was to identify what molecular pathways were differentiating our two inherent subtypes of ccRCC, ccA and ccB, and causing the survival differences that we observed. As expected, both groups overexpress cell cycle genes and underexpress oxidative phosophorylation compared to normal tissue. Interestingly, angiogenesis and hypoxia genes were predominantly overexpressed by ccA tumors, while ccB tumors tended to overexpress genes involved in epithelial-to-

mesenchymal transition and proliferation. ccB tumors underexpress glycolytic genes, compared to normal tissue. However, *VHL* mutations and methylation, as well as HIF1 protein expression, were observed in both subtypes. These results increase our understanding that the better prognostic ccA tumors appear to have more of a classic ccRCC phenotype of angiogenesis and hypoxia, while the poor survival group ccB displays a program of increased proliferation and aggression.

Results

Analysis of pathway differences between two core clusters. The previous identification of ccRCC subtypes[97] provides an opportunity to identify biological differences within the spectrum of ccRCC. SAM (Significance Analysis of Microarrays) analysis identified 2701 and 3512 probes over-expressed in ccA and ccB, respectively (Figure 3.1A). This result confirms the gene expression profile heterogeneity observed in previous studies[92]. The functional classification program, DAVID, was used to functionally categorize the probes identified in our analysis. A demonstration of the gene ontologies and pathways found to be differentially regulated between ccA and ccB tumors is provided in supplementary material on the Genes and Cancer website. Additionally, SAM Gene Set Analysis, a more statistically robust way of identifying correlated gene groups, was performed using curated gene sets, providing similar results. The most notable genes, gene sets, and gene ontologies associated with cluster ccA were involved in angiogenesis (Figure 3.1B), the beta-oxidation pathway (Figure 3.1C), organic acid metabolism, fatty acid metabolism (Figure 3.1D), and pyruvate metabolism. In contrast, core cluster ccB tumors overexpressed genes associated with cell differentiation, epithelial to mesenchymal transition (EMT) (Figure 3.1E), the mitotic cell cycle, TGF beta (Figure 3F), response to wounding, and Wnt targets (Figure 3.1G).

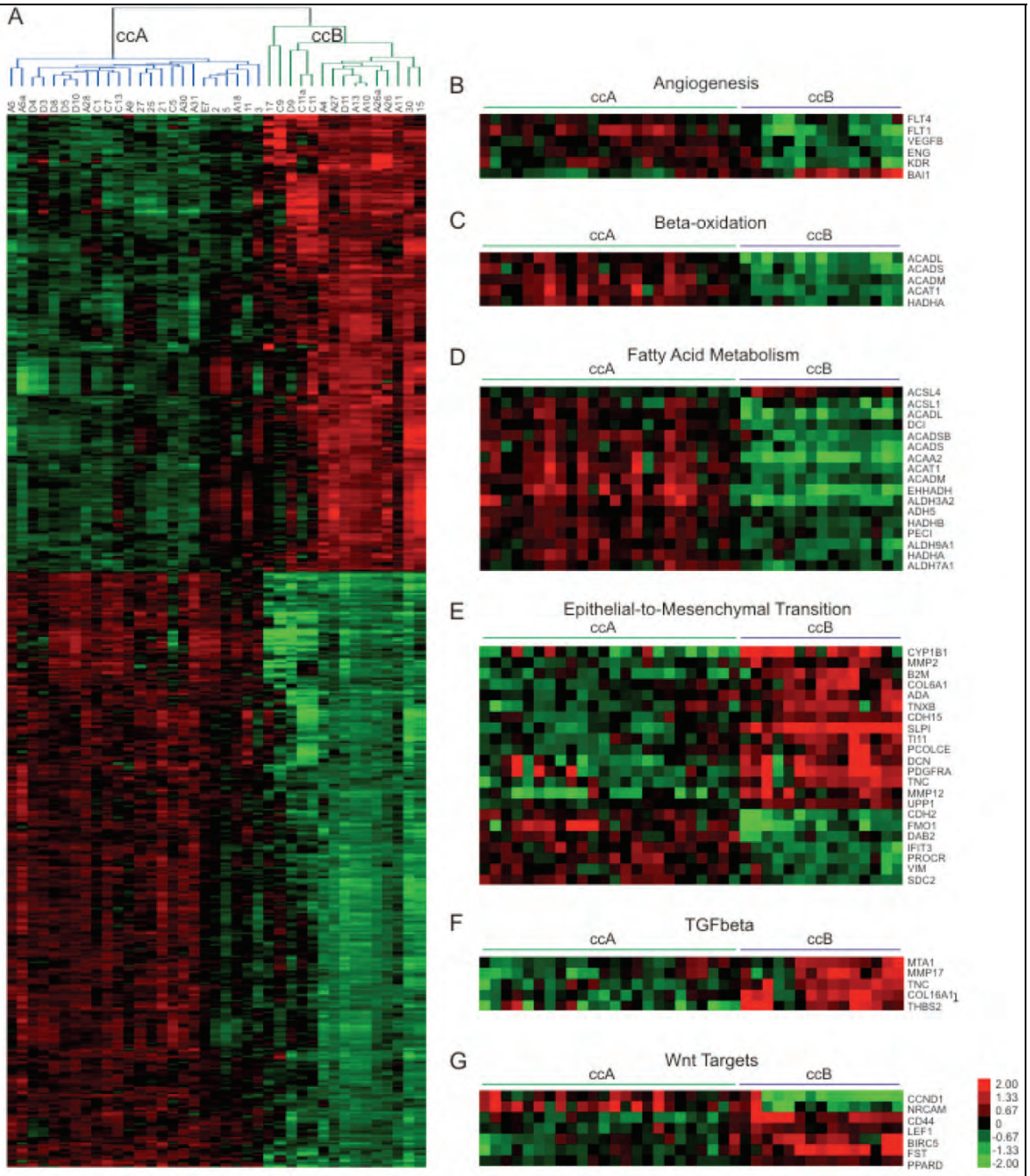

Figure 3.1 Pathway analysis of subtypes shows that ccA and ccB differentially express many genes

(**A**) Heat map of the 6,213 probes differentially expressed between ccA and ccB as determined by SAM analysis; false discovery rate (FDR) < 0.000001. (**B-G**) Magnified heat maps of the genes from (**A**) that populate the ccA (**B-D**) or ccB (**E-G**) overexpressed Molecular Signatures Database curated gene sets of Brentani angiogenesis (**B**), beta-oxidation (**C**), HSA00071 fatty acid metabolism (**D**), epithelial to mesenchymal transition (EMT) up (**E**), transforming growth factor beta (TGFβ) C4 up (**F**), and Wnt targets (**G**).

Confirmation of pathway analysis results on a validation set. We next confirmed these pathway differences between the subtypes using the 143 assigned validation tumors from the Brooks group[92] (Figure 3.2). ccA tumors continued to overexpress genes involved in fatty acid metabolism, glycolysis, and angiogenesis. Note that the Myc-related gene set in ccA is comprised of genes that are expressed by wild type levels of Myc as compared to transgenic overexpression[149]. In contrast, ccB tumors overexpress genes related to epithelial-to-mesenchymal transition (EMT), c-myc, cell cycle, Wnt targets, and TGF-beta.

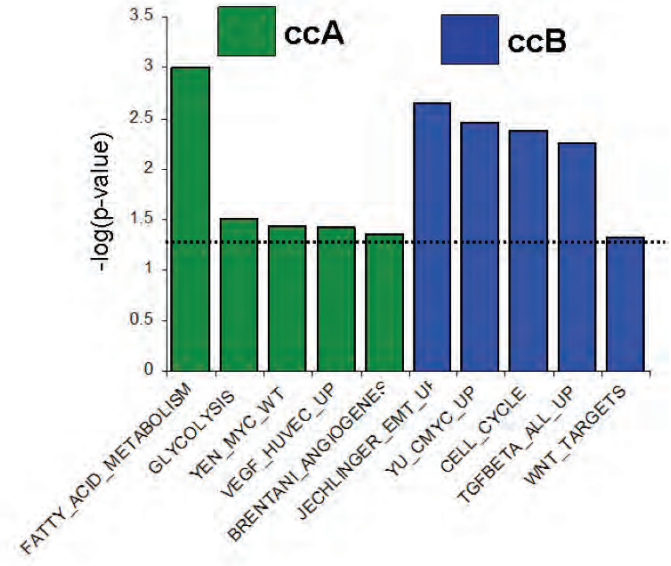

Figure 3.2 Pathway analysis of validation data subtypes mimics training data.
SAM Gene Set Analysis of ccA versus ccB tumors using Molecular Signatures Database curated gene sets shows similar pathway expression to those found in the training set (see Figure 3.1). The dashed line is equivalent to p=0.05.

Characterization of subtypes compared to normal tissue. The above analyses were performed comparing ccA and ccB tumors in relation to each other. To better understand the two subtypes, we wanted to compare each subtype to normal kidney tissue. Doing so would allow us insight into what different molecular changes are occurring in the subtypes during carcinogenesis, in addition to what makes the subtypes different from each other.

As expected, SAM analysis showed that both subtypes overexpress hypoxia-response related genes common to ccRCC, such as VEGF, EGLN3 and CAIX, while underexpressing *VHL*. In total, 9112 probes were differentially expressed in ccA tumors compared to normals, while 13, 571 probes were identified in ccB tumors.

When analyzed by SAM-GSA, a more complete picture is uncovered. Both subtypes overexpress cell cycle pathways, as well as p53 and caspase pathways (Figure 3.3A). Both subtypes underexpress oxidative phosphorylation and normal human kidney pathways, as expected for RCC. The ACE2 or angiotensin pathway is also similarly decreased in both. This pathway inhibits an increase in blood flow, thereby preventing

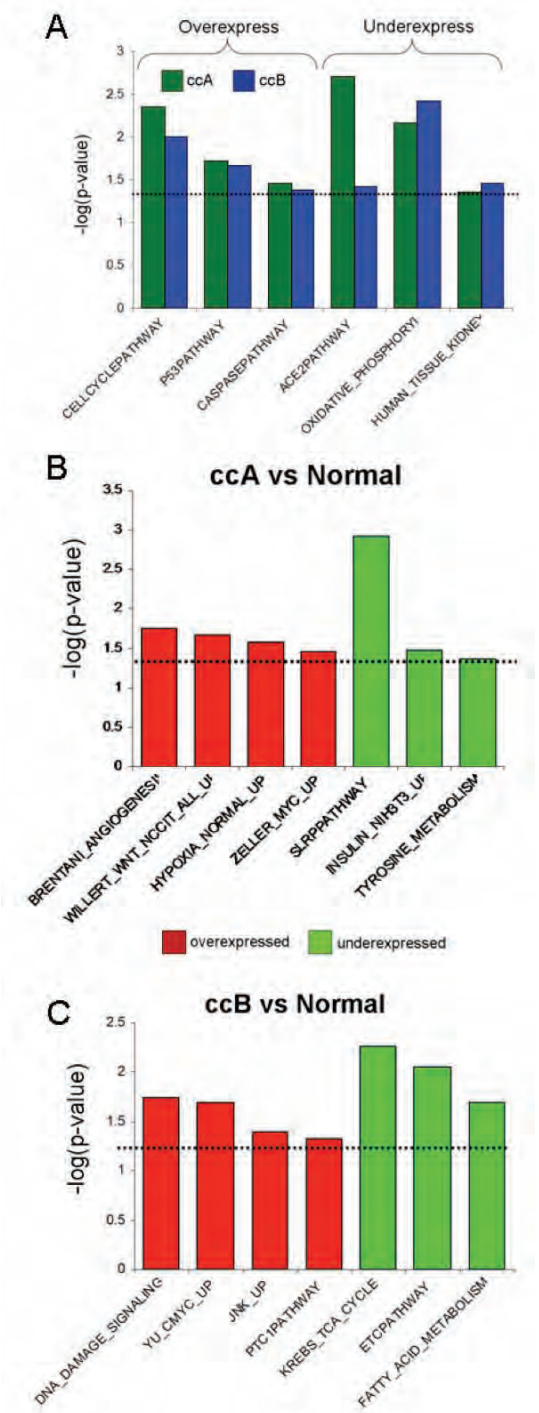

Figure 3.3 Pathway expression in subtypes compared to normal shows similarities and vast differences.
Subtypes were compared to normal tissue, and similarities between the subtypes were apparent (A) as well as the differences (B,C). ccA tumors are still angiogenic and hypoxic, while more aggressive ccB tumors underexpress metabolism pathways. The dashed line is p=0.05.

angiogenesis, a necessary step for tumorigenesis.

However, the differences are more telling. Only ccA overexpresses angiogenesis and hypoxia pathways (Figure 3.3B). This result is particularly interesting given that both subtypes overexpress VEGF and underexpress VHL, as described above. A group of Wnt3 targets are overexpressed in ccA versus normal, despite Wnt targets being overexpressed in ccB vs ccA. The SLRP pathway is underexpressed in ccA and is involved in the arrangement of collagen in the extracellular matrix; changes in this pathway are frequently associated with disease.

Surprisingly, ccA does not overexpress glycolysis and fatty acid metabolism genes as we previously thought; instead, ccB underexpresses these pathways. Given that ccRCC is classically known to overexpress both angiogenic and glycolytic pathways as they are both regulated via HIF, this observed split seems contrary to classical thought.

Myc responsive genes are overexpressed by both subtypes, but the genes in these two gene sets (Zeller Myc Up for ccA and Yu CMyc Up for ccB) are completely different. ccB tumors overexpress genes in the DNA damage signaling, JNK, and Patched1 pathway (Figure 3.3C), again suggesting a more aggressive and immature tumor. Overall, these results suggest that while both subtypes express certain common ccRCC gene and pathway alterations, ccB tumors have undergone more molecular modifications than ccA tumors.

VHL pathway analysis. As described above, we had found that several of the pathways overexpressed in ccA tumors are typically considered as being perturbed in ccRCC (*i.e.*, angiogenesis and hypoxia is considered a defining feature of ccRCC). A number of genes (*e.g.* EPAS1, EGLN3, PDGFC, HIG2, and CA9) tightly correlated

with aspects of *VHL* inactivation and hypoxia inducible factor (HIF) signaling were found to be overexpressed in ccA relative to ccB.

To further analyze the VHL pathway in our ccRCC subtypes, we classified each tumor of our previously published dataset[39] that was well annotated for *VHL* inactivation. Out of the 21 tumors, 10 were predicted to be ccA, 6 as ccB, and 5 as unclassified (Table 3.1). In each category, there were *VHL* wild type tumors, HIF1 and HIF2 overexpressing tumors and HIF2 only overexpressing tumors. Our own analysis of *VHL* status also demonstrated the presence of *VHL* mutations and/or methylation in both the ccA and ccB clusters (Table 3.1). These data suggest that ccA and ccB, despite both displaying *VHL* inactivation, might have activation of different dominant biologic pathways, resulting in distinct patterns of gene expression.

Table 3.1 Classification of HIF annotated Gordan et al.[39] tumors
Tumors from Gordan et al. were classified as ccA or ccB, and their HIF and VHL status assessed. *VHL* wild type tumors were negative for HIF1 or HIF2 expression. Each subtype has both tumors expressing HIF1, as well as those being wild type for *VHL*.

	HIF1 and HIF2	**HIF2 only**	***VHL* wild-type**
ccA	5	3	2
ccB	2	2	3
Unclassified	1	2	1

HIF1 protein is overexpressed in both subtypes. While we saw HIF1 and HIF2 overexpressing tumors in both subtypes in the Gordan et al. data above, we wanted to validate this wasn't a sample bias. Therefore, we performed IHC for both these HIF molecules on all available core tumors from the training data set (Figure 3.4). Again, we found a rather similar pattern of expression in both subtypes (Table 3.2).

From this result, we concluded that the presence of HIF1 protein may not be the primary distinguishing factor driving the pathway differences in the subtypes.

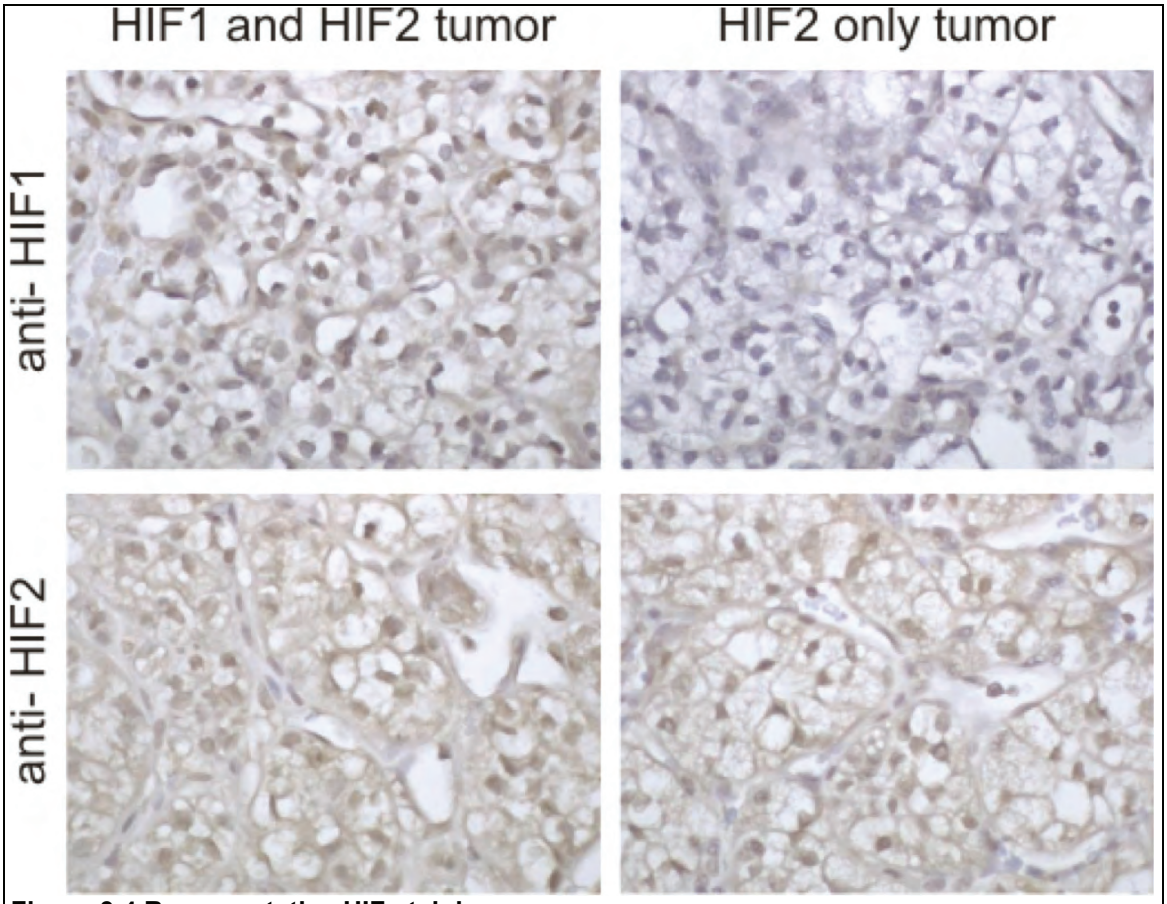

Figure 3.4 Representative HIF staining
HIF1/HIF2 and HIF2 only expressing tumors were found in both subtypes.

Table 3.2 Similar percents of HIF1/HIF2 tumors were found in each subtype.
HIF expression based on 5-10 fields with expression in greater that 40% cells. Each tumor was assessed by 3-4 independent histology readers. Tumors labeled n/a lacked slides or were indeterminate.

	HIF1 and HIF2	HIF2 only	HIF Negative	n/a
ccA	14	3	1	5
ccB	8	3	0	2

Discussion

We previously subdivided clear cell Renal Cell Carcinoma into two subtypes, ccA and ccB. While we had shown that patients with ccA tumors fared significantly better, we could not explain the cause of the survival difference.

Pathway analysis showed that the better prognosis ccA group relatively overexpressed genes associated with hypoxia, angiogenesis, fatty acid metabolism, and organic acid metabolism, whereas ccB tumors overexpressed a more aggressive panel of genes that regulate EMT, the cell cycle, and wound healing. This same pattern of pathways was observed in the Brooks validation dataset. When we compared these subtypes to normal tissue, again we saw that only the ccA tumors overexpressed angiogenesis and hypoxia pathways. Surprisingly, ccB tumors underexpressed glycolytic and fatty acid metabolism genes compared to normal.

While it is a subtle difference that ccB underexpresses glycolytic pathways rather than ccA overexpressing them, it is a very important observation as metabolic pathways are becoming more studied in cancer. Additionally, FDG-PET (fluorodeoxyglucose positron emission tomography) uses glucose uptake as a means of measuring tumor size in this generally highly glycolytic tumor type and is becoming a more acceptable marker of response to treatment[28]. Given that ccB tumors underexpress glycolytic genes, this particular marker may not be as effective as with ccA tumors.

Intriguingly, ccA overexpresses genes associated with components of hypoxia and angiogenesis pathways, processes known to be broadly dysregulated in clear cell RCC. *VHL* inactivation and subsequent activation of the hypoxia response pathway is so highly correlated with ccRCC that many of these pathways are expected to be upregulated in virtually all ccRCC tumors. As expected, using both training set tumors and LAD assigned gene expression arrays from Gordan et al.[39] we identified *VHL*

inactivation in both clusters. However, it is still possible that the two subtypes harbor different inactivating mutations of *VHL,* which could alter HIF expression or even affect HIF-independent pVHL activity.

The next obvious question was whether HIF1 expression was different between ccA and ccB. HIF1 transcriptionally regulates expression of glycolytic enzymes, and decreases C-myc activity. Gordan et al. had shown that HIF2 only expressing tumors were more proliferative, but less glycolytic[39], a pattern that closely matches ccB tumors. However, in both our tumors and through the Gordan et al. data, we saw that HIF1 was expressed in both subtypes, suggesting that HIF1 protein expression may not be the driving difference between ccA and ccB. There are several possibilities to address this apparent paradox:

1. There might be subtle differences in HIF1 expression between the two subtypes that we are unable to detect or quantify through IHC, but that shift the balance enough to create the resulting gene expression differences.
2. HIF1 may have specific mutations in ccB tumors that prevent full functionality, either in its transcriptional regulation role or its ability to bind Myc's binding partners. The Futreal group has found inactivating mutations within HIF1[40], which lends credence to this possibility.
3. HIF3 may be directly inhibiting HIF1 through binding of HIF1α or competitively inhibiting it through binding of HIF1ß in ccB tumors. HIF3 transcript is overexpressed in ccB tumors, providing support for this idea. This prospect is discussed further in Chapter 6.
4. ccB may have acquired additional genetic events which supplement *VHL* pathway events, contributing to a more biologically immature and

aggressive phenotype that overwhelms the signature associated with *VHL* inactivation.

ccB tumors overexpress a variety of genes and pathways, such as Myc, Myb, Hedgehog/Patched pathway, and Jnk pathway that are related to proliferation, differentiation, survival and migration. While on its own, each one of these may not create a vastly different expression pattern, these genes and pathways could act in tandem to cause the aggressiveness found in ccB tumors.

Another pathway that could influence the differences between ccA and ccB tumors is Wnt. Wnt targets are overexpressed in ccB tumors compared to ccA tumors. Interestingly, Wnt 3 targets are overexpressed in ccA tumors compared to normal. This contrast may be due to a different Wnt proteins and/or pathways being expressed in the two systems.

In particular, the Wnt kinase Ror2 is overexpressed in ccB tumors. This kinase has been shown to commonly interact with Wnt5[150-153]. Additionally, our lab has previously published[154] that it is an active kinase in clear cell Renal Cell Carcinoma. In tumors, we showed that its expression correlates with extracellular matrix genes and Wnt related genes. Specifically, Twist1 and MMP2 were validated by quantitative real-time as increasing in expression as Ror2 increased. *In vitro* experiments showed that Ror2 expression corresponded with the ability to fill a scratch wound, suggesting that Ror2 plays a role in migration. Ror2 expression was also necessary for proliferation of cells within soft agar, signifying a role in an anchorage independence and therefore invasion. Finally, a xenograft assay illustrated that Ror2 expression was necessary for the formation of discrete tumors *in vivo*. These results portend that Ror2 plays a definitive role in tumor aggressiveness, a key aspect of ccB tumors. Interestingly, this gene is also generally only expressed during development in kidneys, adding to the

immature status of ccB tumors. Further studies are needed to confirm Ror2's role in ccB tumorigenesis compared to ccA.

Overall, ccB tumors express genes and pathways more commonly associated with invasion, which fits well with their association with poor prognosis. In comparison, ccA tumors are predominantly defined by their angiogenic nature. This split suggests that the ccA/ccB classification scheme may also have predictive value, in addition to its prognostic value. Many of the current treatments for ccRCC, such as sunitinib, sorafenib, axitinib, pazopanib, and bevacizumab, are anti-angiogenic agents and were designed to primarily target the VEGF or the VEGF receptor. Given that ccA tumors do highly overexpress angiogenesis genes, it seems likely that these drugs may be more effective against these tumors. Future studies are planned to examine this intriguing possibility.

Methods

Gene expression data. Samples were collected and processed as previously published[97].

Pathway Analysis. SAM was performed, and genes were selected using a cutoff of False Discovery Rate (FDR) < 0.000001. Heat maps were generated using Cluster 3.0 (http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/) and Java Treeview (http://jtreeview.sourceforge.net/). Differentially regulated genes were functionally annotated in DAVID Bioinformatics Database (http://david.abcc.ncifcrf.gov/) with p-value and FDR <0.05. SAM-GSA was also performed on the data using the curated gene sets from MSigDB (http://www.broad.mit.edu/gsea/msigdb/) and $p < 0.05$.

VHL Sequence and Methylation Analysis. DNA was extracted from tumor samples using proteinase K (Roche) and standard phenol/chloroform extraction. VHL exons were PCR-amplified and directly sequenced for mutations with a BigDye Terminator Cycle kit on a 3130xl sequencer (Applied Biosystems). Primers and protocols used were described previously[155]. A CpG Wiz kit (Chemicon) and/or NotI digestion was used for methylation studies[156].

Immunohistochemistry. Immunohistochemistry staining was performed on formalin fixed paraffin embedded sections according to the protocol from Dako Catalyzed Signal Amplification (CSA) kit. Antigen retrieval for all antibodies was done by boiling the slides in citrate buffer (pH 6.0; Dako) for 30min. Endogenous peroxidase

activity was quenched in 3% H_2O_2 for 10min. Antibodies used were: anti-HIF-1A (rabbit polyclonal antibody, 1:2500, Novus NB100-479) , anti-HIF-2A (rabbit polyclonal antibody, 1:2000, Novus NB100-122). Detection for all antibodies was performed using the Dako Catalyzed Signal Amplification (CSA) kit. Slides were labeled as HIF1 or HIF2 positive if greater than 40% of the cells were positive for stain as determined by the majority of 3-4 reviewers.

Chapter Four:

Characterization of clear cell renal cell carcinoma subtypes reveals underlying genetic differences

Abstract

Clear cell renal cell carcinoma is the main subtype of kidney cancer, but we have shown that it can be further divided into two subtypes, ccA and ccB. These subtypes vary both in the primary molecular pathways that they overexpress as well as their natural disease progression in patients. However, we have been unable to understand why this is. Therefore, we sought to examine the underlying genetic changes between these two subtypes in an effort to better understand this disease heterogeneity.

Regions of altered expression were first studied in our training set data using positional gene sets and a technique called computational karyotyping. These approaches provided guideposts as to potential regions of copy number changes. However, as these regions were identified based on gene expression data, these regions may be altered due to other epigenetic or molecular events. For that reason, copy number analysis was then performed on a previously published dataset of ccRCC that were assigned to ccA and ccB subtypes using corresponding gene expression data. Loss of chromosome 3p, espousing the genetic location for *VHL*, was present in the majority of tumors in both subtypes, substantiating previous research and confirming the likelihood that most clear cell tumors arise from *VHL* inactivation. Interestingly, however, fewer ccB tumors contained this chromosomal deletion. Additionally, loss of chromosome 9 and 14 and amplification of 8q and X were more common in subtype ccB. Finally, gene mutation data from sequencing of these tumors was analyzed with respect to subtype, showing a prevalence of mutated histone modifications genes in subtype ccA. These data provide increased insight into the vast molecular, genetic, and epigenetic differences between subtypes ccA and ccB.

Introduction

In the previous chapters, we identified two subtypes of clear cell Renal Cell Carcinoma, termed ccA and ccB, which have a 6.6 year survival difference. Performing pathway analysis on these two groups, ccA tumors overexpressed more classic clear cell genes such as those in angiogenesis and hypoxia pathways, while ccB tumors overexpressed more immature and aggressive genes, including those in cell cycle, cell differentiation, and response to wounding. However, we still lack the answer to why these two subtypes behave in these ways.

As discussed in chapter 1, many groups have attempted to answer the question of what genetic events cooperate with VHL loss to accelerate tumor progression. Much as we do early in this chapter, Furge et al. attempted to understand the underlying genetics through predicted cytogenetic profiles based on gene expression data[6]. They found that loss of 14q correlated with higher stage and poor survival.

Additionally, copy number analysis and sequencing has become more common, allowing a more detailed examination of the underlying genetic changes leading to the biological impact of the tumors. Using unsupervised hierarchical clustering on comparative genomic hybridization (CGH) data from 51 ccRCC tumors, the Kanai group identified 2 distinct subclasses with significant survival differences[63]. They found that both clusters had frequent loss of 3p and gain of 5q and 7, but one of the poor prognosis clusters additionally sported loss of 1p, 4, 9, 13q, and 14q. Interestingly, this cluster also had increased DNA methylation on 9 analyzed genes.

More recently Beroukhim et al. and the Futreal group studied a larger number of ccRCC tumors using both gene expression analysis and single nucleotide polymorphisms [40,65]. Beroukhim et al. analyzed 54 sporadic ccRCC tumors as well as 36 tumors from patients with von Hippel-Lindau disease. Overall, they identified loss of 3p

in almost all tumors, and gain of 5q in most. Additionally, they found amplifications in 1q, 2q, 7q, 8q, 12q and 20q, and deletions in 1p, 4q, 6q, 8p, 9p, and 14q. Sporadic tumors tended to undergo more copy number changes than VHL disease associated tumors, but share many of the same copy number changes.

The Futreal group analyzed 96 tumors by copy number analysis, and found frequent losses of 3p (in greater than 80% of cases), 4, 6p, 8p, 9p and 14q, and gains of 1q, 2, 5q (in 50% of tumors), 7, and 12. Importantly, they found two main subgroups within their expression data of these tumors, with 82% of their tumors exhibiting a hypoxia signature. The Futreal group also performed sequencing on 3544 genes and identified mutations in a number of histone modification genes, the majority of these changes being found in the subgroup with a more hypoxic expression pattern.

The Kanai group identified two clusters based on copy number, but we cannot confirm that these correspond to our subgroups without gene expression data. The other two groups (and many others, as discussed in chapter 1) focused on overall genetic changes compared to normal[40], and the differences between sporadic and germline mutations [65].

However, with our panel of 120 genes, we can classify tumors into our ccA and ccB subgroups to define the genetic steps that guide the heterogeneity that is seen in many gene expression studies and in the clinic. Therefore, in this chapter, we began by assessing possible copy number changes from our gene expression data through both SAM and a new technique called computational karyotyping. To confirm these results, we used the data from the Futreal group[40] to perform copy number analysis. In doing so, we determined that ccB tumors are more likely to have deletions on chromosome 9 and 14, regions associated with decreased survival, while ccA tumors are slightly more likely to have 3p deletions, an alteration associated with better prognosis[66]. Finally, using Futreal's list of mutated genes for each tumor, we identified alterations to histone

modification genes in both subtypes, but far more in the ccA subtype. These results help us to have a better understanding of the underlying genetic and epigenetic changes driving the molecular and clinical differences observable in our two subtypes.

Results

Analysis of chromosomal changes based on expression. We previously identified substantial transcript and pathway differences between ccA and ccB tumors. However, we still could not definitively state the cause of these changes. Therefore, we wanted to determine whether the transcripts with altered expression clustered in specific chromosomal regions, suggestive of amplification or deletion. If so, these results might give us further information as to other genetic changes besides VHL inactivation that lead to tumorigenesis and the varied outcomes between subtypes.

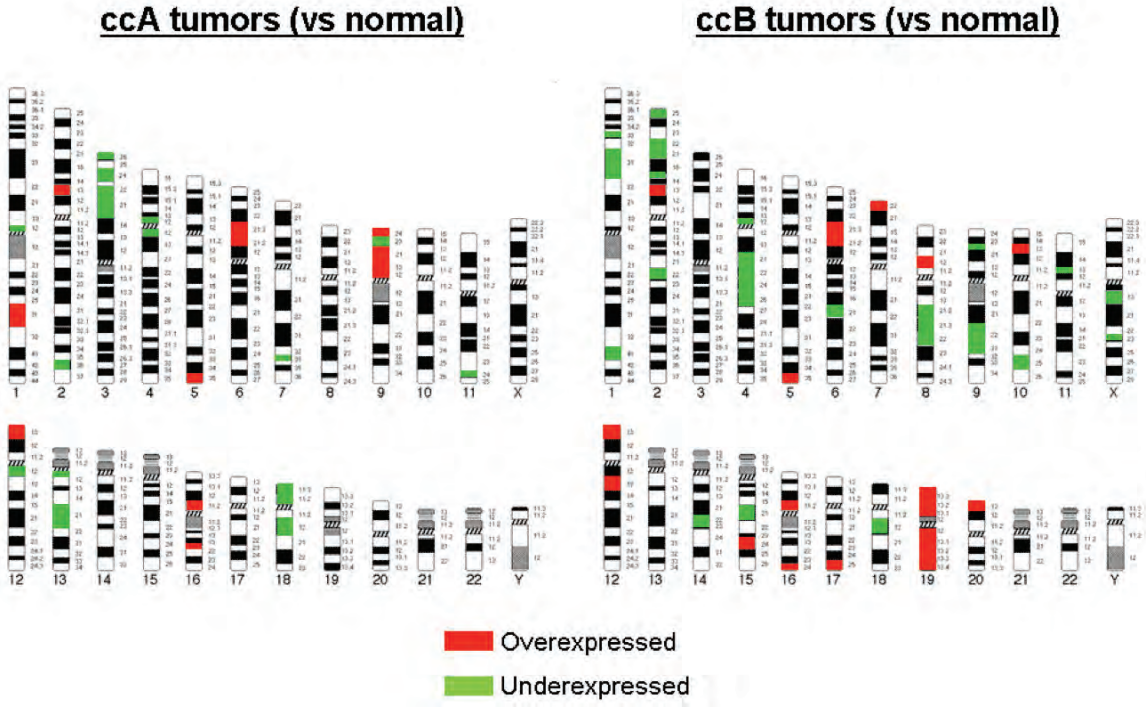

Figure 4.1 Chromosomal regions of differential gene expression.
Regions of differentially expressed genes compared to normal tissue were identified using SAM-GSA and the Broad Institute's MSigDB positional gene sets, $p<0.05$.

SAM gene set analysis was performed on our training set data compared to normal tissue using the Broad's positional gene sets (Figure 4.1, Table 4.1). Very few regions were similar between the two subtypes, ccA and ccB. Interestingly, ccA tumors

show underexpression of genes in much of chromosome 3p, the location of *VHL*. Deletions of 3p have been shown to correlate with increased survival[66]. ccB tumors show decreased expression on chromosome 14, near the location of HIF1. While the identified regions of over- and underexpression may be indicative of chromosomal changes, it is also possible that pathway alterations or epigenetic modifications could be driving these differences.

Computational karyotyping of training set data. Use of the Broad Institute positional gene sets with SAM-GSA is based on probes with matching gene names within specific chromosomal bands. Thus, probes without any gene names and genes not listed within certain gene sets will not provide added information. Therefore, we turned to a method we named computational karyotyping to provide more detailed information via a sliding window technique. Computational karyotyping compares probe expression in tumors compared to normal tissue, then it identifies regions along the chromosome where there is an enrichment of outliers compared to background. We could then examine regions that were different between ccA and ccB.

When we applied this technique our original training data, many regions that were previously identified by SAM-GSA were also apparent (Table 4.1). These regions, therefore, are most likely to truly be regions of changed expression. Chromosome 3p, in particular, stood out to us, since there was almost no change in expression compared to normal for ccB tumors, yet up to 96% of ccA tumors showed underexpression in this region (Figure 4.2). As mentioned above, loss of 3p has been shown to correspond with increased survival[66]. Additionally, genes located within this region include *VHL;* the TGF-beta receptor 2 (TGFBR2), which regulates transcription of proliferation genes; Beta-catenin, which regulates c-Myc, cyclin D1 and many other genes; SETD2, a

histone methyltransferase; Wnt5a, which has been shown to interact with Ror2; and ADAMTS9, a metalloprotease and anti-angiogenesis gene. Loss of these genes could easily cause a tumor to be more angiogenic and less aggressive.

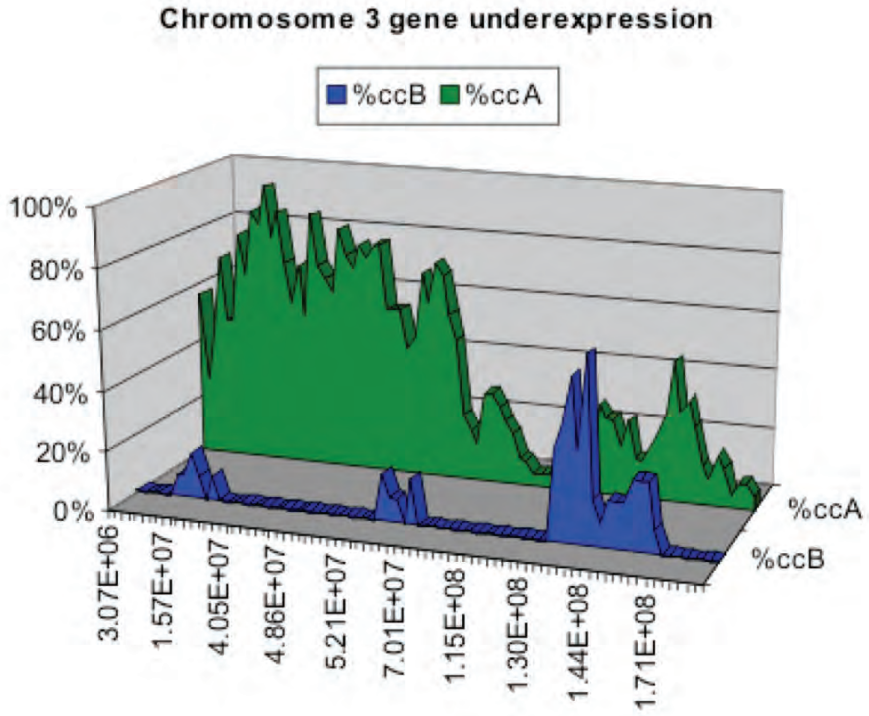

Figure 4.2 Chromosome 3 underexpression of UNC data shows significant difference between ccA and ccB tumors.
Chromosomal karytopying of UNC expression data identified chromosome 3 as having significant changes in regional gene expression. Additionally, ccA tumors show far greater loss of expression in this region, *t* test $p<0.0001$.

One other region of interest that was identified by both SAM-GSA and computational karyotyping is underexpression in 14q in ccB tumors. Loss of this region has been identified to be associated with poor prognosis[66]. However, as mentioned above, this identified regions is based on expression data, which could be caused by either copy number changes, pathway alterations, or epigenetic modifications.

Table 4.1 Regional expression changes of training set data by computational karyotyping and SAM-GSA

UNC training set expression data was used to locate regions of over- and underexpression in ccA and ccB tumors compared to normal via SAM-GSA and computational karyotyping. SAM-GSA regions were chosen with p<0.05. Regions listed under computational karyotyping with q<0.05 and greater than 25% of tumors within a subtype. Bolded regions are identified by both techniques.

Overexpression				**Underexpression**			
SAM-GSA		Computational karyotyping		SAM-GSA		Computational karyotyping	
ccA	ccB	ccA	ccB	ccA	ccB	ccA	ccB
1q31				1p12	**1p31**	1p31.1-p22.3	**1p31.3-p22.3**
2p13	2p13			2q36	1p33		**2q22.3-24.3**
5q35	5q35	5q32-35.3		**3p21**	1q41		2q31.1-32.2
6p21	**6p21**	6p22.2-21.31	**6p22.2-21.31**	**3p22**	2p15	**3p26.2-q11.2**	3q22.2-25.1
	7p22		7q22.1	**3p24**	2p21	4p14-q12	**4p16.1-q12**
	8p12		8q24.3	**3p26**	2p22	4q21.22-q23	**4q13.3-q27**
9p			9q34.13-34.3	**4p12**	2p25		4q31.3-35.2
	10p13			**4q12**	**2q22**		**6q21-q23.1**
			11p15.5	7q33	**4p11**		7q22.3-31.33
12p13	12p13			9p23	**4p12**	8q13.3-q21.3	**8q13.1-q22.2**
16p11	12q13			11q24	**4q21**		**9q22.33-q32**
	15q24			12q12	**4q22**		11p15.1-p11.2
	16p11	16p11.2	**16p13.3-11.2**	13q11	**4q23**		12p13.1-p11.22
16q21	**16q24**		**16p24.1-24.2**	13q21	**4q25**		13q13.3-q14.2
			17q12	18p	**4q26**		**14q12-q23.2**
	17q25		**17q25.1-25.3**	18q12	**6q21**		**15q15.3-q21.3**
	19p13	19p13	**19p13.3**		**8q21**		**18p11.21-q21.2**
	19q		19p13.2-p13.12		**8q22**	X21.1-22.2	X21.1-22.2
			19q13.32-13.43		9p22		
	20p13		**20p13**		**9q31**		
			20q13.33		10q25		
			22p11.1-11.22		11p13		
			Xp11.3-11.23		**14q22**		
			Xq26.3-28		**14q23**		
					15q21		
					18q12		
					xq13		
					xq23		

Assigning subtypes in a validation dataset. Because we could not determine whether these regional expression changes were due to chromosomal loss in our data, we turned to a previously published dataset of gene expression and SNP data by the Futreal group[40]. To identify regions differentially expressed by each subtype, we

first classified each tumor as ccA or ccB using gene expression data from 100 genes corresponding to the 120 LAD probes we previously identified[97]. This data was clustered with ConsensusCluster and ambiguous arrays were set aside, to allow identification of copy number results characteristic for each subtype. The resulting consensus matrix and PCA plot of 44 ccA and 17 ccB tumors shows two distinct and robust subtypes (Figure 4.3). Given our previous analysis that ccA tumors overexpress hypoxia genes and that the majority of Futreal tumors did likewise, this preponderance of ccA tumors makes sense.

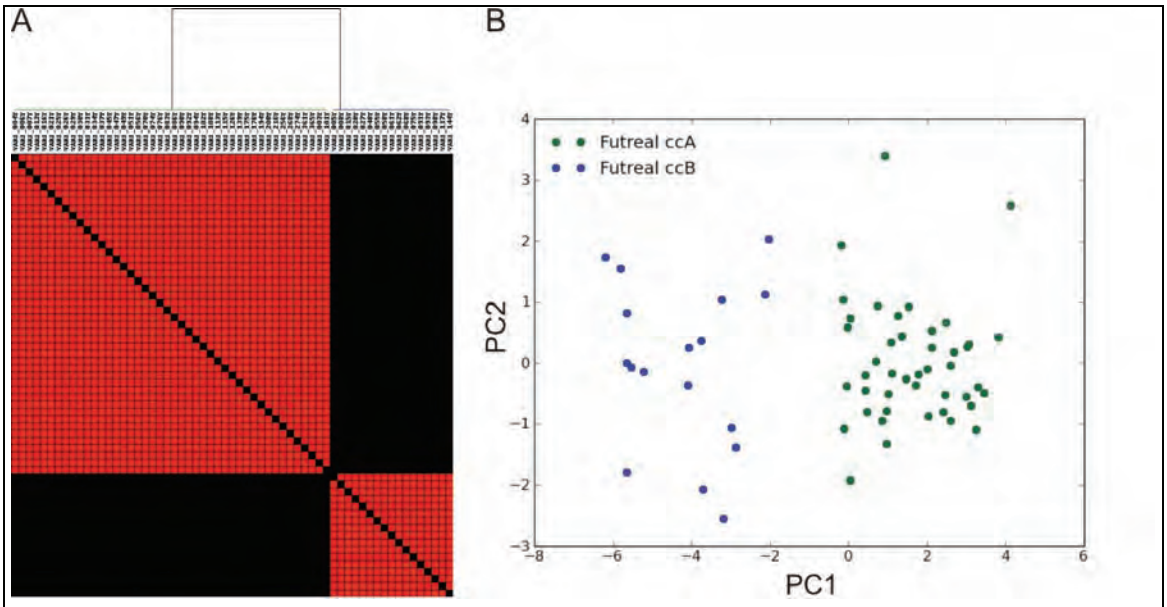

Figure 4.3 Consensus matrix and PCA plot demonstrate distinct clusters in Futreal data. A consensus matrix (A) and PCA plot resulting from 100 genes corresponding to 120 LAD genes show two distinct clusters – 44 ccA tumors and 17 ccB tumors.

Computational karyotyping of Futreal data. To confirm the results we had seen before, we wanted to first perform computational karyotyping on the Futreal data. Doing so would allow us to determine if there are regions with significant regions of gene expression changes that match our previously observations. Additionally, if some of the

regions don't match by copy number, we will need to consider epigenetic changes or whether several genes within altered pathways are located in one region.

Looking at the computational karyotyping data, three regions were of particular interest (Figure 4.4): On chromosome 3, there are still fewer ccB tumors than ccA tumors that contain underexpression. However, this dataset contained a large percentage of ccB tumors with underexpression, suggesting that the extreme loss of 3p in our dataset may have just been sample bias.

Chromosome 14q was a region of gene underexpression recognized in our training set data. As previously observed, more ccB tumors display underexpression in this region than ccA tumors. The same pattern is observed for chromosome 9p. Both of these regions are associated with decreased survival.

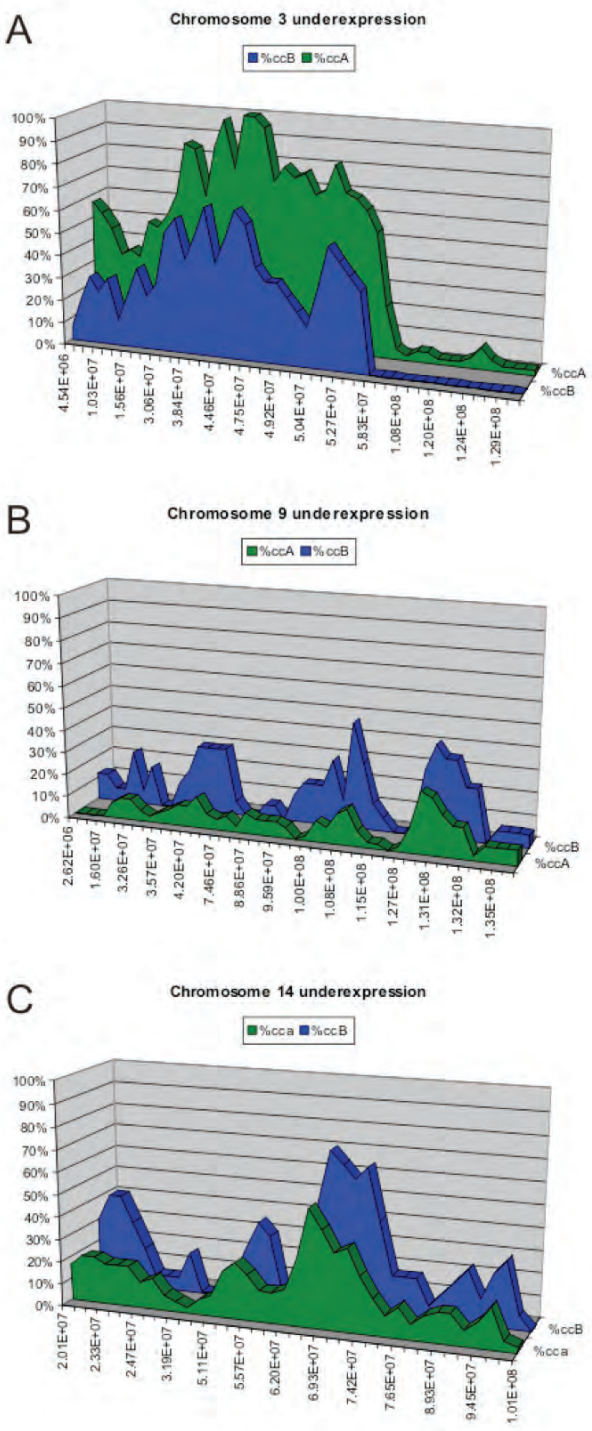

Figure 4.4 Computational karyotyping of Futreal expression data
Regions of gene underexpression as compared to normal ($q<0.05$) for chromosomes 3, 9, and 14. These regions are significantly different between ccA and ccB, t test $p<0.001$.

Deciphering chromosomal changes with SNP data. We next performed copy number analysis on the Futreal data using the program SWITCHdna (https://genome.unc.edu/pubsup/SWITCHdna/) to determine whether the regions of over- and, particularly, underexpression seen above are caused by amplification or deletion events, respectively. As can be seen in Figure 4.5, the majority of amplifications and deletions are extremely similar between the two subtypes. In particular, the majority of tumors in each subtype have deletions of chromosome 3p, as observed in the chromosomal karyotyping. This result is consistent with previous classical cytogenetic, CGH, and SNP reports identifying 3p as a common deletion for renal cell carcinoma. Additionally, this result lends support to the idea that almost all ccRCC tumors arise through VHL loss/inactivation. However, 3p is still deleted in approximately 30% more ccA tumors than ccB tumors.

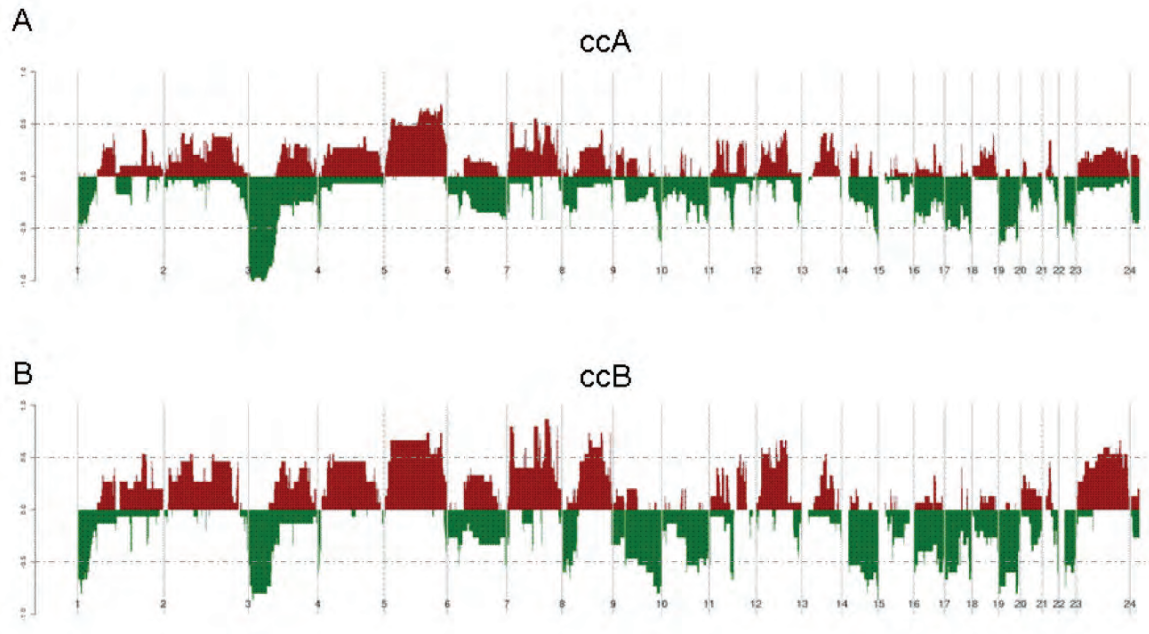

Figure 4.5 Copy number analysis identies regions of dissimilarity between subtypes. Copy number analysis was performed on SNP data from the Futreal group[66] using a cutoff of $p<0.0001$. A greater percentage of ccB tumors are contain deletions of chromosome 9 and 14, and amplifications of chromosomes 8 and X.

Looking at amplifications and deletions beyond chromosome 3p, the genome is far more stable, both in general and between subtypes. There are a few notable exceptions, however: ccB tumors show increased deletions on chromosomes 9 and 14 and amplification of 8q and the X chromosome compared to normal. Statistical analysis is necessary to determine if other copy number is different in other regions. Additionally, further analysis is necessary to determine what genes may be targeted in these regions of amplification and deletion.

Mutation analysis suggests epigenetic differences between subtypes.

Chromosome amplification and deletion alone is unlikely to completely explain the subtype differences we have found. Therefore, we analyzed the list of mutated genes for each of the Futreal tumors, which were derived from exon-sequencing of 3,544 genes. Each subtype has a median number of 2 mutations per tumor, with a total of 84 different genes mutated in the 44 ccA tumors and 58 genes in the 17 ccB tumors (Table 4.2). Only 3 genes are mutated in both subtypes: AKAP4, which binds protein kinase A; SETD2, a histone H3K36 methyltransferase; and *VHL*. Mutations in *VHL* occur equally in ccA and ccB (52% and 47%, respectively), and AKAP4 was mutated only once in each subtype. In contrast, SETD2 was mutated in 7 ccA tumors and only 1 ccB tumor (16% and 6%, respectively). Interestingly, histone H3K4 demethylase *KDM5C*, encoding JARID1C, was mutated twice, and histone H3K4 methyltransferases MLL and MLL2 were each mutated once in ccA. Two other histone modification genes, NSD1 and RNF20 and two other genes associated with histone modification, BTG2 and BCOR, were mutated once in ccA. In contrast, ccB tumors have only the one SETD2 mutation, one mutation of a histone H3K4 methyltransferase gene, MLL4, and one mutation of a

gene associated with a histone modification complex, UTX. The prevalence of mutations in histone modification genes, particularly in ccA tumors, suggests that epigenetic modification may be driving some of the biological differences between the two subtypes.

Table 4.2 Mutated genes in each subtype
A prevalence of histone modification genes (bolded) are present in ccA tumors. For this analysis, 44 ccA tumors and 17 ccB tumors were analyzed. If the gene was mutated in multiple tumors, the count is listed after the gene.

Subtype	Mutated genes
ccA	ADAM32, ADAMTS18, AK3L1, AKAP4, APC, ARNT, **BCOR**, BIRC7, BMPR1A, **BTG2**, C19orf2, CAD, CADM2, CDKN2C, CLSPN, CSMD3, CTSZ, DDB2, DDX20, DDX23, DDX27, DEK, DGKK, DKK1, DNAJC18, ENPP2, ERCC3, FGD5, FRS2, GPLD1, ICK, IER2, IGBP1, IPO4, ITGA10, ITPR2, **JARID1C (2)**, KCNV1, MAFG, MAGI1, MCF2L, MDC1, MDH1, MED1, MERTK, MGA, **MLL, MLL2**, MMP10, N4BP2, NOTCH2, NOV, **NSD1**, NUP188, NUP98, OCRL, PDHX, PFTK2, PINX1, PIP5KL1, PLCB2, PLEKHA5, PTBP1, PTPLB, PTPN11, RABGAP1, RIOK2, **RNF20**, SBK1, SEMA4B, SERPINB10, **SETD2 (7)**, SMG6, STCH, TEK, TOPBP1, TRPS1, UBA5, USP24, USP53, VHL (23), VPS13B, WNK3, XPO4
ccB	AFF4, AKAP3, ARFGEF1, ARHGEF11, ASB8, BUB1B, CCR3, CNKSR1, DHX8, ESRRG, FAM5C, FBXO28, GDF11, KLK3, **MLL4**, MMP16, MMP3, MYBL2, NCAPD2, NLRP5, PC, PGM1, POU2AF1, PPP2R2C, PTPN22, PTPRF, PTPRJ, **SETD2**, SPHK1, TMEM74, TNKS2 (2), TRAD, TRIM32, USP24, **UTX**, VAV1, VHL, VTN, VHL (9)

Discussion

We had previously indentified two distinct molecular subtypes of clear cell renal cell carcinoma, with significant survival outcomes (8.6 years for ccA tumors vs. 2 years for ccB tumors), and substantially different expression of pathways. However, we wanted to more fully understand what is driving the molecular and survival differences between the two subtypes. Therefore, we sought to identify what underlying genetic changes were occurring in ccA and ccB.

In the absence of copy number data for our training set, we began by using SAM-GSA and a technique called computational karyotyping. Both of these methods present means of identifying chromosomal regions of significant over- and underexpression of genes. Chromosomal karyotyping appears to present a more detailed means of identifying altered regions as it works through a sliding window technique. Additionally, the sliding window allows you to asses the impact of a specific number of genes on the proposed expression pattern by examining which genes fall in the overlapping windows. Computational karyotyping is a good technique to start with in the absence of copy number data.

However, we were able to obtain copy number data. SNP analysis of these subtypes in a previously published dataset identified differences in copy number on chromosomes 9, 14 and X. Through analyzing which genes are mutated in the tumors in each of these subtypes, we noted a prevalence of histone modification genes are mutated in ccA. These results provide us with important means of understanding the biological heterogeneity present in clear cell renal cell carcinoma.

The increase percent of deletions of chromosomes 9 and 14 in subtype ccB is of particular interest. Chromosome 9 or 9p deletion, in particular, has been previously identified as being associated with metastases and/or prognosis by several groups[7,47,52-

[54,60,157]. More recently, Klatte et al. performed a prospective study using classic cytogenetics on tumors from 246 ccRCC patients and discovered that both 9p and 14q loss were prognostic by univariate analysis, but only 9p remained under multivariate analysis[66]. An expanded study by this same group confirmed the prognostic significance of 9p deletions and provided recurrence data[67].

A combined SNP and gene expression analysis of familial and sporadic tumors identified the only gene within chromosome 9's to be focally and homozygously deleted as *CDKN2A*, which encodes known tumor suppressors p16 and mdm2[65]. *CDNK2B* was also indicated by deletion analysis, but not by the corresponding gene expression analysis which also failed to detect *VHL* underexpression. Intriguingly, *CDKN2A* is overexpressed in both subtypes compared to adjacent normal in our gene expression analysis. This discrepancy may be caused by tumors which lack the 9p deletion skewing the results. Alternatively, these genes may be more strongly targeted for deletion in tumors caused by germline *VHL* mutations, a large percentage of the Beroukhim data. Regardless, this region does deserve further study.

Deletions of 14q also hold interest, primarily due to the location of HIF1 at 14q23.2, a region deleted in 60% of the Futreal ccB tumors and 25% of the Futreal ccA tumors. Gordan et al has shown that HIF1 inhibits C-Myc activity[39], which aligns with the ccB phenotype. However, we have previously assigned the arrays from the Gordan et al study and have shown that ccA and ccB do not split along HIF1 and HIF2 vs. HIF2 only lines[97]. These results are confirmed again in this study, where a mix of HIF1/HIF2 and HIF2 only tumors is present in both subtypes. One possibility is that subtle expression changes of HIF1 and HIF2, undetectable by IHC, could be creating a transcriptional imbalance and causing the Myc dysregulation apparent in ccB tumors. Another explanation is that Myc may be commonly amplified in ccB tumors, a possibility given Myc's overexpression and the increased amplification of 8q in ccB tumors. A

combination of both HIF expression subtleties and Myc overexpression could be providing a strong impetus towards the pathway divide between ccA and ccB. Also, as we saw in chapter 3, both ccA and ccB overexpress Myc target genes as compared to normal tissue, but these specific target genes are different in the two groups.

Yet another hypothesis of a cause of the rift between ccA and ccB tumor biology may be epigenetic changes. This option bears much consideration, particularly given the prevalence of mutated histone modification genes in ccA tumors. Whether all of these mutations are complete loss-of-function is currently unknown and will need to be followed up. However, the histone methyltransferase SETD2, the gene that is most commonly mutated in ccA is located on chromosome 3p in a region deleted 20% more frequently in ccA tumors than ccB. While the inactivation or deletion of this particular gene might be a side-effect of inactivating VHL (also located on 3p), the preponderance of mutations in other histone modification genes suggests that these genetic changes are important for the progression of the tumor.

Additionally, as mentioned earlier, the Kanai group has shown that tumors with increased deletions of chromosome 9 and 14q also had increased DNA methylation on CpG islands[63]. This group of tumors also showed poorer prognosis. While histone methylation and DNA methylation work via different means and create opposing transcriptional outcomes, their work leads further credence to the clear cell subtypes being epigenetically different.

Methylation arrays, even on a limited scale, will provide a great deal more information about these subtypes. The arrays will be focused on regions where gene expression and copy number changes do not correspond. For example, SAM-GSA of gene expression data suggested that several regions of chromosome 2 and 4 in ccB tumors are underexpressed, but SNP analysis shows little or no changes in these regions. While this discrepancy may be caused by sample set variation, it may also be

due to epigenetic modification. More information about the epigenetic landscape of these tumor subtypes could give additional guidance in drug discovery.

ccA and ccB tumors share a great number of similarities in genetic and molecular makeup, causing them to look the same histologically. However, as this paper elucidates, their differences are caused by large chromosome changes and very possibly unique epigenetic modifications. These differences likely cause the inherent heterogeneity seen in natural progression, cytokine therapy, and molecularly targeted therapy. Additionally, our results suggest that the putative subtypes identified by varying groups through different technological means are the same. Understanding the points of convergence and divergence of these subtypes could lead to enhanced clinical decision making.

Materials and Methods

Gene Expression Data. UNC and Stanford samples were collected and data processed as previously published[97]. Futreal samples were collected as previously published[40]. Futreal series II raw gene expression data was downloaded from Gene Expression Omnibus (GEO) and mas5 normalized in Genespring v11 (Agilent, Santa Clara, CA). By both Genespring quality control and ConsensusCluster (http://code.google.com/p/consensus-cluster/), three arrays (Vari_038T, Vari_046T, and Vari111T) were found to be very distinct from the others. Data was reentered into Genespring without these arrays and filtered such that 41/92 arrays must have present calls to force inclusion of *VHL*. By ConsensusCluster, tumors Vari_052T, Vari_398T, Vari_071T, Vari_032T, Vari_078T, Vari_087T, Vari_352T, and Vari_085T clustered with normal tissue and were removed. Data was re-annotated with Affymetrix v30 annotation file. Duplicate probes with fewer present flag and lower median values were removed. Data was standard normalized by subtracting array mean and dividing by array standard deviation.

Computational karyotyping. Computational karyotyping is an adaptation of Cancer Outlier profile Analysis (COPA[158]), a method of locating genes with aberrant expression levels in a small subset of samples by looking for outlier values for each probe set across all samples and then ranking them according to their frequency. In order to detect chromosomal amplifications/deletions we modified this algorithm to look for outliers across the whole genome for each sample separately. Since this particular data set contains normal controls, outliers were defined with respect to the normal set if they were in the top or bottom 25% quartile for high or respectively low outliers. Outlier profiles thus obtained are organized in two binary matrices B_1 and B_2 corresponding to

high and respectively low outliers. For both matrices, *B(i,j) = 1* if *gene i* is an outlier in *sample j* and *B(i,j) = 0* otherwise. We then ordered the genes according to their chromosomal position and looked for regions enriched in outliers as an indication of chromosomal aberration. For each column in matrix *B*, outlier enrichment is computed with Fisher Exact test[159] in a sliding window 50 genes wide with a pace of 10 genes. Varying the size of the window and/or pace by 5-10 genes does not significantly affect the result. Benjamini-Hochberg method[160] is used to control the false discovery rate to 5% by converting the p-values from the Fisher Exact test to q-values. For each array chromosomal regions with q-values < 0.05 are marked as potential amplifications and then ordered by their frequency amongst the sample set.

Pathway and Positional Analysis. Significance Analysis of Microarrays (SAM, http://www-stat.stanford.edu/~tibs/SAM/) was performed, and genes were selected using a cutoff of False Discovery Rate (FDR) < 0.000001. SAM-GSA was also performed on the data using the curated and positional gene sets from the Broad Institute's MSigDB (http://www.broad.mit.edu/gsea/msigdb/), p < 0.05.

Distance Weighted Discrimination (DWD) of Futreal and UNC Data. To assign Futreal tumors to subtype ccA or ccB, Futreal and UNC data was DWD-combined. UNC data[97] was re-annotated from Agilent's 20100115 annotation file, and duplicate genes (by Entrez GeneID) by lower median were removed. Gene lists for Futreal and UNC were compared, and non-matching genes were removed. Remaining data was combined using Java DWD (https://genome.unc.edu/pubsup/dwd/), using non-standardized DWD (to prevent column standardization) and centering to the mean of the UNC data. Figure 4.6 displays PCA plots, showing that the data was successfully adjusted.

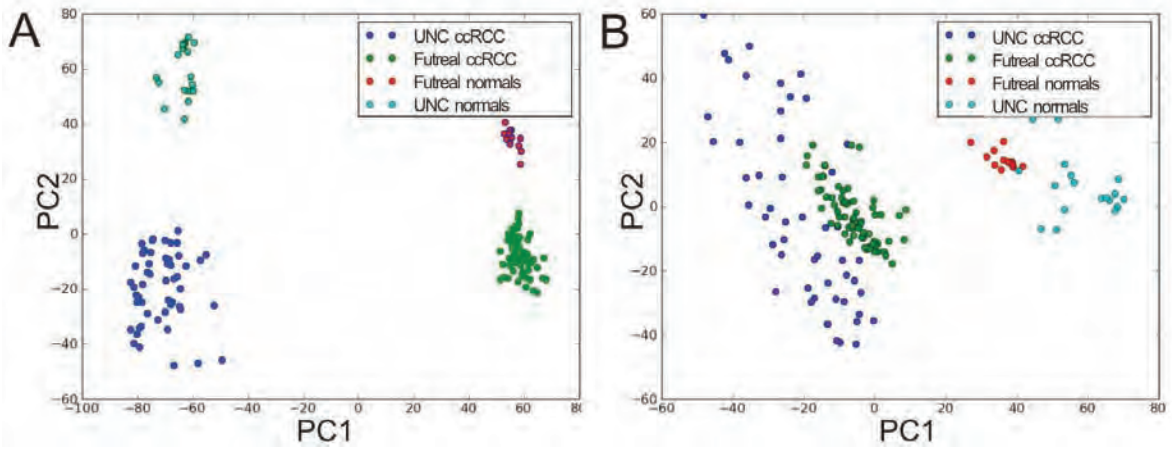

Figure 4.6 DWD adjustment of UNC and Futreal data
PCA plot of the first two principle components, showing the UNC and Futreal data widely variable (A) and adjusted by DWD to be comparable (B).

SNP analysis. SNP data was as previously published[40]. Raw data corresponding to series II clear cell tumors and matched normal tissue was loaded into Genespring v11. PD2135a was removed due to extreme dissimilarity by PCA. Copy number analysis was run, pairing tumors against matched normal tissue, and LogR data (ratio of observed to expected hybridization intensity) was exported for SNPs with data values for all 51 tumors. This data was analyzed in SWITCHdna (https://genome.unc.edu/pubsup/SWITCHdna/), using an F-statistic threshold of 11% and 10% chromosome size cutoff for gain/loss. Resulting copy number changes were calculated per subtype, using an intensity threshold of 0.1 and a Z-score cutoff of 5 ($p<0.0001$).

Acknowledgements

A great deal of thanks goes to Grace Silva for teaching me SWITCHdna and to Jeremy Simon for teaching me Cygwin and writing a PERL script. Kate Hacker was kind enough to get Switch running on the server for me from a stable Ethernet connection, while I was at home trying to write this thesis. She also began running SAM-GSA on her computer since it's tremendously faster than mine.

Chapter Five:

Development of an FFPE-based biomarker to classify clear cell renal cell carcinoma

Abstract

Kidney cancer is newly diagnosed in over 50,000 people each year, and the majority of these tumors are classified as clear cell renal cell carcinoma (ccRCC). Despite being given the same label, the natural history of these tumors can be quite variable, with some patients cured post-nephrectomy and others recurring within a short amount of time. Many different prognostic biomarkers for ccRCC have been proposed, but they do not fully address the underlying molecular or genetic differences leading to the variation in survival. We had previously identified a panel of genes that could distinguish between two molecular classes, ccA and ccB, which were characterized by distinct molecular and genetic changes and had a 6.6 year survival difference. In this chapter, we began development of an assay based on this panel of genes for use on formalin-fixed paraffin-embedded tissue. After careful consideration, we chose to use NanoString Technologies for the quantification of gene expression. A draft FFPE extraction technique was tested and found to be quite reliable for this assay, with a median correlation of 99% between duplicate samples and 98% between replicates. Stable housekeeping genes were calculated from current microarray data and validated by quantitative Real Time PCR (qRT-PCR). A final gene list, with additional ccB, angiogenesis genes, and other biomarker genes was compiled. The resulting custom CodeSet, which we have named ClearCode, was quality tested on RNA from snap-frozen tumors and FFPE lysates for matching tumors. Overall, 96% of the probes were above background for all samples, the correlation for the replicate sample was 97%, and the median correlation between fresh frozen and FFPE samples was 89%. These results suggest that we have a suitable assay to begin to create expression level cutoffs for classifying tumors and ccA and ccB. From here, the biomarker panel can be validated with a group of well annotated tumors to confirm the prognostic value and

potentially assess any possible predictive value. This assay holds a great deal of promise for providing well-needed information to clinicians and patients alike.

Introduction

Kidney cancer affects 1 out of every 67 people in the US[3], making it the 7th leading cause of cancer death in men and the 8th in women[131]. Ninety percent of these cases are renal cell carcinoma, and 60-80% of these are histologically labeled as clear cell renal cell carcinoma. Despite this subtype currently and historically being treated as one disease, the progression is quite variable between patients. At diagnosis, 20% of patients already have metastatic disease and face a 10.6% five-year survival rate[3]. Of the remaining patients, approximately two-thirds will be surgically cured through nephrectomy. This math leaves one-third of the patients to recur. However, at diagnosis, there is currently no confirmed way to distinguish these two outcomes.

Abundant research has gone into trying to determine patient survival and risk of recurrence, as discussed in detail in chapter 1. In 2007, Nogueira and Kim reviewed a list of 63 individual biomarkers, identified to be prognostic through univariate and multivariate analyses[113]. The list has continued to grow since then. Some groups retained the approach of survival-based biomarker selection, but switched to a net-like approach to pin down a prognosis. For example, Zhao et al. identified 259 genes that could be used to predict survival outcome. This focus on survival has considerable advantages, such as helping a patient decide whether it is worthwhile to start on treatments that may decrease their quality of life. However, it does not answer the question of *why* these tumors behave so differently.

We and others tried to answer that question, by attacking the problem from an entirely different angle and seeking to identify any underlying divergence of ccRCC tumor biology that might be creating this clinical variability[39,85,87,88,90,92,95,97]. Once these groups were identified through molecular, genetic or cytogenetic means, then biomarkers for these groups could be determined. Survival data was analyzed where

available, and as might be expected was shown to be significant. However, survival was a secondary question. Underlying tumor biology was the primary goal.

In the last several chapters, we have laid out the identification of two main molecular subtypes of clear cell renal cell carcinoma (ccA and ccB), caused by underlying genetic and possibly epigenetic differences. These changes instigate the expression of different biological pathways, leading to significantly different survival outcomes. Importantly, we identified a panel of 120 probes that could distinguish between these two subtypes.

From this work, our goal is to develop an assay that can be taken into the clinic to differentiate between these two prognostic groups, ccA and ccB, so that clinicians and patients have predicted survival outcome for treatment planning. Therefore in this chapter, we worked towards developing an assay using these 120 probes to classify ccRCC tumors as ccA or ccB, especially using formalin-fixed, paraffin-embedded (FFPE) tumors. Our fixation on the classification of FFPE tissue is because it is the mainstay for preservation and analysis of tumors for the majority of hospitals. It is our ideal goal that a patient having a tumor removed at any hospital could have access to this assay.

To develop this assay, we turned to NanoString Technologies[161]. NanoString CodeSets consist of two probes, the barcode probe and the capture probe, which each have 30-50 base pairs that are complementary to the target mRNA. Both are supplied in excess and mixed simultaneously with sample mRNA. The capture probe is biotin-streptavidin labeled to first allow for purification during removal of excess probes, and second for binding of the complex to the coated surface for measurement. The barcode probe is labeled with one of 5 different flourophores in 7 contiguous locations, and this specific ordering of flourophores is what allows for the identification and quantification of target mRNAs. There are no enzymes or amplifications steps in the process.

We chose this technology over the more common practice of quantitative real time PCR (qPCR) or other available technology, such as High Throughput Genomics, for a number of reasons:

1. qPCR requires the isolation of high quality RNA, a difficult task from FFPE tissue. NanoString is able to provide quantification of more fragmented RNA.
2. With our isolated RNA from FFPE tissue, we had learned that approximately 50ng of starting material was necessary for each replicate of one gene. NanoString recommended 100ng for the quantification of up to 550 genes (now 800 genes).
3. The preparation steps for NanoString analysis is substantially fewer, decreasing risk of degradation or error.
4. There are no reverse transcription, other enzymatic events, or amplification steps. NanoString measures what is present, rather than introducing possible steps of skewing results.
5. To analyze expression of 150 genes by NanoString, it would cost just under $200 per sample.
6. NanoString “barcode” technology means that there is a direct measure of each molecule of RNA, thereby allowing the detection of both very small fold changes and also large ranges of expression differences.

When we began this project, no one on our campus had performed NanoString Analysis. Additionally, NanoString could not provide us with experimental results of using FFPE lysate rather than using RNA isolated from FFPE tissue. Therefore, we needed to begin from scratch, testing the FFPE lysate extraction protocol and assaying the quality of results. Once we affirmed this, we identified the most stable housekeeping genes for our dataset and clarified the biomarker genes that we wanted to use. Upon

receipt of the CodeSet, we again had to perform a quality control test. Only now can we begin the process of determining which genes are best for classifying tumors as ccA or ccB and setting cutoff levels. Overall, NanoString Technologies seems like an excellent choice for the development of our assay to define a ccRCC tumor as ccA or ccB.

This chapter, therefore, will describe the optimization and development of a NanoString CodeSet, which we have named ClearCode, for analysis of FFPE embedded ccRCC samples. In the future, ClearCode will be most helpful in predicting survival outcome patients newly diagnosed with ccRCC. Additionally, given the pathway differences discussed in chapter 3, it is possible that ClearCode may also have value for predicting response to anti-angiogenesis treatments.

Results

Confirmation of extraction technique. Before investing in a custom NanoString CodeSet, it was necessary to confirm that the technique would reliably work on our FFPE samples. Therefore, lysates were prepared from 3 different tumor FFPE blocks: Two tumors, C11 (ccA) and D8 (ccB), were removed in 2008; one tumor, 2 (ccB), was removed in 2001. Slides from all three had been cut six months previous to lysates preparation. All three tumors were extracted using two different lysis buffers, MES and PKD. Lysates in duplicate and a reference RNA were hybridized to a Customer Assay Evaluation CodeSet of 48 commonly assayed genes.

Overall, 97.6% of genes were measured above background, with a range of 93.8%-100% detection per sample, establishing that our extraction technique was efficient using either buffer. The median percent difference of each probe within duplicate samples was 5.0%. The median squared correlation coefficient (r^2) between duplicate samples was 0.99, and r^2 for the 2001 tumor was 0.91, mostly caused by low expressing genes. Comparing the two lysates buffers, r^2 ranged from 0.94-0.99. Interestingly, by hierarchical clustering, the genes in the evaluation CodeSet were already splitting ccA and ccB tumors (Figure 5.1).

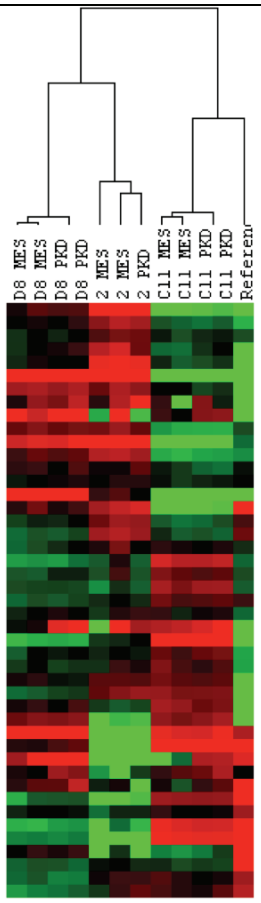

Figure 5.1 Heatmap representation of NanoString test run
NanoString analysis of FFPE lysates shows consistency between duplicates and replicates. The tumors also cluster based on subtype.

The high correlation of technical and sample replicates values demonstrated that NanoString was providing reliable reads from our FFPE lysates.

Identification of housekeepers. Once we determined that NanoString was a viable option for quantifying gene expression from FFPE lysates, we needed to design a custom CodeSet of our biomarkers and housekeeping genes. With regards to housekeeping genes, we wanted to confirm these genes would be stable across our own kidney tumor and normal dataset. Therefore, we retrieved the expression data[97] for the housekeeping genes suggested by NanoString, High Throughput Genomics, and a study by Popovici, et al[4]. Using calculations delineated by de Jonge et al.[162] and Popovici et al., genes were ranked for stability across the tumor and normal tissue arrays. From this data, we chose ABCF1, CD63, C14orf166, RPLP1, and TBP. Additionally, SNRPD2 was chosen due to its extremely low coefficient of variation using de Jonge's technique in our entire dataset of genes.

To confirm these genes would qualify as housekeepers, cDNA from tumors, normals, and a reference were analyzed by real-time PCR using primers for these genes (Figure 5.2). The 18S gene was used as a control, given its history of being a stable housekeeping gene in our hands. Due to multiple peaks in the dissociation curve (data not shown), RPLP1 was immediately removed as an option. The median standard deviation of the remaining genes for the tumors and normal tissue was 1.00 cycles, ranging from 0.67 (TBP) - 1.17 (SNRPD2) cycles. The standard deviation for the 18S probe was 1.22 cycles. Therefore, the genes chosen as housekeepers were deemed to be satisfactory.

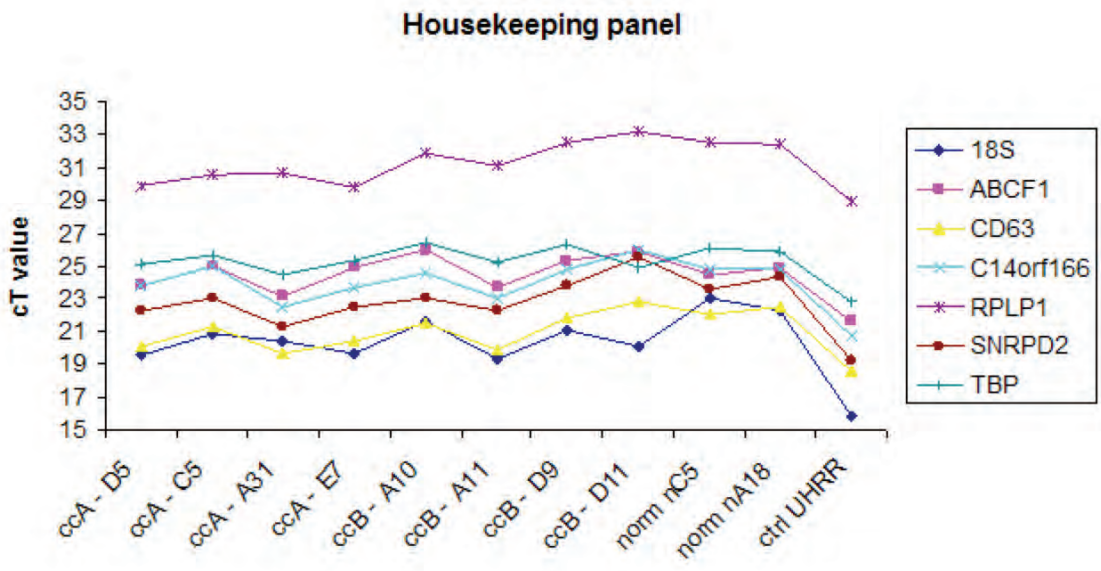

Figure 5.2 cT values of putative housekeeper genes.
Semi-quantitative real time PCR was performed on 4 ccA tumors, 4 ccB tumors, 2 normal tissues and an RNA reference (UHRR). Standard deviations of all test genes except RPLP1 were less than that of 18S.

Finalization of NanoString gene list. By using NanoString's custom CodeSet option, the final gene list (Table 5.1) could be tailored to our exact specifications. Our previously published biomarker panel consisted of 120 probes, corresponding to 110 genes[97]. Of these genes, 86 are overexpressed in ccA tumors and 24 in ccB tumors. Using Agilent's annotation file released in January, 2010, 13 genes had different or missing RefSeq IDs. Nine of these genes (SEPT8, DNCH2, FBI4, FLJ23834, KIAA1043, KIAA1648, PURA, TTLL3, and ZNF292) were either not present or not clear in subtype distinction in the Brooks lab validation data[92] and were, therefore, removed. Three of the remaining 4 genes (FLJ23867, GALNT10, and IMP-2) were overexpressed by ccB tumors, and IMP-2 showed some ability to discriminate between ccA and ccB tumors by semi-quantitative RT-PCR. Therefore, we decided to leave these genes in the first round of NanoString testing. The last gene, FLT1, was definitely able to discriminate between ccA and ccB, so was kept.

To increase the number of genes overexpressed by ccB tumors, we turned to three other analyses: 1) SAM analysis of UNC tumors, 2) SAM analysis of the Brooks group tumors, and 3) genes identified by LAD to distinguish ccA from ccB in grade 2 and 3 Brooks lab tumors. For the SAM analysis, top genes were retrieved both by score and fold change. All genes were rank scored and the rank product was calculated. Six genes ranked highly present in either 4 or 5 categories were chosen – TGB1, SERPINA3, MOXD1, SRPX2, SLC4A3, and FOXM1. Two additional genes from SAM analysis that seemed promising by qRT-PCR were also chosen – GPR87 and LAMB3. With these changes, our final discriminating panel had 79 ccA genes and 30 ccB genes.

In order to compare our biomarker panel to previously published studies, we added 4 other ccRCC prognostic transcript markers. Three (EDNRB, RGS5, and VCAM1) were identified by Yao et al. as indicators of better prognosis in a study with 386 tumors[93] and are overexpressed in ccA tumors. Expression of the fourth, Survivin or BIRC5, correlates with decreased survival[90] and is overexpressed in ccB tumors.

Finally, we wanted to use this panel to gain a greater understanding of the tumors studied. Given the angiogenic and hypoxic molecular phenotype of ccA tumors, we included a panel of genes to confirm this result in a larger set of tumors: ARNT, CDH5, ENG, EPAS1, KDR, NRP1, and VEGFC. We also included Ror2, which is overexpressed in ccB tumors[97] and is associated with increased invasion[154]. With these additions, our final gene panel consisted of 126 genes.

Table 5.1 Custom NanoString CodeSet ClearCode
Final list of genes chosen for the NanoString CodeSet, ClearCode

Category	Genes
House-keepers	ABCF1,C14orf166,CD63,SNRPD2,TBP
ccA[97]	ACAA2,ACADL,ACAT1,ACBD6,ADFP,AFG3L2,ALDH3A2,AQP11,ARSE,B3GNT6, BAT4,BNIP3L,C11orf1,C13orf1,C9orf87,CWF19L2,DREV1,DSCR5,ECHDC3, EHBP1,ESD,FAHD1,FAM44B,FLJ11200,FLJ11588,FLJ13646,FLJ14054,FLJ14146, FLJ14249,FLJ22104,FLT1,FZD1,GALNT4,GHR,GIPC2,HIRIP5,HOXA4,HOXC10, HSPA4L,ITGA6,KCNE3,KIAA0436,LEPROTL1,LOC119710,LOC134147, LOC57146,LOC90624,MAOB,MAP7,MAPT,MGC32124,MGC33887,MRPL21, NETO2,NMT2,NPR3,NUDT14,OSBPL1A,PDGFD,PHYH,PMM1,PRKAA2,PTD012, RAB3IP,RBMX,RDX,RNASE4,SLC1A1,SLC4A1AP,SLC4A4,ST13,STK32B,TCEA3, TCN2,TIGA1,TLR3,TUSC1,YME1L1,ZADH1
ccB[97]	ALDH1A2,AP4B1,B3GALT7,BCL2L12,C5orf19,CDH3,CYB5R2,FLJ23867, GALNT10,IMP-2,KCNK6,KCNN4,MATN4,MGC40405,NCE2,NPM3,SAA4,SLPI, SYTL1,TPM4,UNG2,USP4
ccB	FOXM1,GPR87,LAMB3,MOXD1,SERPINA3,SLC4A3,SRPX2,TGFB1
Other markers	BIRC5,EDNRB,RGS5,VCAM1
Pathway	ARNT,CDH5,ENG,EPAS1,KDR,NRP1,ROR2,VEGFC

Quality control for the custom CodeSet. Before the custom CodeSet ClearCode could be used to set cutoff levels for discerning ccA vs. ccB tumors, quality control was necessary. Twelve samples were run: an RNA reference, a 2008 sample from above (D8) in both lysate buffers as a replicate test, the remaining two samples from above (C11 and 2) in MES buffer, and two additional 2008 tumors (D5 and D11). Overall, 95.6% of genes were measurable above background (91.2%-99.9% per sample). Surprisingly, one gene that was repeatedly at or below background was the housekeeper SNRPD2, and was, therefore, removed for the analysis. There was a 97.1% correlation (r^2) of the D8 replicates in the two buffers (Figure 5.3A). In comparing results of RNA from snap frozen tissue to lysate from FFPE tissue, the median correlation (r^2) was 89.3% (83.9%-91.3% per sample). Again, much of the discrepancy between the two samples was created by low expressing genes (Figure 5.3B). Depending on the performance of these genes in classifying ccA from ccB

tumors, they will likely be removed from future iterations of ClearCode. However, given the vastly different nature of starting materials, these quality control tests results using our custom CodeSet ClearCode suggest that the probes are reliable enough to move forward with creating cutoff values for classifying tumors as ccA vs. ccB.

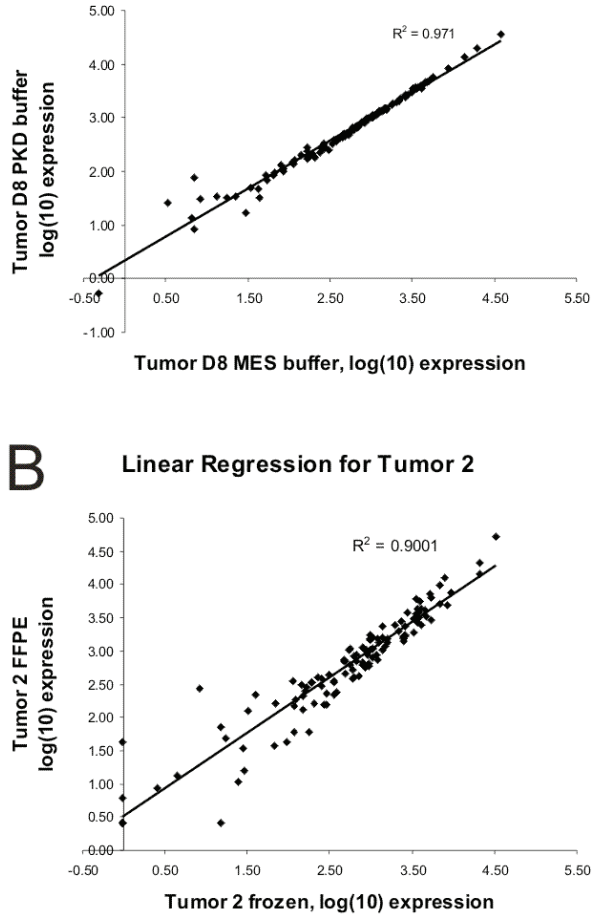

Figure 5.3 Linear regression plots of NanoString data
Plots of tumor D8 extraction replicates (A) and tumor 2 frozen RNA vs. FFPE extract show high correlation values.

Discussion

While microarray data provides enough information to fully classify ccRCC tumors as ccA or ccB, performing gene expression analysis on each new tumor is not time or cost-effective. We had previously identified 120 probes that could discern whether a tumor was ccA or ccB[97]. It is critical that we move forward with these experimental biomarkers for clinical validation. In this chapter, we made great strides towards doing exactly that.

Hospitals most commonly store tissue through formalin-fixation and paraffin-embedding (FFPE), so we wanted our assay to be focused on this type of preserved tissue. However, formalin fixation causes crosslinking between nucleic acids and protein and addition of mono-methylol groups to amino acids which can lead to methylene bridges between amino groups[163]. Additionally, nucleic acids degrade over the time of storage. Therefore, RNA from FFPE tissue is of substantially smaller size and lower quality and does not undergo reverse transcription efficiently.

To overcome these problems, we chose to employ NanoString Technologies as the basis of our assay. NanoString requires a minimal amount of RNA, with lower quality cutoffs. A capture probe and barcode probe directly bind the target mRNAs, eliminating the need for enzymatic or amplification steps. This technique allows for the direct measurement of target transcripts, resulting in measurements of small fold changes as well measurements over a large range of values.

In this chapter, we verified that NanoString's draft FFPE protocol worked well and was reproducible, allowing us to move forward with our plans of creating a custom CodeSet, named ClearCode, based on our biomarker panel. The first step of the CodeSet design was to choose appropriate housekeeping genes. We have found that one of the most common housekeepers, beta-actin, is unreliable in our systems (data

not shown). The housekeeping gene that we most commonly use, 18S ribosomal RNA, is expressed at too high of a copy number to be useful in the NanoString system. Many other common housekeepers are part of the glucose metabolism pathway, a pathway commonly perturbed in kidney cancer due to HIF1 transcriptionally activating several key components, such as PGK and LDHA. Therefore, we had decided to use our gene expression data to calculate the most stable genes across our tumors and normal tissues. Tentative genes were further tested by qPCR, and then employed in NanoString. From this work, we identified ABCF1, CD63, C14orf166, and TBP to be suitable housekeeping genes for kidney cancer research. These genes may be of interest to other groups struggling to find stable housekeeping genes in their research.

We designed ClearCode to include the majority of our previously identified 120 probes that can classify a tumor as ccA or ccB. To this, we added additional genes overexpressed by ccB tumors and genes involved with angiogenesis (a pathway overexpressed by ccA tumors). We also added 4 genes that have previously been shown to be prognostic for ccRCC. When testing 5 different tumors, this gene panel showed an average of 88% correlation between FFPE lysate extracts and RNA from snap-frozen tissue. Given the differences in source material and the relatively low quality of RNA from FFPE tissue, this correlation is more than acceptable.

The next step in creating the ccA-ccB subtyping assay is to determine whether genes are still discriminatory using FFPE tissue instead of snap-frozen tissue, as well as to create cutoff levels for classifying a tumor as ccA or ccB. For this, we will hybridize the rest of the well-defined ccA and ccB tumors (16 and 12, total, respectively) using their FFPE tissue. Using the same tumors that the 120 panel was devised from will give us the best indication as to whether the gene expression will be discriminatory enough in FFPE tissue. Genes that are no longer discriminatory will be removed from future ClearCode syntheses. Additionally, universal reference will be run 2 more times to best

define quality control standards. Once this is done, the reference will only need to be run with every newly synthesized ClearCode batch and for 1-2% of total arrays overall.

Following the creation of expression levels for each gene, ClearCode will be tested on a set of 12 tumors not used in identifying the biomarker panel. Full gene expression analysis will also be run on these 12 tumors and classified using the resulting microarray data. This will validate whether NanoString can properly assign a new tumor to subtype ccA or ccB.

From there, we can turn to a much larger set of FFPE tissue that is clinically well-annotated. We have shown this panel of genes to be prognostic, and patients who have ccA tumors having a median disease-specific survival of 8.6 years versus 2 years with ccB tumors[97]. However, this was analyzed using microarray data from snap-frozen tumors. We will need to show that ClearCode is prognostic on FFPE tissue, and how it compares and/or adds to the clinical data that is currently commonly used to predict risk of recurrence.

Additionally, we purposely included 4 other genes that have been shown to have prognostic value: The overexpression of EDNRB, VCAM1, and RGS5, genes also overexpressed by ccA tumors, correlates with increased survival[93]. In contrast, the overexpression of survivin (BIRC5), a gene overexpressed in ccB tumors, correlates with poor prognosis[90,99]. The expression pattern of these genes in our tumor subtypes mimics the prognostic breakdown of our subtypes, lending increased credence to our observations. Inclusion of these genes within ClearCode will allow us to directly compare these two groups' prognostic models to our own gene panel. Potentially, we may find the greatest prognostic value lies with the combination of several biomarker models and clinical information.

Finally, this gene panel was not chosen for its prognostic value; that result was purely added value. This gene panel was selected as the most able to classify tumors

into their two inherent molecular subtypes, subtypes that are marked by distinct molecular pathways. It is our supposition that these pathway differences may cause the differential response to treatment. This potential predictive value will be tested retrospectively on 30 tumor samples from a trial in which patients were treated with sorafenib for the management of metastatic disease[164]. Dependent on the results, this assay will also be used in a prospective trial here at UNC.

Overall, the NanoString assay of our ccA/ccB biomarkers, ClearCode, is progressing superbly and shows splendid promise for both providing prognostic information for patients and clinicians, as well as possibly helping to guide treatment decisions to improve response.

Materials and Methods

FFPE lysate extraction. Protocol was adapted from NanoString's draft FFPE lysate protocol. Tissue sections from formalin fixed paraffin embedded tumor samples were sliced 5-7 microns onto slides by the UNC Tissue Procurement Facility. All samples were retrieved with appropriate university IRB approval. Total surface area of the tissue section was a minimum of 1 cm^2. Xylene was added to remove paraffin and washed away twice with 100% ethanol, before air-drying to remove any residual ethanol. Pellets were resuspended in either 10mM MES pH 6.5 or PKD buffer (Qiagen). 0.5% SDS and 5ul Proteinase K (20mg/ml) was added to both buffer options. Unless specified, tissues were extracted in MES buffer. These suspensions were incubated at 55ºC, the proteinase K was then inactivated at 80ºC for 15 minutes each. Supernatant from this step was used for hybridization.

NanoString hybridization and data collection. The UNC genomics core processed 5 microliters lysate and 100 micrograms RNA (extraction as previously published) for hybridization against NanoString CodeSets, post-hybridization in the nCounter Prep Station, and data collection with the nCounter Digital Analyzer (NanoString, Seattle, WA). The initial test run used the Customer Assay Evaluation CodeSet nCounter48_C285E. Thereafter, the custom CodeSet Brannon1_C595 was used.

NanoString data analysis. The totals of positive controls for each sample were averaged, and this average was divided by the sample's total to create a spike-in correction factor for each sample. To calculate background, the average of all negative

controls was multiplied by three. This background value was subtracted from the value for each gene, and the result multiplied by the correction factor. Any negative or zero values were changed to 1 for log purposes. Percent present probes were calculated as the total number minus the number of 1's present and divided by the total. Data was then normalized to housekeeping genes: The geometric mean of the housekeeping genes for each sample was calculated, and the geometric mean of all these was acquired. The overall geometric mean was divided by the sample's geometric mean to create a housekeeping correction factor for each sample. The housekeeping correction factor was then multiplied against the previously normalized value. Data was then logged (base 10). Correlation was calculated by linear regression of the data.

Housekeeping gene calculations: Previously published data[154] was analyzed for stable housekeepers. For identification of SNRPD2, the antilog(2) of the data was calculated, and the coefficient of variation (CV=Standard Deviation /average) and maximum fold change (MFC=maximum/minimum) were calculated. For the remaining 5 housekeeping genes, expression data was culled for the suggested housekeepers from NanoString, High Throughput Genomics, and the top 100 of Popovici et al.'s kidney list[4]. SD, CV and MFC were calculated; genes were sorted on each of these variables and ranked accordingly. Duplicate probes with lower SD were removed, to keep worse case scenario. A stability score was calculated according to Popovici et al.: PSS =αLOG(MAX(average-ß,0),2)-stdev, where α is a coefficient to control mean expression vs. SD (we set it to 0.25 as the paper did) and ß is the mean expression cutoff, which we set to the 25th percentile following the paper or -0.3908466934. The PSS was sorted and ranked, with genes that had calculation errors due to negative values being given a rank of 87. The rank product was then calculated as RP=product(ranks)^(1/(n)), where n

is the number of ranks available, and RP was sorted. The top 2 overall (C14orf166) and top 2 NanoString probes (TBP and ABCF1) were chosen. Additionally, CD63 was chosen for having the highest mean value.

Semi-quantitative real time PCR. Tumor cDNA was as previously published[97], and normal tissue cDNA was made using the same protocol. The UHRR reference is Stratagene Universal Human RNA Reference (San Diego, CA). Five nanograms of tumor cDNA and ten nanograms of UHRR cDNA were used per reaction amplified using Absolute SYBR Green ROX mix (Thermo Scientific, Epsom, Surrey, UK) on the Applied Biosystems ABI 7900HT Sequence Detection System (Carlsbad, CA). 18S rRNA primers (Applied Biosystems) were used as a control. Primers were designed using IDT (http://www.idtdna.com/): ABCF1 (CGCCAAGCCATGTTAGAAAATG and TGCCATGAGCGGAGATGCTGAA), C14orf166 (TCGGATTTTGGTTCAGGAGC and TGTCTAAAGCAACAGGTAAGCC), CD63 (AACGAGAAGGCGATCCATAAG and ACAAAAGCAATTCCAAGGGC), RPLP1 (ATCTGCAATGTAGGGGCC and GCTTCCACTTTCTTCTCCTCAG), SNRPD2 (AATAAGAAACTCCTGGGCCG and CTCAGTCCACATCTCCTTCAC), and TBP (CCCGAAACGCCGAATATAATC and GCACACCATTTTCCCAGAAC).

Chapter Six:

Conclusions and Discussions

Overall summary

Renal cell carcinoma is disease with variable natural history and poor treatment options. In the previous chapters, we have demonstrated that there are two primary and inherent molecular subtypes of clear cell renal cell carcinoma, which we have named ccA and ccB. These two subtypes have significant survival differences, where patients with ccA tumors have a median disease-specific survival of 8.6 years vs. 2 years for patients with ccB tumors ($p=0.002$). By both univariate and multivariate analysis with common clinical measures, this classification is significantly associated with survival. This division therefore allows us to explore what underlying molecular and genetic differences are causing the clinical heterogeneity, as well as take the prognostic breakdown into the clinic.

As we began to examine the ccA and ccB subtypes molecularly and genetically, we saw vast differences between them. Using the gene expression data, ccA tumors were shown to overexpress genes in the angiogenesis and hypoxia pathways. Seeing as these are the classic phenotypes associated with ccRCC, it is particularly striking that one subclass should express these pathways more highly. It also suggests that these tumors may be more responsive than ccB tumors to anti-angiogenic agents, one of the main classes of molecularly targeted treatments for ccRCC. In comparison, ccB tumors overexpress genes related to cell cycle, cell differentiation, TGF-Beta, Wnt targets, epithelial to mesenchymal transition, and response to wounding. These tumors have clearly undergone additional or different molecular changes to become a far more aggressive tumor. Interestingly, ccB tumors underexpress metabolism and glycolysis genes compared to both ccA tumors and normal tissue. This result is highly unusual as many tumor types, especially ccRCC, are known for being highly glycolytic, thus allowing

for the use of a radioactive glucose mimetic to be used in imaging (FDG-PET) for the diagnosis and monitoring of tumors.

Genetically, ccA and ccB subtypes continue to show their immense differences. More ccB tumors have deletions on chromosomes 9 and 14, changes previously shown to be associated with poor survival. These regions may harbor key tumor suppressor genes and will need to be further explored. Overall, these results support ccB tumors being more aggressive tumors, fitting with the decreased survival outcome seen above. Intriguingly, ccA tumors contain mutations in a number of different histone modification genes, suggesting that there are also epigenetic differences between the two subtypes.

Finally, we have delineated a biomarker panel that can discriminate between ccA and ccB tumors based on microarray data. These genes, along with additional genes of interest for these subtypes and biomarkers previously identified by other groups, are the basis for a new assay, called ClearCode, for FFPE tissue using NanoString Technologies. We have shown that this technology produces reproducible results using extracts from our FFPE samples and acceptable correlation between FFPE and fresh frozen samples. We are now ready to move forward into delineating expression levels necessary to classify a tumor as ccA or ccB. This critical advance opens the door to prospective tumor assignment and translation of this technology to clinical use.

This research helps to explain why certain patients progress, while others are completely cured by nephrectomy. The survival data suggests that patients with indolent disease have ccA tumors, while those that recur have ccB tumors. The pathway analyses above suggest that ccB tumors tend to be more invasive in nature, while the main hallmark of ccA tumors is angiogenesis and hypoxia. Further research into this underlying tumor biology differences between these two subtypes may provide insight for better treatment options. Additionally, once completed and validated, the FFPE assay should supply important prognostic predictions for the clinician and patient.

Comparison to previous work

One of the most important aspects of research is whether it is novel and important, and our studies definitely meet those criteria. Previous work on the topic falls into 4 main camps (Table 1.1):

1) Early groups focused primarily on ccRCC versus normal tissue and the molecular changes between the two, both to serve as diagnostic markers and to understand disease development[73-78].

Our work in chapter 3 does analyze tumors with respect to normal tissue, but we do this to get a better understanding of the differences between ccA and ccB tumors. This analysis is what helped us to understand that ccA tumors do not overexpress metabolism genes compared to normal as we had thought from the ccA vs. ccB analysis. Rather, ccB tumors underexpress these genes and pathways.

2) A number of early groups also studied all types of RCC, looking for genetic or molecular markers in order to perform "molecular histology" [6,80-85,89,91].

These studies examined the molecular differences between clear cell, papillary, chromophobe, and oncocytoma histologies, providing extra insight into the tumor biology of these prognostically different tumors. Additionally, these histologic subtypes appear identical by radiologic examination, so the identification of diagnostic biomarkers will be important as core, and even fine-needle, biopsies become more common.

When enough clear cell tumors were present in the mix, it generally became apparent that there were possibly two clusters of clear cell. Though, questions were

then raised whether this grouping was being created by either the presence of mixed tumors or the selection of genes which are most variable across all tumor types.

Interestingly, however, we discovered that a papillary tumor mislabeled as clear cell had been a ccB core tumor, clustering tightly with the other ccB tumors. When we do include papillary tumors in with our core tumors, ccB tumors do cluster closely with the papillary tumors (Figure 6.1). Some papillary tumors even cluster with the ccB tumors, depending on the tumors present. This result does present the question as to whether our ccB tumors represent a transition state tumor or are more similar in some respects to papillary RCC, which could potentially be important information for therapy decisions. Given that ccB tumors are the poor prognostic group and that papillary tumors tend to be more aggressive, it is also possible that ccB and papillary tumors underwent parallel evolution (or devolution, which may be a more apropos term given the dedifferentiation that occurs in tumors) to express similar genes and pathways to survive in an environment that is meant to be hostile to proliferating cells. If so, perhaps both ccB tumors and papillary tumors could respond to the same drug therapies. This line of questioning is an area of research that could provide some interesting answers.

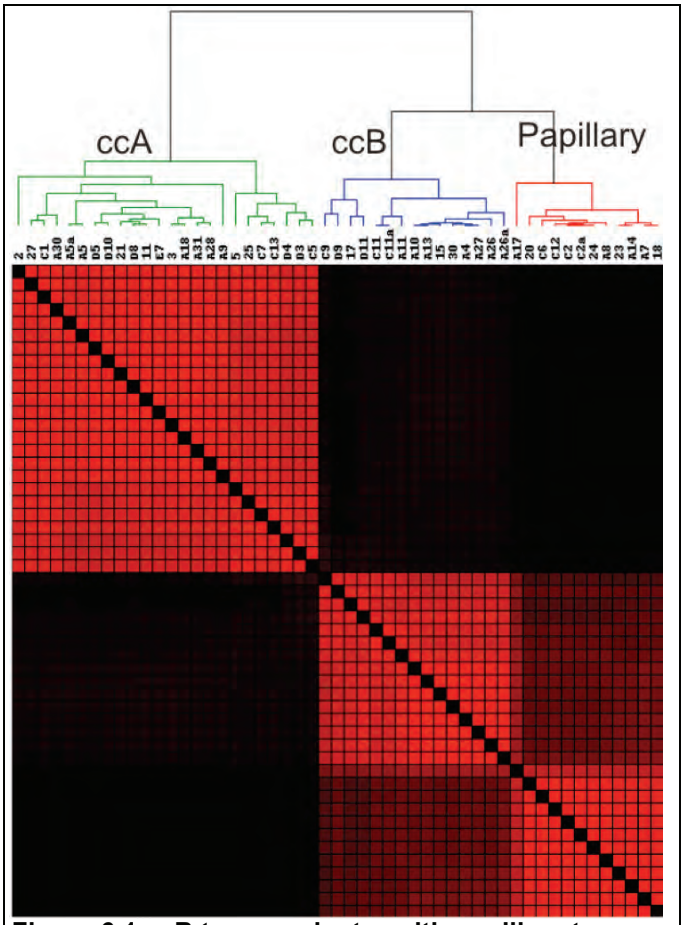

Figure 6.1 ccB tumors cluster with papillary tumors Consensus matrix of ccB and papillary tumors clustering together apart from ccA tumors.

Clearly, we do believe that these studies were showing correctly showing the presence of two groups of clear cell. However, this result was not the focus of the above studies, and it was not further researched.

3) A number of studies ended up being more clinically driven, regardless of whether they began with an unsupervised analysis[87,88,90,92].

Of these, Takahashi et al. was the earliest study and did see that tumors tended to cluster unsupervised into two groups[87]. These groups were based predominantly upon 5 year disease-specific survival, and genes that could differentiate based on survival then became the focus. Unfortunately, no work was done to understand the underlying pathway changes causing this clustering and prognostic result. Vasselli et al. followed a similar strategy, finding two natural clusters with survival differences but then choosing genes entirely based on survival[88]. Kosari et al. also saw two clusters, predominantly broken down into aggressive (patients died of disease or developed metastases in less than 4 years post-nephrectomy) and non-aggressive tumors, and then chose genes to discriminate between aggressive vs non-aggressive tumors.

Zhao et al. analyzed the largest number of tumors, 177 in all[92], and we used their data for validation and prognostic information. In their data, they found 5 different clusters, within 2 main clusters, with significant survival differences. Additionally, they saw overexpression of angiogenesis and metabolism genes in one of the main subsets. Again, rather than building a biomarker panel based on these inherent groups, they chose genes based purely on survival outcome.

When we originally analyzed the Zhao et al. data using ConsensusCluster, we saw four (and possibly a fifth) different clusters immediately apparent (Figure 6.2). However, the 177 tumors were arrayed on 10 different print batches of chips. We combined these batches using DWD and found that only two clusters were present in the adjusted data. This result helped confirm our data that two subtypes of ccRCC dominate. Although it still remains possible that with larger sample studies, additional heterogeneity will emerge within ccA and ccB or that the unclassified tumors will emerge as their own class.

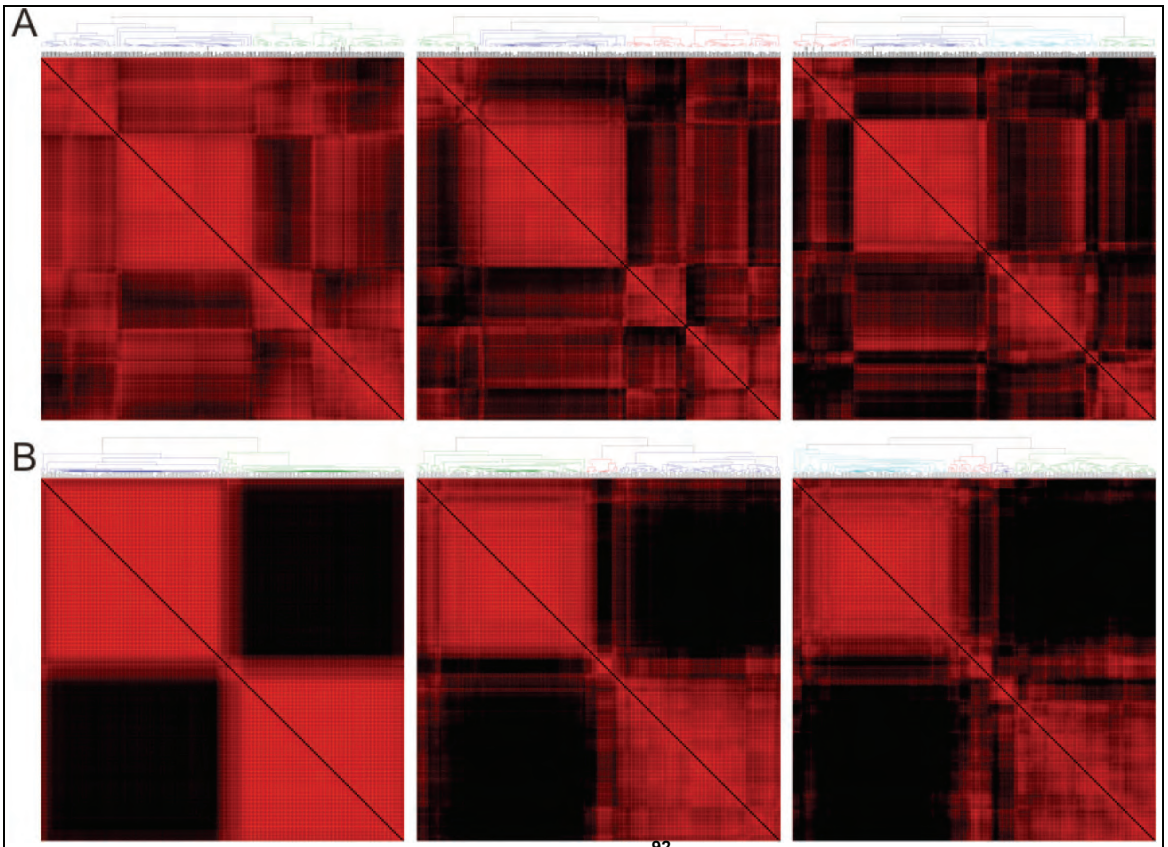

Figure 6.2 Effect of DWD adjustment on Zhao et al.[92] data
Data from Zhao et al. was analyzed by ConsensusCluster without (A) and with (B) DWD adjustment to control for batch effect caused by 10 print runs.

These groups had a goal to find genes associated with survival and recurrence, whether it was a time frame of 4 year, 5 year or just continuous. The gene sets have the

potential to become useful in the clinic, helping a patient to better gauge how much time they have remaining. However, all remain to be prospectively validated.

In contrast, when we began, we certainly had hopes of prognostic import (especially given the results shown above). That was not our overriding goal, however. We wanted to determine whether our tumors also created two (or more) clusters, determine if they correlate with patient survival, and then fully explore those inherent groups. This information has the power to really help explain *why* tumors behave so differently. It might also provide insight into which patients will respond to specific treatments, and/or provide new avenues for drug development. These were our goals.

4) The final group of studies did approach ccRCC as we did – focused on the underlying biology[39,95,96].

The first study, by Skubitz et al., was like several other studies mentioned above and predominantly dismissed due to small sample sizes. They looked at 16 tumors ccRCC tumors, and found 2 subtypes, ccRCC-A and ccRCC-B[95]. Three of their ccRCC-B samples had sarcomatoid features, suggesting a far more aggressive disease trajectory for the entire group. Interestingly, 2 of their ccRCC-A markers (GIPC2 and MAP7) and 1 of their ccRCC-B markers are found in our groups as well (SLPI). Similar to ours, the ccRCC-A tumors overexpressed metabolism genes (compared to ccRCC-B) tumors, while ccRCC-B tumors overexpressed genes related to the extracellular matrix. Our study tripled the number of tumors and did not involve any sarcomatoid tumors, in order to avoid skewing the results as this histologic feature is known to portend a dismal prognosis. (Though, when working with the Futreal gene expression data as normalized by them, we did find that the sarcomatoid tumors were labeled as ccB tumors (data not shown).) Additionally, our studies were able to provide more molecular and genetic information discriminating the two subtypes, as well as prognostic information.

Gordan et al. also concentrated on the underlying biology of ccRCC tumors, but not by studying the inherent clusters present in the data. Instead, they focused on the primary pathway dysregulated ccRCC, pVHL inactivation leading to HIF overexpression[39]. One of the most important results from this study is the *in vivo* confirmation that HIF1 and HIF2 expressing tumors are molecularly distinct from HIF2 only expressing tumors and that HIF2 only tumors overexpress c-Myc, leading to increased proliferation. The relation of this study to ours will be discussed in the next section.

Finally, the Zhao et al. group later attempted to find biological significance to their survival gene set. Not surprisingly, they found that tumors from patients who survived longer were more like normal tissue, while those from poor survival patients exhibited a wound-healing signature[96]. We have also seen that ccB tumors, which have a decreased survival outcome, overexpress genes related to wound healing.

HIF expression versus ccA/ccB?

As discussed above, *VHL* inactivation is found in the overwhelming majority of ccRCC tumors, leading to the overexpression of either HIF1 and HIF2 (H1H2 tumors) or just HIF2 (H2 tumors). Gordan et al. showed that this delineation also leads to a change in C-myc activity and increased cellular proliferation in H2 tumors, while H1H2 tumors show increased expression of glycolysis genes. These expression profile differences strongly mimic what we see in our ccA/ccB breakdown, constantly raising the question of whether ccA tumors are H1H2 and ccB tumors are H2 only. Yet, as seen in Chapter 3, this does not seem to be the case. We found that there were H1H2 and H2 only tumors in both ccA and ccB, using both our dataset and the Gordan et al. set. One possibility is that there are minute differences in expression of these proteins that we are unable to

detect, but are still enough to be causing these pathway outcomes. Emerging evidence also points to the possibility that microdeletions of splice variants occur in HIF biology, which would be undectable by IHC. Looking at C-Myc for example, given that HIF-1 and HIF-2 act as counterbalances for Myc expression, a slight increase or decrease in HIF-1 would shift that expression, so the possibility remains that some more subtle HIF influence contributes to ccA/ccB physiology.

How *VHL* is inactivated can impact the expression levels of HIF and therefore the growth of the tumor[37,38]. When we looked at *VHL* in our tumors, we saw the presence of both mutation and methylation in both subtypes. More tumors in ccA than in ccB contained alterations to *VHL*'s genetic code though, which could possibly be causing the shift. Not much can be drawn from these results in the current environment, however, for four reasons: One, we did not perform laser-capture microdissection or even razor sectioning to make sure that the sample was >90% tumor. PCR should still have been able to detect the presence of mutations, even without a pure sample, though. Two, we saw discrepancies in mutational analyses from two groups that we used. As Dr. Kate Nathanson is well-respected in the field for the quality of her analyses, we went with her results. However, the differences in the results raise the possibility of regional differences within the tumor. VHL mutations are generally an early step in ccRCC tumor development, but there is a slight chance that certain regions of the tumor may harbor differ mutations. Three, our sample size is extremely limited. In order to answer the question of the *VHL*'s and HIF's role in the ccA/ccB subtypes, far more samples need to be critically analyzed. Additionally, it would be important to analyze the type of *VHL* inactivation, to see whether one subtype is more likely to have different types of inactivating mutations or whether mutations are spatially located in different regions. Fourth, the molecular impact of most VHL missense mutations on HIF or other biology is unknown.

HIF expression levels are generally accepted to only be modulated post-translationally by pVHL. However, one group has shown that HIF-2α mRNA (EPAS1) expression inversely correlates with both TNM stage and nuclear grade[165]. Another group identified it as one of only 35 transcripts that were necessary to discriminate between non-aggressive and aggressive tumors, and EPAS1 was overexpressed in the non-aggressive tumors[90]. In our tumors, the Brooks tumors, and the Futreal tumors, EPAS1 (HIF-2α) is overexpressed in the ccA tumors as compared to ccB tumors, correlating HIF-2α mRNA expression with better survival. What is leading to this difference in expression level is unknown, as well as how exactly this difference is impacting HIF2 protein expression or the tumor. These questions are fodder for future experiments.

Third HIF's the charm?

One more player in the *VHL*/HIF pathway that generally gets dismissed is HIF-3. HIF-3α splice variants 1-3 may also transcriptionally activate specific genes, but they may not be as efficient as HIF1 and HIF2 given the lack of a C-terminal transactivation domain. Instead, HIF-3α1 contain an LZIP domain, which functions in DNA binding, and all 3 contain LXXL domains, which promote protein-protein interactions[22,166,167]. HIF-3α1-2, however, do seem to inhibit effective transcriptional regulation of genes by HIF-1α, likely due to competitive binding of HIF-1β, and this inhibition is particularly the case when HIF-1β is limiting [167,168]. HIF-3α4 acts to dominantly negatively regulate HIF-1 and HIF-2 by interacting with both alpha subunits and HIF-1β, but tends to be downregulated in ccRCC[22,31,32] There are also up to 6 other splice variants with unknown roles in the cell, and even the existence of variants 3 and 5 are disputed[168].

Perhaps we really should not be ignoring HIF3, however. As described above, our ccA and ccB tumors stratify in a way that suggests that HIF1 is underexpressed in ccB tumors, yet we don't see that change in HIF1 by IHC. However, looking at gene expression data, we do see that ccB tumors overexpress HIF3 (compared to ccA) and underexpress ARNT/HIF-1β (compared to normal tissue). It is unclear from the microarray data which splice variant of HIF-3α is being overexpressed; however, given Maynard's downregulation results, it is unlikely to be HIF-3α4. Yet, two other splice variants have been shown to inhibit HIF-1 activity, particularly when HIF-1β is decreased, which is true in ccB tumors. Additionally, since this inhibition occurs through direct binding of HIF-1 and competitive binding of HIF-1β, there would be no resulting decrease in HIF-1 protein expression. While pure speculation, this hypothesis (Figure 6.3) seems like a plausible way of explaining why ccA and ccB do not stratify based on HIF1 and HIF2 expression, as one would expect. HIF-3 expression and interactions should really be examined in these two tumor subtypes.

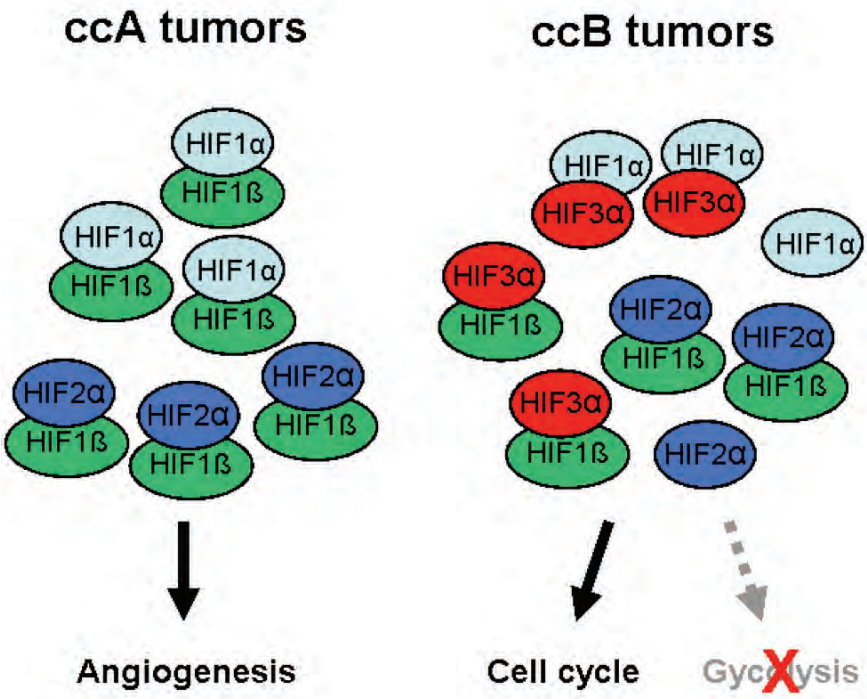

Figure 6.3 Model of HIF protein interactions in ccA and ccB tumors
The pathways expressed by these two tumor subtypes are extremely reminiscent of those described by Gordan et al. as caused by HIF1 and HIF2 expression differences. However, both HIF1 and HIF2 protein are expressed in both ccA and ccB tumors as seen by IHC. By gene expression, HIF3 is upregulated in ccB tumors while HIF-1β is decreased. The shift in these two proteins could drive ccB tumors to express a more HIF2 centric expression pattern.

The Ror2 of the wild ccB

One more interesting gene that is dysregulated between these subtypes is the tyrosine kinase Ror2, which is overexpressed in ccB tumors. Our lab recently discovered that Ror2 is a tumor intrinsic kinase for ccRCC, and we showed by microarray data that its expression correlates with extracellular matrix remodeling proteins. Additionally, expression of Ror2 protein correlates with cellular migration, anchorage independent growth, and tumor growth in xenografts[154]. Since then, Ror2 expression has been shown to correlate with migration, invasion, and/or metastases in

melanoma, osteosarcoma, squamous cell carcinoma, gastric cell carcinoma, and prostate cancer[169-175].

The overexpression of Ror2 in ccB tumors makes perfect sense given that Ror2 expression correlates with an invasive phenotype and ccB is the poor prognosis group, which would indicate recurrence/ metastases of ccRCC. However, we have been unable to directly correlate Ror2 expression with any clinical data. Therefore, Ror2 was included in the NanoString custom codeset. The resulting data will answer two questions: 1) Does Ror2 directly correlate with clinical measures of survival, namely with decreased survival or increased stage and grade? 2) Is Ror2 a good marker for ccB? Given the above mentioned research, we anticipate that the answer will be yes to both questions. Since Ror2 is tumor intrinsic kinase in ccRCC, another member of the lab is working on a collaboration to identify potential drugs to target Ror2, and therefore ccB tumors.

Deep felt losses

In chapter 4, we started to explore the underlying genetic differences between ccA and ccB tumors. Several regions stood out as having distinct changes in copy number between the two subtypes. In particular, ccB tumors seem to have more deletions of regions on chromosome 9 and 14. These regions are of great interest, because they have been previously shown to associated with decreased disease-specific survival[6,7,47,53,60,66,67,176-178]. Loss of 14/14q has not always generally retained independent prognostic significance, but loss of 9p has. These previous studies confirm that we are on the correct track with ccB being a more aggressive and deadly form of ccRCC.

We would still like to know how and why these regions are associated with poor survival, however. The next steps will require us to first confirm that these regions are statistically different between the two subtypes, then start looking within these regions for genes that seem to be in peaks of highest deletion. Copy number analysis will also be combined with gene expression data in order to correlate results more strongly.

The most obvious tumor suppressor on chromosome 9p is *CDKN2A*, the gene that encodes p16INK4A/ARF and functions to inhibit the cycle protein CDK4 and stabilize p53. Beroukhim et al. find that this gene is located at the point of highest deletion and lowest expression in their data, 40% of which are from patients with VHL disease[65]. However, our gene expression data shows that there is no difference in *CDKN2A* transcript levels between the subtypes and, in fact, is overexpressed in both subtypes compared to normal. Another tumor suppressor located nearby and identified by Beroukhim et al. through copy number peaks (but not expression data) is *CDKN2B*. The resulting protein, p15, also inhibits function of CDK4 or CDK6, preventing activation of the cell cycle by cyclin D1. The transcript of this protein is overexpressed in ccA compared to normal, but its expression is not altered in ccB tumors. The role, or rather lack thereof, of this protein in ccB tumors should be further investigated.

Located on chromosome 14 is HIF1, which would fit our expressed pathways, but as discussed above, there was a lack of observable differences in protein expression by IHC. Looking for other targets, through gene expression data, Beroukhim et al. suggest NRXN3, which encodes neurexin 3, a gene that functions as a cell adhesion molecule and receptor. However, this gene is not dysregulated in our microarray data. Overall, we will need to look more closely for which genes are dysregulated in this region.

One more region of deletion in ccB tumors is in chromosome 1p. This region was also marked as a distinguishing copy number alteration between the two clusters of tumors in the study by Arai et al. and was specifically deleted in the poor prognostic

group[63]. By both peak and expression analysis, Beroukhim et al. identified RUNX3 as being a potential target of this deletion. RUNX3 is a transcription factor that promotes transcription of p21, another cell cycle inhibitor[179,180]. p21 specifically functions by inhibiting the activity of the cyclin E/CDK2 and the cyclin D/CDK4 complexes. While this is an extremely appealing idea for a target, RUNX3 is overexpressed in our data in both subtypes compared to normal. Other targets will need to be explored.

Only on the surface

The copy number changes described above and earlier are certainly causing some of the pathway differences between ccA and ccB tumors. However, as was apparent from the data shown in chapter 4, there are regions of differential expression that do not correlate with amplification or deletion. This result, particularly in combination with the large number of mutations in histone modification genes, suggests that epigenetic modifications play a large role in the differences between the subtypes as well. The Futreal group had noticed that the majority (88%) of the tumors with *SETD2* or *JARID1C* mutations also contained a mutation in *VHL* or exhibited a hypoxia phenotype[40]. Similarly, the better prognostic group identified by Arai et al. showed a decreased number of methylated CpG islands[63]. These data correlate well with our better surviving ccA tumors, which contain the majority of the histone modification gene mutations and overexpress pathways related to hypoxia. Analysis of the two subtypes by methylation arrays and/or ChIP-chip assays may provide additional answers to the differences between ccA and ccB tumors.

An interesting related question is whether genes that are similarly overexpressed or underexpressed by both subtypes are regulated in different ways. I.e., a gene may be

deleted in one subtype but may be epigenetically modified for repression in the other subtype.

The problem in the pathways

We have presented a lot of pathway results in these chapters, pathways that distinguish these two subtypes and help bring understanding to why ccB tumors have such a decreased survival outcome. However, with the exception of the Ror2 data mentioned above, all of these pathways remain constructs of gene expression data. This limited analysis does not cause the results to be untrue, especially given that they have been validated in other datasets. Rather, it just makes the data feel one-sided and not fully substantiated.

Ideally, ccA cell lines and ccB cell lines would be analyzed and pathways validated by means of perturbation by shRNA, known inhibitors, or overexpression. Western blots, quantitative Real Time PCR, wound healing assays, foci formation assays, soft agar assays, would shortly follow. These original and altered cell lines could even be placed into nude mice to determine whether tumors were less likely to form under certain modifications. However, ccRCC cell lines are few in number, and attempts to classify these lines by subtype have failed. Individual genes or pathways, such as was done in Ror2, are still feasible, but for now, full analysis of subtypes via current cell lines is not feasible.

Three main options, therefore, present themselves:

First and most obvious, tumors that have already been solidly identified as ccA or ccB could be analyzed by immunohistochemistry (IHC), metabolomics, or proteomics. The main problem with this means of addressing the question is limited sample. In fact,

many of the core tumors have no more available sample. New tumors can always be classified, but sample will always be in limited quantity.

Second, fresh tumor tissue could be retrieved, and cell lines could be made via growth on plastic tissue culture plates. If successful, both the original tissue and newly derived cell lines could be arrayed to confirm subtype and whether dramatic shifts had been made due to culturing techniques that would prevent analysis of key pathways. These new and classified cell lines would provide the most flexible means of exploring the pathways inherent to each subtype. However, tissue culture dishes fail to mimic tissue microenvironment and certain details may not be possible to ascertain.

The final, and most immediately feasible, option presents itself in collaboration with Dr. William Kim. Dr. Kim has generated xenograft lines from ccRCC tumors. Original tumor tissue is available for arraying and subtyping. While not as replenishable as traditional cell lines, this technique would most closely mimic the environment within the human body. An immediate question that could be answered is which subtype is most sustainable upon transplantation. Which subtype responds to various treatments is the question on the forefront of many minds, and this system might provide strong clues to that answer. Why they respond to these specific treatments might be also answerable. IHC of fixed tissue, as well as RNA and protein extraction from fresh tissue could confirm the pathways identified by gene expression analysis. These pathways could even be perturbed via drug treatments or lentiviral shRNA vectors, possibly providing clues as to which pathways are most important in delineating these two subtypes. Overall, although there are difficulties in maintaining such a system, there are tremendous benefits as well. Additionally, of the three options, this one provides an immediate avenue for dissection of the pathways implicated in ccA and ccB tumors.

That's a nice assay

With our understanding that clear cell RCC made up of two distinct groups and having a list of genes to distinguish between these two groups, we have begun the process of creating an assay to easily subtype new tumors. We have chosen to design the assay around formaldehyde-fixed paraffin embedded (FFPE) tissue, as that is the most common means of tissue preservation and analysis. We have also chosen to work with NanoString Technologies, for a variety of reasons, but foremost because this technology allows the use of 100ng (or less) of fragmented RNA to analyze up to 800 genes at a time. We have verified that extraction of RNA from FFPE is unnecessary, as NanoString provides consistent results from FFPE lysates. Currently, this process has shown that the majority of our probes produce reproducible results and correlate reasonably well between snap-frozen RNA and FFPE lysates. Therefore, we are in the process of delineating expression level cutoffs for the purpose of classifying a tumor as ccA or ccB. We will then test 12 tumors not used in the selection of our biomarker gene panel by both microarray and NanoString to verify that this assay can correctly subtype unknown tumors.

Progressive, bifocal, or an entire second set

An obvious question that arises as we examine the ccA and ccB subtypes is whether ccB tumors are just ccA tumors that have progressed further, ccA and ccB tumors are just another two aspects of the single ccRCC disease, or if they are two completely different diseases. It is easy to envision ccB tumors as more advanced ccA tumors, with the unclassified tumors being a transitional state. If this were true, one would expect that all patients diagnosed early (at a low stage and grade) would have

ccA tumors and never recur. To some extent, this is true, as having a low stage and grade tumor confers a lower risk of recurrence. However, patients with ccB tumors tend to be diagnosed younger than those with ccA tumors (median of 63 vs. 70 years old, $p<0.01$). Therefore, while progression is an attractive idea, it seems somewhat unlikely.

Throughout this dissertation, we've been treating ccA and ccB as subclasses of the overarching clear cell umbrella. Histologically speaking, this would be true. Both look identical to a pathologist. Based on the Futreal data, both also share 3p deletions, as well as global copy number pattern. Given these definitions, one could define them as being in the same species of ccRCC, but perhaps different genera.

However, it is tempting to entertain the thought that these ccA and ccB tumors are two different diseases. Chromophobe and oncocytoma look similar to pathologists and even cluster together compared to clear cell and papillary tumors based on gene expression[83,84,89]; nonetheless, they are regarded as two different diseases. It is possible that ccA and ccB tumors did both begin their tumorigenesis path through deletion of 3p and/or inactivation of *VHL,* but then diverged shortly thereafter into two different species. After all, the differences between ccA and ccB tumors are almost as large as those between ccA tumors and normal tissue (6213 vs 9112 probes differentially expressed). One might argue that this question is a matter of semantics; however, it could affect how seriously clinicians and researchers address the differences between ccA and ccB tumors.

Two for one deal

This entire dissertation has focused on the division of ccRCC into the two subtypes, which is very important. However, perhaps in fully understanding these two subtypes, we can also better understand clear cell as a single disease as well. At the

same time that cancers get further and further subdivided, research is also turning away from such divisions and trying to find out what rules them all. Specifically, the question being asked is which common, actionable molecular or genetic changes are inherent in the cancers. When looking at a set of all clear cell tumors compared to normal, we are ignoring the differences between ccA and ccB. Throughout the last chapters, it should have become clear that these differences are important and may be of great boon to the clinic. However, in looking at each subtype compared to normals, we can also fully see the genes and pathways that are similarly overexpressed and underexpressed. If this is instead done with all tumors against normal, there's no way to know whether one subtype is skewing the results. When the results are skewed, we are left again with an understanding that really only applies to a select group of tumors. Instead, by looking at the similarities between the groups, we can decrease noise and feel fully confident that we have an understanding of what is true to clear cell as a single disease. This approach is an area that bears follow-up, and mayhap, personalized medicine in the not so distant future need not be quite so personalized.

In conclusion….

This research presented within this dissertation has shown that we can take one disease and break it into two. Many studies have mentioned the presence of two clusters/classes/groups of ccRCC tumors, whether by molecular, genetic, or cytogenetic means. Particularly in this chapter, we have shown how our work supports and fleshes out these previous results. It is important to consider that all these disparate laboratories are likely discussing the same two subtypes. With this understanding, the knowledge garnered from the different studies could be combined to lead to a powerful episteme

that could strongly influence clinical decision making and drug development for the better.

The two groups of clear cell renal cell carcinoma, ccA and ccB, are almost as different from each other as clear cell is from normal tissue. ccA tumors follow the classic RCC pattern and express an angiogenic and hypoxic molecular signature. ccB tumors are the darker side of the disease, showing genetic losses common in poor prognostic groups and molecular pathways of proliferation, wound healing and epithelial to mesenchymal transition. Inasmuch, patients with ccA tumors show a median survival of 8.6 years to the 2 years seen for ccB patients. Essentially, as a friend says, ccA tumor good, ccB tumor bad. We will soon have an assay to distinguish between these two, to provide more information for researchers, clinicians, and patients. Then, clear cell renal cell carcinoma may become a little clearer.

References

1. Brannon, A.R. & Rathmell, W.K. Renal Cell Carcinoma: Where Will the State-of-the-Art Lead Us? *Current Oncology Reports* **12**(2010).

2. Ferlay, J. et al. GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet]. (International Agency for Research on Cancer, Lyon, France, 2010).

3. Altekruse SF, K.C., Krapcho M, Neyman N, Aminou R, Waldron W, Ruhl J, Howlader N, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Cronin K, Chen HS, Feuer EJ, Stinchcomb DG, Edwards BK (eds). *SEER Cancer Statistics Review, 1975-2007, http://www.seer.cancer.gov/*, (National Cancer Institute, Bethesda, MD, 2010).

4. Popovici, V. et al. Selecting control genes for RT-QPCR using public microarray data. *BMC Bioinformatics* **10**, 42 (2009).

5. Waalkes, S. et al. Obesity is associated with improved survival in patients with organ-confined clear-cell kidney cancer. *Cancer Causes Control* (2010).

6. Furge, K.A. et al. Robust classification of renal cell carcinoma based on gene expression data and predicted cytogenetic profiles. *Cancer Res* **64**, 4117-21 (2004).

7. Brunelli, M. et al. Loss of chromosome 9p is an independent prognostic factor in patients with clear cell renal cell carcinoma. *Mod Pathol* **21**, 1-6 (2008).

8. Zisman, A. et al. Improved prognostication of renal cell carcinoma using an integrated staging system. *J Clin Oncol* **19**, 1649-57 (2001).

9. Kattan, M.W., Reuter, V., Motzer, R.J., Katz, J. & Russo, P. A postoperative prognostic nomogram for renal cell carcinoma. *J Urol* **166**, 63-7 (2001).

10. Frank, I. et al. An outcome prediction model for patients with clear cell renal cell carcinoma treated with radical nephrectomy based on tumor stage, size, grade and necrosis: the SSIGN score. *J Urol* **168**, 2395-400 (2002).

11. Sorbellini, M. et al. A postoperative prognostic nomogram predicting recurrence for patients with conventional clear cell renal cell carcinoma. *J Urol* **173**, 48-51 (2005).

12. Motzer, R.J. et al. Survival and prognostic stratification of 670 patients with advanced renal cell carcinoma. *J Clin Oncol* **17**, 2530-40 (1999).

13. Mekhail, T.M. et al. Validation and extension of the Memorial Sloan-Kettering prognostic factors model for survival in patients with previously untreated metastatic renal cell carcinoma. *J Clin Oncol* **23**, 832-41 (2005).

14. Leibovich, B.C. et al. A scoring algorithm to predict survival for patients with metastatic clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. *J Urol* **174**, 1759-63; discussion 1763 (2005).

15. Isbarn, H. & Karakiewicz, P.I. Predicting cancer-control outcomes in patients with renal cell carcinoma. *Curr Opin Urol* **19**, 247-57 (2009).

16. Latif, F. et al. Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science* **260**, 1317-20 (1993).

17. Gnarra, J.R. et al. Mutations of the VHL tumour suppressor gene in renal carcinoma. *Nat Genet* **7**, 85-90 (1994).

18. Shuin, T. et al. Frequent somatic mutations and loss of heterozygosity of the von Hippel-Lindau tumor suppressor gene in primary human renal cell carcinomas. *Cancer Res* **54**, 2852-5 (1994).

19. Lisztwan, J., Imbert, G., Wirbelauer, C., Gstaiger, M. & Krek, W. The von Hippel-Lindau tumor suppressor protein is a component of an E3 ubiquitin-protein ligase activity. *Genes Dev* **13**, 1822-33 (1999).

20. Cockman, M.E. et al. Hypoxia inducible factor-alpha binding and ubiquitylation by the von Hippel-Lindau tumor suppressor protein. *J Biol Chem* **275**, 25733-41 (2000).

21. Maxwell, P.H. et al. The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271-5 (1999).

22. Maynard, M.A. et al. Multiple splice variants of the human HIF-3 alpha locus are targets of the von Hippel-Lindau E3 ubiquitin ligase complex. *J Biol Chem* **278**, 11032-40 (2003).

23. Ivan, M. et al. HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. *Science* **292**, 464-8 (2001).

24. Jaakkola, P. et al. Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science* **292**, 468-72 (2001).

25. Masson, N., Willam, C., Maxwell, P.H., Pugh, C.W. & Ratcliffe, P.J. Independent function of two destruction domains in hypoxia-inducible factor-alpha chains activated by prolyl hydroxylation. *Embo J* **20**, 5197-206 (2001).

26. Kondo, K., Klco, J., Nakamura, E., Lechpammer, M. & Kaelin, W.G., Jr. Inhibition of HIF is necessary for tumor suppression by the von Hippel-Lindau protein. *Cancer Cell* **1**, 237-46 (2002).

27. Iliopoulos, O., Kibel, A., Gray, S. & Kaelin, W.G., Jr. Tumour suppression by the human von Hippel-Lindau gene product. *Nat Med* **1**, 822-6 (1995).

28. Cowey, C.L., Fielding, J.R. & Rathmell, W.K. The loss of radiographic enhancement in primary renal cell carcinoma tumors following multitargeted

receptor tyrosine kinase therapy is an additional indicator of response. *Urology* **75**, 1108-13 e1 (2010).

29. Semenza, G.L. Targeting HIF-1 for cancer therapy. *Nat Rev Cancer* **3**, 721-32 (2003).

30. Arany, Z. et al. An essential role for p300/CBP in the cellular response to hypoxia. *Proc Natl Acad Sci U S A* **93**, 12969-73 (1996).

31. Maynard, M.A. et al. Dominant-negative HIF-3 alpha 4 suppresses VHL-null renal cell carcinoma progression. *Cell Cycle* **6**, 2810-6 (2007).

32. Maynard, M.A. et al. Human HIF-3alpha4 is a dominant-negative regulator of HIF-1 and is down-regulated in renal cell carcinoma. *Faseb J* **19**, 1396-406 (2005).

33. Edgren, M., Lennernas, B., Larsson, A. & Nilsson, S. Serum concentrations of VEGF and b-FGF in renal cell, prostate and urinary bladder carcinomas. *Anticancer Res* **19**, 869-73 (1999).

34. Berse, B., Brown, L.F., Van de Water, L., Dvorak, H.F. & Senger, D.R. Vascular permeability factor (vascular endothelial growth factor) gene is expressed differentially in normal tissues, macrophages, and tumors. *Mol Biol Cell* **3**, 211-20 (1992).

35. Cowey, C.L. & Rathmell, W.K. Using molecular biology to develop drugs for renal cell carcinoma. *Expert Opinion on Drug Discovery* **3**, 311-327 (2008).

36. Gordan, J.D. & Simon, M.C. Hypoxia-inducible factors: central regulators of the tumor phenotype. *Curr Opin Genet Dev* **17**, 71-7 (2007).

37. Lee, C.M. et al. VHL Type 2B gene mutation moderates HIF dosage in vitro and in vivo. *Oncogene* **28**, 1694-705 (2009).

38. Rathmell, W.K. et al. In vitro and in vivo models analyzing von Hippel-Lindau disease-specific mutations. *Cancer Res* **64**, 8595-603 (2004).

39. Gordan, J.D. et al. HIF-alpha effects on c-Myc distinguish two subtypes of sporadic VHL-deficient clear cell renal carcinoma. *Cancer Cell* **14**, 435-46 (2008).

40. Dalgliesh, G.L. et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**, 360-3 (2010).

41. Rathmell, W.K., Martz, C.A. & Rini, B.I. Renal cell carcinoma. *Curr Opin Oncol* **19**, 234-40 (2007).

42. Isaacs, J.S. et al. HIF overexpression correlates with biallelic loss of fumarate hydratase in renal cancer: novel role of fumarate in regulation of HIF stability. *Cancer Cell* **8**, 143-53 (2005).

43. Selak, M.A. et al. Succinate links TCA cycle dysfunction to oncogenesis by inhibiting HIF-alpha prolyl hydroxylase. *Cancer Cell* **7**, 77-85 (2005).

44. Richard, D.E., Berra, E., Gothie, E., Roux, D. & Pouyssegur, J. p42/p44 mitogen-activated protein kinases phosphorylate hypoxia-inducible factor 1alpha (HIF-1alpha) and enhance the transcriptional activity of HIF-1. *J Biol Chem* **274**, 32631-7 (1999).

45. Dowling, R.J., Topisirovic, I., Fonseca, B.D. & Sonenberg, N. Dissecting the role of mTOR: lessons from mTOR inhibitors. *Biochim Biophys Acta* **1804**, 433-9 (2010).

46. Meloni-Ehrig, A.M. Renal cancer: cytogenetic and molecular genetic aspects. *Am J Med Genet* **115**, 164-72 (2002).

47. Moch, H. et al. Genetic aberrations detected by comparative genomic hybridization are associated with clinical outcome in renal cell carcinoma. *Cancer Res* **56**, 27-30 (1996).

48. Bugert, P. & Kovacs, G. Molecular differential diagnosis of renal cell carcinomas by microsatellite analysis. *Am J Pathol* **149**, 2081-8 (1996).

49. Thrash-Bingham, C.A., Salazar, H., Freed, J.J., Greenberg, R.E. & Tartof, K.D. Genomic alterations and instabilities in renal cell carcinomas and their relationship to tumor pathology. *Cancer Res* **55**, 6189-95 (1995).

50. Beroud, C. et al. Correlations of allelic imbalance of chromosome 14 with adverse prognostic parameters in 148 renal cell carcinomas. *Genes Chromosomes Cancer* **17**, 215-24 (1996).

51. Gunawan, B. et al. Prognostic impacts of cytogenetic findings in clear cell renal cell carcinoma: gain of 5q31-qter predicts a distinct clinical phenotype with favorable prognosis. *Cancer Res* **61**, 7731-8 (2001).

52. Presti, J.C., Jr. et al. Allelic loss on chromosomes 8 and 9 correlates with clinical outcome in locally advanced clear cell carcinoma of the kidney. *J Urol* **167**, 1464-8 (2002).

53. Schraml, P. et al. CDKNA2A mutation analysis, protein expression, and deletion mapping of chromosome 9p in conventional clear-cell renal carcinomas: evidence for a second tumor suppressor gene proximal to CDKN2A. *Am J Pathol* **158**, 593-601 (2001).

54. Schullerus, D., Herbers, J., Chudek, J., Kanamaru, H. & Kovacs, G. Loss of heterozygosity at chromosomes 8p, 9p, and 14q is associated with stage and grade of non-papillary renal cell carcinomas. *J Pathol* **183**, 151-5 (1997).

55. Kinoshita, H. et al. Contribution of chromosome 9p21-22 deletion to the progression of human renal cell carcinoma. *Jpn J Cancer Res* **86**, 795-9 (1995).

56. Yoshimoto, T. et al. High-resolution analysis of DNA copy number alterations and gene expression in renal clear cell carcinoma. *J Pathol* **213**, 392-401 (2007).

57. Chen, M. et al. Genome-wide profiling of chromosomal alterations in renal cell carcinoma using high-density single nucleotide polymorphism arrays. *Int J Cancer* **125**, 2342-8 (2009).

58. Pei, J. et al. Combined classical cytogenetics and microarray-based genomic copy number analysis reveal frequent 3;5 rearrangements in clear cell renal cell carcinoma. *Genes Chromosomes Cancer* **49**, 610-9 (2010).

59. Toma, M.I. et al. Loss of heterozygosity and copy number abnormality in clear cell renal cell carcinoma discovered by high-density affymetrix 10K single nucleotide polymorphism mapping array. *Neoplasia* **10**, 634-42 (2008).

60. Bissig, H. et al. Evaluation of the clonal relationship between primary and metastatic renal cell carcinoma by comparative genomic hybridization. *Am J Pathol* **155**, 267-74 (1999).

61. Gronwald, J. et al. Comparison of DNA gains and losses in primary renal clear cell carcinomas and metastatic sites: importance of 1q and 3p copy number changes in metastatic events. *Cancer Res* **57**, 481-7 (1997).

62. Jiang, F. et al. Construction of evolutionary tree models for renal cell carcinoma from comparative genomic hybridization data. *Cancer Res* **60**, 6503-9 (2000).

63. Arai, E. et al. Genetic clustering of clear cell renal cell carcinoma based on array-comparative genomic hybridization: its association with DNA methylation alteration and patient outcome. *Clin Cancer Res* **14**, 5531-9 (2008).

64. Zhang, Z., Wondergem, B. & Dykema, K. A Comprehensive Study of Progressive Cytogenetic Alterations in Clear Cell Renal Cell Carcinoma and a New Model for ccRCC Tumorigenesis and Progression. *Adv Bioinformatics*, 428325 (2010).

65. Beroukhim, R. et al. Patterns of gene expression and copy-number alterations in von-hippel lindau disease-associated and sporadic clear cell carcinoma of the kidney. *Cancer Res* **69**, 4674-81 (2009).

66. Klatte, T. et al. Cytogenetic profile predicts prognosis of patients with clear cell renal cell carcinoma. *J Clin Oncol* **27**, 746-53 (2009).

67. La Rochelle, J. et al. Chromosome 9p deletions identify an aggressive phenotype of clear cell renal cell carcinoma. *Cancer* (2010).

68. Perou, C.M. et al. Molecular portraits of human breast tumours. *Nature* **406**, 747-52 (2000).

69. Sorlie, T. et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**, 10869-74 (2001).

70. van de Vijver, M.J. et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347**, 1999-2009 (2002).

71. Paik, S. et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* **351**, 2817-26 (2004).

72. Dunn, L. & Demichele, A. Genomic predictors of outcome and treatment response in breast cancer. *Mol Diagn Ther* **13**, 73-90 (2009).

73. Boer, J.M. et al. Identification and classification of differentially expressed genes in renal cell carcinoma by expression profiling on a global human 31,500-element cDNA array. *Genome Res* **11**, 1861-70 (2001).

74. Gieseg, M.A. et al. Expression profiling of human renal carcinomas with functional taxonomic analysis. *BMC Bioinformatics* **3**, 26 (2002).

75. Skubitz, K.M. & Skubitz, A.P. Differential gene expression in renal-cell cancer. *J Lab Clin Med* **140**, 52-64 (2002).

76. Lenburg, M.E. et al. Previously unidentified changes in renal cell carcinoma gene expression identified by parametric analysis of microarray data. *BMC Cancer* **3**, 31 (2003).

77. Liou, L.S. et al. Microarray gene expression profiling and analysis in renal cell carcinoma. *BMC Urol* **4**, 9 (2004).

78. Hirota, E. et al. Genome-wide gene expression profiles of clear cell renal cell carcinoma: identification of molecular targets for treatment of renal cell carcinoma. *Int J Oncol* **29**, 799-827 (2006).

79. Dalgin, G.S., Holloway, D.T., Liou, L.S. & Delisi, C. Identification and characterization of renal cell carcinoma gene markers. *Cancer Inform* **3**, 65-92 (2007).

80. Young, A.N. et al. Expression profiling of renal epithelial neoplasms: a method for tumor classification and discovery of diagnostic molecular markers. *Am J Pathol* **158**, 1639-51 (2001).

81. Yamazaki, K. et al. Overexpression of KIT in chromophobe renal cell carcinoma. *Oncogene* **22**, 847-52 (2003).

82. Takahashi, M. et al. Molecular subclassification of kidney tumors and the discovery of new diagnostic markers. *Oncogene* **22**, 6810-8 (2003).

83. Higgins, J.P. et al. Gene expression patterns in renal cell carcinoma assessed by complementary DNA microarray. *Am J Pathol* **162**, 925-32 (2003).

84. Schuetz, A.N. et al. Molecular classification of renal tumors by gene expression profiling. *J Mol Diagn* **7**, 206-18 (2005).

85. Sultmann, H. et al. Gene expression in kidney cancer is associated with cytogenetic abnormalities, metastasis formation, and patient survival. *Clin Cancer Res* **11**, 646-55 (2005).

86. Rogers, C.G. et al. Microarray gene expression profiling using core biopsies of renal neoplasia. *Am J Transl Res* **1**, 55-61 (2009).

87. Takahashi, M. et al. Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification. *Proc Natl Acad Sci U S A* **98**, 9754-9 (2001).

88. Vasselli, J.R. et al. Predicting survival in patients with metastatic kidney cancer by gene-expression profiling in the primary tumor. *Proc Natl Acad Sci U S A* **100**, 6958-63 (2003).

89. Jones, J. et al. Gene signatures of progression and metastasis in renal cell cancer. *Clin Cancer Res* **11**, 5730-9 (2005).

90. Kosari, F. et al. Clear cell renal cell carcinoma: gene expression analyses identify a potential signature for tumor aggressiveness. *Clin Cancer Res* **11**, 5128-39 (2005).

91. Yao, M. et al. Gene expression analysis of renal carcinoma: adipose differentiation-related protein as a potential diagnostic and prognostic biomarker for clear-cell renal carcinoma. *J Pathol* **205**, 377-87 (2005).

92. Zhao, H. et al. Gene expression profiling predicts survival in conventional renal cell carcinoma. *PLoS Med* **3**, e13 (2006).

93. Yao, M. et al. A three-gene expression signature model to predict clinical outcome of clear cell renal carcinoma. *Int J Cancer* **123**, 1126-32 (2008).

94. Wuttig, D. et al. Gene signatures of pulmonary metastases of renal cell carcinoma reflect the disease-free interval and the number of metastases per patient. *Int J Cancer* **125**, 474-82 (2009).

95. Skubitz, K.M., Zimmermann, W., Kammerer, R., Pambuccian, S. & Skubitz, A.P. Differential gene expression identifies subgroups of renal cell carcinoma. *J Lab Clin Med* **147**, 250-67 (2006).

96. Zhao, H. et al. Alteration of gene expression signatures of cortical differentiation and wound response in lethal clear cell renal cell carcinomas. *PLoS One* **4**, e6039 (2009).

97. Brannon, A.R. et al. Molecular Stratification of Clear Cell Renal Cell Carcinoma by Consensus Clustering Reveals Distinct Subtypes and Survival Patterns. *Genes and Cancer* **In press**(2010).

98. Shioi, K. et al. Vascular cell adhesion molecule 1 predicts cancer-free survival in clear cell renal carcinoma patients. *Clin Cancer Res* **12**, 7339-46 (2006).

99. Parker, A.S. et al. High expression levels of survivin protein independently predict a poor outcome for patients who undergo surgery for clear cell renal cell carcinoma. *Cancer* **107**, 37-45 (2006).

100. Parker, A.S. et al. Development and evaluation of BioScore: a biomarker panel to enhance prognostic algorithms for clear cell renal cell carcinoma. *Cancer* **115**, 2092-103 (2009).

101. Rioux-Leclercq, N. et al. Immunohistochemical analysis of tumor polyamines discriminates high-risk patients undergoing nephrectomy for renal cell carcinoma. *Hum Pathol* **35**, 1279-84 (2004).

102. Rioux-Leclercq, N. et al. Value of immunohistochemical Ki-67 and p53 determinations as predictive factors of outcome in renal cell carcinoma. *Urology* **55**, 501-5 (2000).

103. Bui, M.H. et al. Prognostic value of carbonic anhydrase IX and KI67 as predictors of survival for renal clear cell carcinoma. *J Urol* **171**, 2461-6 (2004).

104. Visapaa, H. et al. Correlation of Ki-67 and gelsolin expression to clinical outcome in renal clear cell carcinoma. *Urology* **61**, 845-50 (2003).

105. Delahunt, B., Bethwaite, P.B., Thornton, A. & Ribas, J.L. Proliferation of renal cell carcinoma assessed by fixation-resistant polyclonal Ki-67 antibody labeling. Correlation with clinical outcome. *Cancer* **75**, 2714-9 (1995).

106. de Riese, W.T. et al. Prognostic significance of Ki-67 immunostaining in nonmetastatic renal cell carcinoma. *J Clin Oncol* **11**, 1804-8 (1993).

107. Dudderidge, T.J. et al. Mcm2, Geminin, and KI67 define proliferative state and are prognostic markers in renal cell carcinoma. *Clin Cancer Res* **11**, 2510-7 (2005).

108. Kallio, J.P. et al. Renal cell carcinoma MIB-1, Bax and Bcl-2 expression and prognosis. *J Urol* **172**, 2158-61 (2004).

109. Aaltomaa, S. et al. Expression of cyclins A and D and p21(waf1/cip1) proteins in renal cell cancer and their relation to clinicopathological variables and patient survival. *Br J Cancer* **80**, 2001-7 (1999).

110. Moch, H. et al. p53 protein expression but not mdm-2 protein expression is associated with rapid tumor cell proliferation and prognosis in renal cell carcinoma. *Urol Res* **25 Suppl 1**, S25-30 (1997).

111. Tannapfel, A. et al. Incidence of apoptosis, cell proliferation and P53 expression in renal cell carcinomas. *Anticancer Res* **17**, 1155-62 (1997).

112. Dahinden, C. et al. Mining tissue microarray data to uncover combinations of biomarker expression patterns that improve intermediate staging and grading of clear cell renal cell cancer. *Clin Cancer Res* **16**, 88-98 (2010).

113. Nogueira, M. & Kim, H.L. Molecular markers for predicting prognosis of renal cell carcinoma. *Urol Oncol* **26**, 113-24 (2008).

114. Perez-Gracia, J.L. et al. Identification of TNF-alpha and MMP-9 as potential baseline predictive serum markers of sunitinib activity in patients with renal cell carcinoma using a human cytokine array. *Br J Cancer* **101**, 1876-83 (2009).

115. Rini, B.I. et al. Antitumor activity and biomarker analysis of sunitinib in patients with bevacizumab-refractory metastatic renal cell carcinoma. *J Clin Oncol* **26**, 3743-8 (2008).

116. Deprimo, S.E. et al. Circulating protein biomarkers of pharmacodynamic activity of sunitinib in patients with metastatic renal cell carcinoma: modulation of VEGF and VEGF-related proteins. *J Transl Med* **5**, 32 (2007).

117. Nakada, C. et al. Genome-wide microRNA expression profiling in renal cell carcinoma: significant down-regulation of miR-141 and miR-200c. *J Pathol* **216**, 418-27 (2008).

118. Chow, T.F. et al. Differential expression profiling of microRNAs and their potential involvement in renal cell carcinoma pathogenesis. *Clin Biochem* (2009).

119. Juan, D. et al. Identification of a MicroRNA Panel for Clear-cell Kidney Cancer. *Urology* (2009).

120. Liu, H. et al. Identifying direct mRNA targets of microRNA dysregulated in cancer: with application to clear cell Renal Cell Carcinoma. *BMC Systems Biology* **(In revision)**(2010).

121. Petillo, D. et al. MicroRNA profiling of human kidney cancer subtypes. *Int J Oncol* **35**, 109-14 (2009).

122. Sinha, S., Dutta, S., Datta, K., Ghosh, A.K. & Mukhopadhyay, D. Von Hippel-Lindau gene product modulates TIS11B expression in renal cell carcinoma: impact on vascular endothelial growth factor expression in hypoxia. *J Biol Chem* **284**, 32610-8 (2009).

123. Nelson, P.T. et al. RAKE and LNA-ISH reveal microRNA expression and localization in archival human brain. *Rna* **12**, 187-91 (2006).

124. Chen, X. et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res* **18**, 997-1006 (2008).

125. Resnick, K.E. et al. The detection of differentially expressed microRNAs from the serum of ovarian cancer patients using a novel real-time PCR platform. *Gynecol Oncol* **112**, 55-9 (2009).

126. Mitchell, P.S. et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* **105**, 10513-8 (2008).

127. Lawrie, C.H. et al. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. *Br J Haematol* **141**, 672-5 (2008).

128. Gilad, S. et al. Serum microRNAs are promising novel biomarkers. *PLoS One* **3**, e3148 (2008).

129. Kim, H.L. et al. Using tumor markers to predict the survival of patients with metastatic renal cell carcinoma. *J Urol* **173**, 1496-501 (2005).

130. Ficarra, V. et al. Multiinstitutional European validation of the 2002 TNM staging system in conventional and papillary localized renal cell carcinoma. *Cancer* **104**, 968-74 (2005).

131. *Cancer facts and figures 2009*, (American Cancer Society, Inc, Atlanta, GA, 2009).

132. Banks, R.E. et al. Genetic and epigenetic analysis of von Hippel-Lindau (VHL) gene alterations and relationship with clinical variables in sporadic renal cancer. *Cancer Res* **66**, 2000-11 (2006).

133. Nickerson, M.L. et al. Improved identification of von Hippel-Lindau gene alterations in clear cell renal tumors. *Clin Cancer Res* **14**, 4726-34 (2008).

134. Lam, J.S. et al. Postoperative surveillance protocol for patients with localized and locally advanced renal cell carcinoma based on a validated prognostic nomogram and risk group stratification system. *J Urol* **174**, 466-72; discussion 472; quiz 801 (2005).

135. Monti, S., Tamayo, P., Mesirov, J. & Golub, T. Consensus Clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning Journal* **52**, 91-118 (2003).

136. Dalgin, G.S. et al. Portraits of breast cancer progression. *BMC Bioinformatics* **8**, 291 (2007).

137. Alexe, G., Dalgin, G.S., Ramaswamy, R., DeLisi, C. & Bhanot, G. Data Perturbation Independent Diagnosis and Validation of Breast Cancer Subtypes Using Clustering and Patterns. *Cancer Informatics* **2**, 243-274 (2006).

138. Alexe, G. et al. High expression of lymphocyte-associated genes in node-negative HER2+ breast cancers correlates with lower recurrence rates. *Cancer Res* **67**, 10669-76 (2007).

139. Young, A.N., Master, V.A., Paner, G.P., Wang, M.D. & Amin, M.B. Renal epithelial neoplasms: diagnostic applications of gene expression profiling. *Adv Anat Pathol* **15**, 28-38 (2008).

140. Reddy, A. et al. Logical Analysis of Data (LAD) model for the early diagnosis of acute ischemic stroke. *BMC Med Inform Decis Mak* **8**, 30 (2008).

141. Kass, R.E. & Raftery, A.E. Bayes Factors. *JASA* **90**, 773-795 (1995).

142. Seiler, M., Huang, C.C., Szalma, S. & Bhanot, G. ConsensusCluster: a stand-alone software tool for unsupervised cluster discovery in numerical data. *OMICS*, (in press) (2009).

143. Jolliffe, I.T. *Principal Component Analysis*, 487 (Springer-Verlag, New York, 2002).

144. Wall, M.E., Rechtsteiner, A. & Rocha, L.M. Singular value decomposition and principal component analysis. in *A Practical Approach to Microarray Data Analysis* (eds. Berrar, D.P., Dubitzky, W., Granzow, M. & Norwell, M.A.) 91-109 (Kluwer Academic Publishers, Boston, MA, 2003).

145. Everitt, B.S. & Dunn, G. *Applied Multivariate Data Analysis*, (Hodder Arnold Publication, London, 2001).

146. Kohonen, T. *Self-Organizing Maps*, (Springer, New York, 2001).

147. Crama, Y., Hammer, P.L. & Ibaraki, T. Cause-Effect Relationship and Partially Defined Boolean Functions. *Annals of Operation Research* **16**, 299-326 (1988).

148. Hammer, P.L. & Bonates, T.O. Logical Analysis of Data - An Overview: From combinatorial optimization to medical applications. *Annals of Operation Research* **148**, 203-225 (2006).

149. Ellwood-Yen, K. et al. Myc-driven murine prostate cancer shares molecular features with human prostate tumors. *Cancer Cell* **4**, 223-38 (2003).

150. Schambony, A. & Wedlich, D. Wnt-5A/Ror2 regulate expression of XPAPC through an alternative noncanonical signaling pathway. *Dev Cell* **12**, 779-92 (2007).

151. MacLeod, R.J., Hayes, M. & Pacheco, I. Wnt5a secretion stimulated by the extracellular calcium-sensing receptor inhibits defective Wnt signaling in colon cancer cells. *Am J Physiol Gastrointest Liver Physiol* **293**, G403-11 (2007).

152. Nishita, M. et al. Ror2/Frizzled complex mediates Wnt5a-induced AP-1 activation by regulating Dishevelled polymerization. *Mol Cell Biol* **30**, 3610-9 (2010).

153. Feike, A., Rachor, K., Gentzel, M. & Schambony, A. Wnt5a/Ror2-inducedupregulation of xPAPCrequiresxShcA. *Biochem Biophys Res Commun* (2010).

154. Wright, T.M. et al. Ror2, a developmentally regulated kinase, promotes tumor growth potential in renal cell carcinoma. *Oncogene* **28**, 2513-23 (2009).

155. Stolle, C. et al. Improved detection of germline mutations in the von Hippel-Lindau disease tumor suppressor gene. *Hum Mutat* **12**, 417-23 (1998).

156. Herman, J.G. et al. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A* **91**, 9700-4 (1994).

157. Cairns, P., Tokino, K., Eby, Y. & Sidransky, D. Localization of tumor suppressor loci on chromosome 9 in primary human renal cell carcinomas. *Cancer Res* **55**, 224-7 (1995).

158. Tomlins, S.A. et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-8 (2005).

159. Fisher, R. On the Interpretation of χ2 from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society.* **85**, 87-94 (1922).

160. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological).* **57**, 289-300 (1995).

161. Geiss, G.K. et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* **26**, 317-25 (2008).

162. de Jonge, H.J. et al. Evidence based selection of housekeeping genes. *PLoS One* **2**, e898 (2007).

163. Masuda, N., Ohnishi, T., Kawamoto, S., Monden, M. & Okubo, K. Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. *Nucleic Acids Res* **27**, 4436-43 (1999).

164. Hahn, O.M. et al. Dynamic contrast-enhanced magnetic resonance imaging pharmacodynamic biomarker study of sorafenib in metastatic renal carcinoma. *J Clin Oncol* **26**, 4572-8 (2008).

165. Sandlund, J. et al. Hypoxia-inducible factor-2alpha mRNA expression in human renal cell carcinoma. *Acta Oncol* **48**, 909-14 (2009).

166. Gu, Y.Z., Moran, S.M., Hogenesch, J.B., Wartman, L. & Bradfield, C.A. Molecular characterization and chromosomal localization of a third alpha-class hypoxia inducible factor subunit, HIF3alpha. *Gene Expr* **7**, 205-13 (1998).

167. Hara, S., Hamada, J., Kobayashi, C., Kondo, Y. & Imura, N. Expression and characterization of hypoxia-inducible factor (HIF)-3alpha in human kidney: suppression of HIF-mediated gene expression by HIF-3alpha. *Biochem Biophys Res Commun* **287**, 808-13 (2001).

168. Pasanen, A. et al. Hypoxia-inducible factor (HIF)-3alpha is subject to extensive alternative splicing in human tissues and cancer cells and is regulated by HIF-1 but not HIF-2. *Int J Biochem Cell Biol* **42**, 1189-200 (2010).

169. O'Connell, M.P. et al. The orphan tyrosine kinase receptor, ROR2, mediates Wnt5A signaling in metastatic melanoma. *Oncogene* **29**, 34-44 (2010).

170. Yamamoto, H. et al. Wnt5a signaling is involved in the aggressiveness of prostate cancer and expression of metalloproteinase. *Oncogene* **29**, 2036-46 (2010).

171. Kubo, T. et al. Resequencing and copy number analysis of the human tyrosine kinase gene family in poorly differentiated gastric cancer. *Carcinogenesis* **30**, 1857-64 (2009).

172. Enomoto, M. et al. Autonomous regulation of osteosarcoma cell invasiveness by Wnt5a/Ror2 signaling. *Oncogene* **28**, 3197-208 (2009).

173. Morioka, K. et al. Orphan receptor tyrosine kinase ROR2 as a potential therapeutic target for osteosarcoma. *Cancer Sci* **100**, 1227-33 (2009).

174. Kobayashi, M. et al. Ror2 expression in squamous cell carcinoma and epithelial dysplasia of the oral cavity. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* **107**, 398-406 (2009).

175. Ohta, H. et al. Cross talk between hedgehog and epithelial-mesenchymal transition pathways in gastric pit cells and in diffuse-type gastric cancers. *Br J Cancer* **100**, 389-98 (2009).

176. Mitsumori, K. et al. Chromosome 14q LOH in localized clear cell renal cell carcinoma. *J Pathol* **198**, 110-4 (2002).

177. Wu, S.Q. et al. The correlation between the loss of chromosome 14q with histologic tumor grade, pathologic stage, and outcome of patients with nonpapillary renal cell carcinoma. *Cancer* **77**, 1154-60 (1996).

178. Kaku, H. et al. Positive correlation between allelic loss at chromosome 14q24-31 and poor prognosis of patients with renal cell carcinoma. *Urology* **64**, 176-81 (2004).

179. Chi, X.Z. et al. RUNX3 suppresses gastric epithelial cell growth by inducing p21(WAF1/Cip1) expression in cooperation with transforming growth factor {beta}-activated SMAD. *Mol Cell Biol* **25**, 8097-107 (2005).

180. Li, Q.L. et al. Causal relationship between the loss of RUNX3 expression and gastric cancer. *Cell* **109**, 113-24 (2002).