Investigating the genomic underpinnings of renal cell carcinoma, genetic predisposition, and their clinical implications.

KATE I. GLENNON

Department of Human Genetics, Faculty of Medicine and Health Sciences,
McGill University, Montréal, Quebec, Canada
December, 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

TABLE OF CONTENTS

ABSTRACT

Renal cell carcinoma (RCC) comprises the most common form of kidney cancer. Among RCCs, clear cell subtype (ccRCC) accounts for 75-80% of cases, and is characterized by heterogeneous clinical outcomes and extreme incidental variances. Though surgery to remove the tumor (partial or radical nephrectomy) is considered curative for patients with localized disease, 30-40% of these patients experience relapse or metastasis post-nephrectomy. Identifying patients who may benefit from systemic therapy remains difficult, and there are currently no routine biomarkers of disease prognosis. Additionally, global incidence of RCC is highly variable, and little is understood about germline predisposition for sporadic RCC.

Recent genomic studies have suggested that understanding the genetic heterogeneity among RCC tumors may help with prognosis, however, there is a lack of RCC-appropriate genomic assays to facilitate the large-scale screening necessary for biomarker development. This study aimed to investigate the genomic landscape of RCC to deepen our understanding of the genetic evolution of RCC tumors, genetic predisposition to RCC, and the associated clinical consequences.

To facilitate investigating the genomics of RCC progression, we developed a targeted sequencing assay and corresponding bioinformatic workflows to detect somatic mutations in RCC-relevant genes. We demonstrated the ability of the assay to detect somatic tumor mutations through sequencing of patient-matched normal, tumor, and liquid biopsies. Additionally, we interrogated different liquid biopsy fractions, including soluble cell free DNA (cfDNA) and exosomal DNA (evDNA), to compare their ability to capture relevant somatic information from the tumor.

We next explored the role of tumor genomics as biomarkers for risk prediction. We interrogated the somatic mutation status among a large multinational cohort of ccRCC cases, to identify recurrently mutated genes, patterns in concurrently mutated genes, and evaluated their

associations to clinically relevant endpoints. We identified four genomically distinct groups of tumors with divergent rates of relapse. These defined groups account for the ~80% of ccRCC cases with mutations in the *VHL* gene and offer the potential for identifying patients who are at a high risk of relapse and should be prioritized for adjuvant therapy, and those at low risk who could potentially be spared adjuvant therapy.

Finally, we aimed to get a deeper understanding into the germline genetic susceptibility to sporadic RCC. Within a large Canadian cohort, we interrogated germline pathogenic variants to identify potential risk genes for RCC within Canada. We identified *CHEK2* and *ATM* as potential risk-genes for clear cell RCC, and the *FH* gene for non-clear cell subtypes. Notably, we also identified an association between germline pathogenic variants in *BRCA1*, *BRCA2*, and *ATM* genes and the presence of metastasis. Comparing risk genes among large cohort from various countries also revealed global differences in genetic susceptibility to RCC. Lastly, we identified that globally, clinical guidelines for genetic screening for RCC fail to include over 70% of patients with pathogenic variants, highlighting the need to revise referral criteria.

Overall, these findings advance our understanding of the genomic factors underlying clear cell RCC, and their associations to clinical features and disease recurrence. Through the development of NGS methods relevant to RCC, and introducing a genomic classifier for risk-stratification for RCC, this study provides insight into the clinical utility of genome-based prognosis in RCC. Evaluation of germline risk-genes for RCC, along with genetic screening criteria, helps to provide a clearer picture of global differences in RCC susceptibility, and how we can improve precision preventative strategies in RCC.

RÉSUMÉ

Le carcinome rénal (CCR) constitue la forme la plus courante de cancer du rein. Parmi les CCR, le sous-type de cellules claires (CCRcc) représente 75 à 80 % des cas et se distingue par des résultats cliniques hétérogènes et des variances fortuites extrêmes. Bien que la chirurgie visant à enlever la tumeur (néphrectomie partielle ou radicale) soit considérée curative pour les patients atteints d'une maladie localisée, 30 à 40 % de ces patients présentent une rechute ou des métastases après la néphrectomie. L'identification des patients susceptibles de bénéficier d'un traitement systémique reste difficile et il n'existe actuellement aucun biomarqueur de routine pour le pronostic de la maladie. De plus, l'incidence mondiale du CCR est très variable et on sait peu de choses sur la prédisposition germinale aux CCR sporadiques.

Des études génomiques récentes ont suggéré que la compréhension de l'hétérogénéité génétique des tumeurs CCR pourrait aider au pronostic. Cependant, il existe un manque de tests génomiques appropriés au CCR pour faciliter le dépistage à grande échelle nécessaire au développement de biomarqueurs. Cette étude visait à étudier le paysage mutationnel génomique du CCR afin d'approfondir notre compréhension de l'évolution génétique des tumeurs du CCR, de la prédisposition génétique au CCR et des conséquences cliniques associées.

Pour faciliter l'étude génomique de la progression du CCR, nous avons développé un test de séquençage ciblé et des flux de travail bioinformatique correspondants pour détecter les mutations somatiques dans les gènes pertinents pour le CCR. Nous avons démontré la capacité du test à détecter les mutations tumorales somatiques grâce au séquençage de biopsies normales, tumorales et liquides correspondant à chaque patient. De plus, nous avons interrogé différentes fractions de biopsie liquide, notamment l'ADN libre circulant et l'ADN exosomal, pour comparer leur capacité à capturer des informations somatiques pertinentes de la tumeur.

Nous avons ensuite exploré le rôle de la génomique tumorale en tant que biomarqueurs pour la prédiction des risques. Nous avons examiné le statut des mutations somatiques parmi une grande cohorte multinationale de cas de CCRcc, afin d'identifier les gènes mutés de manière récurrente, les gènes mutés de manière concomitante et évalué leurs associations avec des paramètres cliniques pertinents. Nous avons identifié quatre groupes de tumeurs génomiquement distincts avec des taux de rechute divergents. Ces groupes définis représentent environ 80 % des cas de CCRcc présentant des mutations du gène VHL et offrent la possibilité d'identifier les patients qui présentent un risque élevé de rechute et devraient être prioritaires pour un traitement adjuvant, et ceux à faible risque qui pourraient potentiellement être épargnés d'un traitement adjuvant.

Enfin, nous avions pour objectif de mieux comprendre la susceptibilité génétique germinale au CCR sporadique. Au sein d'une vaste cohorte canadienne, nous avons examiné des variants pathogènes germinaux afin d'identifier les gènes de risque potentiel de CCR au Canada. Nous avons identifié *CHEK2* et *ATM* comme les gènes de risque potentiel pour le CCR à cellules claires, ainsi que le gène *FH* pour les sous-types de cellules non claires. Notamment, nous avons également identifié une association entre les variants pathogènes germinaux des gènes *BRCA1*, *BRCA2* et *ATM* et la présence de métastases. La comparaison des gènes de risque au sein d'une grande cohorte de divers pays a également révélé des différences mondiales dans la susceptibilité génétique au CCR. Enfin, nous avons identifié qu'à l'échelle mondiale, les directives cliniques pour le dépistage génétique du CCR n'incluent pas plus de 70 % des patients présentant des variants pathogènes, soulignant la nécessité de réviser les critères de référence.

Dans l'ensemble, ces résultats font progresser notre compréhension des facteurs génomiques sous-jacents au CCR à cellules claires et de leurs associations avec les caractéristiques

cliniques et la récidive de la maladie. Grâce au développement de méthodes NGS pertinents pour le CCR et à l'introduction d'un classificateur génomique pour la stratification des risques pour le CCR, cette étude donne un aperçu de l'utilité clinique du pronostic basé sur le génome dans le CCR. L'évaluation des gènes germinaux à risque pour le CCR, ainsi que les critères de dépistage génétique, contribuent à fournir une image plus claire des différences mondiales en matière de susceptibilité au CCR et à la manière dont nous pouvons améliorer les stratégies préventives précises dans le CCR.

LIST OF ABBREVIATIONS

Gene and protein names follow standard nomenclature.

Abbreviation	Definition
ACMG	American College of Medical Genetics
AKT	Protein kinase B
AUC	Animal Use Protocol
BAP1	BRCA1 associated protein-1
BHD	Birt-Hogg-Dubé syndrome
BRCA1	Breast cancer gene 1
CSS	Cancer specific survival
CAGEKID	Cancer Genomics of the Kidney
CCAC	Canadian Council of Animal Care
ccRCC	Clear cell renal cell carcinoma
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A
cfDNA	Cell free DNA
cfRNA	Cell free RNA
chRCC	Chromophobe renal cell carcinoma
CI	Confidence interval
CNA	Copy number alteration
CTC	Circulating tumor cell
ctDNA	Circulating tumor DNA
CTLA-4	Cytotoxic T lymphocyte-associated protein 4
ddPCR	Digital droplet PCR
DDR	DNA damage response
DFS	Disease-free survival
DNA	Deoxyribonucleic acid
EGA	European Genome-Phenome Archive
ELOC	Transcription elongation factor B (also known as *TCEB1*)
EV	Extracellular vesicles
evDNA	Exosomal DNA
FFPE	Formalin-Fixed Paraffin-Embedded
FH	Fumarate hydratase
FLCN	Folliculin
G4	Fuhrman grade 4
gDNA	Genomic DNA
H3K36	Histone H3 lysine 36
HIF1α	Hypoxia-inducible factors 1α
HIF2α	Hypoxia-inducible factors 2α
HLRCC	Hereditary papillary renal cell carcinoma

Abbreviation	Definition
HPRCC	Hereditary leiomyomatosis and kidney cell cancer
HR	Hazards ratio
ICI	Immune checkpoint inhibitor
IFNα	Interferon-α
IGV	Integrative genomics viewer
IL-2	Interleukin-2
indels	Insertion and deletions
ITH	Intratumor heterogeneity
IVC	Inferior vena cava
KCRNC	Kidney Cancer Research Network of Canada
Ki67	Staining positive for Ki67
LOD	Limit of detection
LOH	Loss of heterozygosity
MET	Mesenchymal Epithelial Transition
mRCC	Metastatic renal cell carcinoma
MRD	Minimal residual disease
mTOR	Mammalian target of rapamycin
mTORC1	Mammalian target of rapamycin complex 1
MUHC REB	McGill University Health Centre Research Ethics Board
MVI	Microvascular invasion
nccRCC	Non-clear cell renal cell carcinoma
NGS	Next-generation sequencing
OR	Odds ratio
OS	Overall survival
PARP	Poly-ADP ribose polymerase
PBRM1	Polybromo 1
PD-1	Programmed-cell-death protein 1
PD-L1	Programmed death-ligand 1
PH	Proportional hazards
PI3K	Phosphoinositide 3-kinases
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
pRCC	Papillary renal cell carcinoma
PTEN	Phosphatase and tensin homolog
PV	Pathogenic variant
pVHL	VHL tumor suppressor protein
RCC	Renal cell carcinoma
REDD1	Protein regulated in development and DNA damage response 1
RFS	Recurrence free survival
RNA	Ribonucleic acid

Abbreviation	Definition
RPKM	Reads per kilobase per million reads
sCNA	Somatic copy number alteration
SD	Standard deviation
SDH-RCC	Succinate dehydrogenase renal cell carcinoma
SETD2	SET domain containing protein 2
SNV	Single nucleotide variations
SSIGN	Stage, Size, Grade, and Necrosis score
sWGS	Shallow whole genome sequencing
TCEB1	Transcription elongation factor B (also known as *ELOC*)
TCGA	The Cancer Genome Atlas
TKI	Tyrosine kinase inhibitor
TMB	Tumor mutational burden
TNM	Tumor–node–metastasis
TP53	Tumor protein p53
tRCC	Translocation renal cell carcinoma
TSC1	Tuberous sclerosis complex subunit 1
TSC2	Tuberous sclerosis complex subunit 2
UISS	UCLA Integrated Staging System
UMI	Unique molecular identifier
VAF	Variant allele frequencies
VEGF	Vascular endothelial growth factor
VEGFR	Vascular endothelial growth factor receptor
VHL	Von Hippel-Lindau
WES	Whole exome sequencing
wGII	Weighted genome instability index
WGS	Whole genome sequencing
WT	Wild-type
3p	Short arm of chromosome 3

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Dr. Yasser Riazalhosseini, for his support, guidance, and mentorship throughout the past 5 years. You have constantly encouraged me to seek valuable opportunities, taught me how to be an effective communicator, and had an immense impact on my development as a scientist.

I would also like to thank the members of my Supervisory Committee, Dr. Celia Greenwood and Dr. Jiannis Ragoussis. You have both provided consistent support and valuable insight throughout my studies. Thank you also to Dr. Yukihide Momozawa, for welcoming me to his lab in Japan for a portion of my research – this was an invaluable experience that I am extremely grateful for.

To the members of the Riazalhosseini lab, past and present: I feel so fortunate to get to work with such an amazing group of scientists, and I am grateful for the friendships we have formed along the way. I'd like to especially thank Madeleine Arseneault, for the substantial amount of support and friendship she has given me since my very first day in the lab.

To the members of the Langlais lab, for being such great lab neighbours. I always look forward to our coffee and snack breaks.

Thank you to my friends, both near and far, who have always provided me with so much encouragement and kept my spirits high with trivia nights, park picnics, cottage weekends away and the occasional cat photo. A special thank you to Brenna, for the daily facetime calls to help me decide what to make for dinner, and for helping me to celebrate every success, big or small.

And finally, thank you to my family. To my dad, Paul, for the endless coffee deliveries and always being my biggest supporter; my mom, Tracy, for always welcoming me home with a glass of good wine and cheering me on every step of the way. My brother, Nick, for helping to turn me into the competitive and ambitious person that I am today. Thank you to all my grandparents for their constant encouragement. My cat, Bella, for joining me (uninvited) on every zoom call; and my beloved dog Cocoa, who I miss immensely, for being the best friend I could ask for.

CONTRIBUTION TO ORIGINAL KNOWLEDGE

The following components of this theses are considered original knowledge and demonstrate the contributions made to the field:

First, Chapter 2 addresses the lack of appropriate targeted NGS panels for investigating renal cancer, and describes the development of a targeted sequencing assay capable of detecting somatic mutations in RCC-relevant genes. The assay provides a reliable platform to conduct parallel mutational analyses of tumor and liquid biopsies, and we provide evidence on the feasibility to capture tumor-associated genetic alterations in blood-based liquid biopsies.

Second, Chapter 3 introduces to the field the largest-to-date cohort of ccRCCs with detailed clinical information and genomic characterization. This dataset represents a valuable resource for investigating patterns in the genomic landscape of ccRCC and their associations to prognosis. Additionally, we define a genomic classifier based on a small panel of genes, and representative of evolutionary trajectories of ccRCC, that can identify groups of patients with diverging risk of disease recurrence.

Lastly, Chapter 4 represents the first investigation into risk-genes for RCC within the Canadian population. We identify *CHEK2*, *ATM*, and *FH* as risk-genes for RCC within Canada, which are not routinely included in genetic screening panels. Notably, we evaluate clinical guidelines for genetic screening and identify that >70% of patients with genetic predisposition to RCC are not captured within current inclusion criteria.

THESIS FORMAT

This thesis is prepared in a manuscript format and is composed of 7 chapters. Chapter 1 is a general introduction and literature review. Chapter 2 consists of a first-author manuscript published in *Cancers,* detailing the development of an RCC-relevant NGS assay for tissue and liquid biopsy analyses. Chapter 3 consists of a co-first author manuscript published in *Clinical Cancer Research* which investigates somatic genomic patterns in ccRCC and introduces a genomic classifier for prognosis. Chapter 4 consists of a first author manuscript, submitted to *JNCI*, which investigates germline susceptibility to RCC and highlights how current guidelines for genetic screening exclude the majority of patients with genetic susceptibility to RCC. Bridging statements are included between manuscript chapters to describe the connections between each manuscript and provide additional context. Chapter 5 is a general discussion of the findings, and Chapter 6 is a final summary and conclusions. Chapter 7 is a master bibliography of references included in the thesis, however reference lists for in-text citations within manuscript chapters are included within the respective chapter.

CONTRIBUTION OF AUTHORS

Of the manuscripts shown, Chapters 2 and 3 were previously published, and Chapter 4 has been submitted for publication. The contributions of authors are indicated below, highlighting my role for each manuscript.

Chapter 2: Rational Development of Liquid Biopsy Analysis in Renal Cell Carcinoma
Glennon, K.I.; Maralani, M.; Abdian, N.; Paccard, A.; Montermini, L.; Nam, A.J.; Arseneault, M.; Staffa, A.; Jandaghi, P.; Meehan, B.; et al. Rational Development of Liquid Biopsy Analysis in Renal Cell Carcinoma. *Cancers* 2021, 13, 5825. https://doi.org/10.3390/cancers 13225825

For this work, I developed and optimized the NGS assay, including the protocols for different tissues (fresh frozen, FFPE), buffy coat blood, and liquid biopsy samples (cfDNA, evDNA) with technical assistance from Antoine Paccard and Alfredo Staffa. I established the bioinformatics methods, analyzed all NGS data included in the study and prepared all figures for publication. Clinical samples and information were provided by Simon Tanguay and Fadi Brimo. Collection of tissue and liquid biopsy samples and nucleic acid isolation was conducted by Narges Abdian and Madeleine Arseneault. Madeleine Arseneault and Alice Nam assisted with validation of somatic variants in IGV. The study was conceived by Yasser Riazalhosseini and Janusz Rak. Isolation of EVs and ddPCR experiments were conducted by Maha Maralani and Laura Montermini, providing data for Figure 1B and 1C. Xenograft models and animal experiments were conducted by Pouria Jandaghi and Brian Meehan, providing data for Figure 1B. Myself, along with Yasser Riazalhosseini wrote the initial draft of the manuscript, which was reviewed and edited by all authors. **Kate Glennon is the first author.**

Chapter 3: Application of Genomic Sequencing to Refine Patient Stratification for Adjuvant Therapy in Renal Cell Carcinoma

Vasudev NS, Scelo G, Glennon KI, Wilson M, Letourneau L, Eveleigh R, Nourbehesht N, Arseneault M, Paccard A, Egevad L, Viksna J, Celms E, Jackson SM, Abedi-Ardekani B, Warren AY, Selby PJ, Trainor S, Kimuli M, Cartledge J, Soomro N, Adeyoju A, Patel PM, Wozniak MB, Holcatova I, Brisuda A, Janout V, Chanudet E, Zaridze D, Moukeria A, Shangina O, Foretova L, Navratilova M, Mates D, Jinga V, Bogdanovic L, Kovacevic B, Cambon-Thomsen A, Bourque G, Brazma A, Tost J, Brennan P, Lathrop M, Riazalhosseini Y, Banks RE. Application of Genomic Sequencing to Refine Patient Stratification for Adjuvant Therapy in Renal Cell Carcinoma. *Clin Cancer Res*. 2023 Apr 3;29(7):1220-1231. doi: 10.1158/1078-0432.CCR-22-1936.

For this study, I contributed to the sample preparation of C3 (N=474 tumor-normal pairs), including library preparation, hybridization capture. Bioinformatic analysis of C3 was conducted by me, with support from Robert Eveleigh. All association and statistical analyses (Kaplan-Meier analysis, Cox-Proportional Hazards models), and figures included in the manuscript were produced by me. Writing of the original manuscript draft was led by Naveen Vasudev, with significant contribution by myself. Naveen Vasudev was responsible for curating and validating the clinical data. Ghislaine Scelo contributed to conceptualizing the study, providing significant resources, and investigating C1 and C2 to identify candidate driver genes. Yasser Riazalhosseini and Rosamonde Banks played key roles in the development of the study, acquiring funding, and supervision. All other authors were involved in manuscript review, and providing resources such as clinical samples. **Kate Glennon, Naveen Vasudev, and Ghislaine Scelo are co-first authors**. Permission from co-authors to include the manuscript in the thesis is included in the Appendix.

Chapter 4: Germline Susceptibility to Renal Cell Carcinoma and Implications for Genetic Screening

Glennon, K.I., Endo, M., Usui, Y., Iwasaki, Y., Breau, R.H., Kapoor, A., Lathrop, M., Tanguay, S., Momozawa, Y., Riazalhosseini, Y.

I performed all experiments and analyses for these experiments. Mikiko Endo assisted with the targeted sequencing experiments, and bioinformatic analysis to identify candidate PVs. Yoshiaki Usui provided support for statistical analyses and experimental design. Yusuke Iwasaki wrote the bioinformatic pipeline used to process NGS data. Yukihide Momozawa and Yasser Riazalhosseini conceived, supervised and financed the study. Rodney Breau, Anil Kapoor, and Mark Lathrop provided resources and were involved in coordinating sample selection and collection. I prepared figures and wrote the manuscript, and all authors were involved in manuscript review. **Kate Glennon is the first author.**

CHAPTER 1. INTRODUCTION

Renal cell carcinoma (RCC) comprises the most common form of kidney cancer. Among RCCs, clear cell subtype (ccRCC) accounts for 75-80% of cases, and is characterized by heterogeneous clinical outcomes and extreme incidental variances[1]. Patients who present with localized tumors undergo curative nephrectomy (radical or partial) to remove the tumor, however, 30-40% of these patients experience relapse or metastasis post-nephrectomy. Due to the indeterminate behaviour of ccRCC and the absence of routine biomarkers, it is difficult to identify patients who are at high-risk for relapse or metastasis and may benefit from adjuvant therapy, and to monitor disease progression. Recent large-scale genomic studies have shed light on the spectrum of genomic abnormalities in ccRCC tumors, and have suggested that somatic genetic heterogeneity between patients may help with prognosis[2]. However, there has not been much success with using genomic information of tumors in precision risk assessment in RCC. Different factors may contribute to this, among which the lack of large-scale studies with matched clinical and genomic data that provide power for association analysis is well acknowledged[2]. Additionally, integrating biological features of the disease such as tumor heterogeneity into the analytical approach would be more likely to generate translational results from association studies.

This thesis aims to investigate in depth associations between the genetic evolution and the repertoire of common somatic genetic alterations in ccRCC, and will examine their potentials for prognosis in ccRCC using the largest to-date single cohort of genomically and clinically annotated samples. Given that ccRCC has specific driver genes, including *VHL*, *PBRM1*, and *BAP1*, which are not included in commercially available targeted sequencing panels, the development of an appropriate genetic screening assay is required to investigate clinically actionable genes relevant to RCC. Furthermore, investigating germline predisposition to RCC among cohorts from large population studies will help to gain insight into the large difference in RCC incidence globally.

1.1 RENAL CELL CARCINOMA

1.1.1 Epidemiology

Globally, there are over 400,000 new RCC diagnoses each year, and 180,000 deaths. While mortality rates have been decreasing in recent decades, incidence rates have been increasing globally since the 1970's[3,4]. Incidence of RCC also varies considerably across the globe, with higher rates in Europe and North America, and lower incidence in Asia and South America[5], though the global variations in incidence are not fully understood. Global variations in mortality also appear to correspond to differences in incidence, ranging from 30-40%, with higher rates of mortality observed in males compared to females[6,7]. Incidence of RCC also increases with age, with approximately half of cases being diagnosed before the age of 65, however for individuals with hereditary RCC syndromes (described in Chapter 1.1.3), diagnosis is often 20 years earlier[6].

Established risk factors for RCC include smoking[8], obesity[9], and hypertension[10,11]. Smoking tobacco is associated with a 30% increased risk of developing renal cancer, while former smokers have a 15% increased risk compared to non-smokers[12]. Obesity is a major risk factor for developing renal cancer, with prospective studies reporting as high as 77% increased risk for obese individuals[13]. A medical history of hypertension or chronic kidney disease is also estimated to increase the risk of developing renal cancer up to 3-fold[14,15]. There is also considerable sex-bias in kidney cancer incidence, with rates in males being double those observed in females[16,17], however the biological factors driving these sex-differences are not yet understood. Aside from lifestyle and medical risk factors, exposure to carcinogens such as trichloroethylene and aristolochic acid is also linked to an increased risk of developing renal cancer[18,19].

1.1.2 Histological Subtypes

Kidney cancers are largely categorized based on the location and cell type within the kidney from which they develop. Urothelial cancers of the kidney (previously called transitional cell carcinomas) originate from urothelial cells within the renal pelvis and ureter, and are treated as bladder cancers. Renal sarcomas develop within the connective tissue surrounding the kidneys, however are rare. Renal cell carcinomas are a heterogeneous group of cancers originating from cells in the nephrons. While the major histological subtype of RCC is clear cell (ccRCC), accounting for 70-75% of diagnoses[20], non-clear cell subtypes include papillary, chromophobe and other more rare subtypes, each having distinct disease etiology, clinical courses, and genomic drivers (**Figure 1**).

RCC subtypes have traditionally been classified based on predominant morphological features; however, as we gain deeper understanding of the molecular features driving each subtype, we are beginning to see the introduction of molecular-driven classifications in renal cancer[20]. Clear cell subtype originates from the epithelial cells of the proximal convoluted tubules, and is named for its 'clear' histological appearance - polygonal cells with a lipid and glycogen-rich clear cytoplasm and small central nuclei[21] (**Figure 1**). ccRCCs are often large, fast-growing tumors, and are also characterized by a variable clinical course and presentation. Compared to papillary and chromophobe subtypes, ccRCCs are more likely to present with metastatic disease, and have poorer cancer specific and overall survival compared to papillary and chromophobe subtypes[22]. The molecular landscape of ccRCC is defined by mutations in the von Hippel-Lindau (*VHL*) gene and the loss of the short arm of chromosome 3 (3p)[23], with somatic alterations also often foun in additional 3p tumor suppressor genes *BAP1*, *PBRM1*, and *SETD2* (described in depth in Chapter 1.2.2).

Among non-clear cell RCCs (nccRCCs), papillary subtype is the most frequent, making up ~15% of all RCC diagnoses[24]. Papillary RCCs originate from epithelial cells of the proximal tubules[24], with cells appearing organized in a spindle-shaped pattern. Until recently, papillary RCCs (pRCCs) were subclassified into Type I and Type II tumors based on morphological features; however, the subcategorization has since been eliminated, recognizing that tumors frequently present with a mixture of Type I and II features[24]. Type I papillary RCCs, also called 'basophilic', are typically lower grade, and detected at earlier stages than Type II pRCCs, thus have a better prognosis. Histologically, they have small cells with scarce, clear cytoplasm and hyperchromatic nuclei, forming a single layer of basophilic cells around the basal membrane[21,24] (**Figure 1**). Type II papillary RCCs are usually detected as high-grade tumors, often with ganglial metastasis present at diagnosis. Histologically, they are described as 'eosiniphilic' as their cells have granular eosinophilic cytoplasm, and prominent nuclei that are often associated with necrotic areas. Genomic characterization of pRCCs has revealed frequent mutations in *MET*, a proto-oncogene, and in tumor suppressor gene *CDKN2A*[20,25].

Chromophobe RCCs (chRCC) are the third most common RCC subtype, representing ~5% of diagnoses[24]. While they are associated with larger tumor size, they are typically early stages (I and II) and less aggressive than ccRCC[26], with metastasis occurring in only 7% of cases. ChRCCs originate from intercalated cells of the distal tubule, and are characterized by large pale cells, reticulated cytoplasm, and perinuclear halos[21]. ChRCCs often have widespread chromosomal losses[27], and frequently harbour mutations in the two well-established tumor suppressor genes, *TP53* and *PTEN*[28]. However, the somatic mutation rate of chRCCs is considerably low compared to most tumors, including ccRCCs[28].

Other rare subtypes of RCC include collecting duct RCC, and medullary RCC which are often extremely aggressive[29]. Collecting duct RCCs are characterized by irregular and infiltrating cells that are arranged in the walls of the collecting ducts[29]. At diagnosis, metastasis is frequently present, and many patients with collecting duct RCCs do not survive two years. Medullary RCCs are more frequently identified in adolescents and young adults, and is aggressive and difficult to treat.

Even with the continuous refinement of the RCC classification system and the identification of additional rare subtypes of RCC, 4-6% of RCCs still cannot be accurately histologically characterized and are labelled as unclassified RCCs. Unclassified RCCs typically have a heterogeneous histology and are frequently high-grade tumors associated with poor prognosis compared to ccRCC[30]. Without distinct classification, these lesions represent a clinical challenging, as it is difficult to stratify risk and develop management strategies.

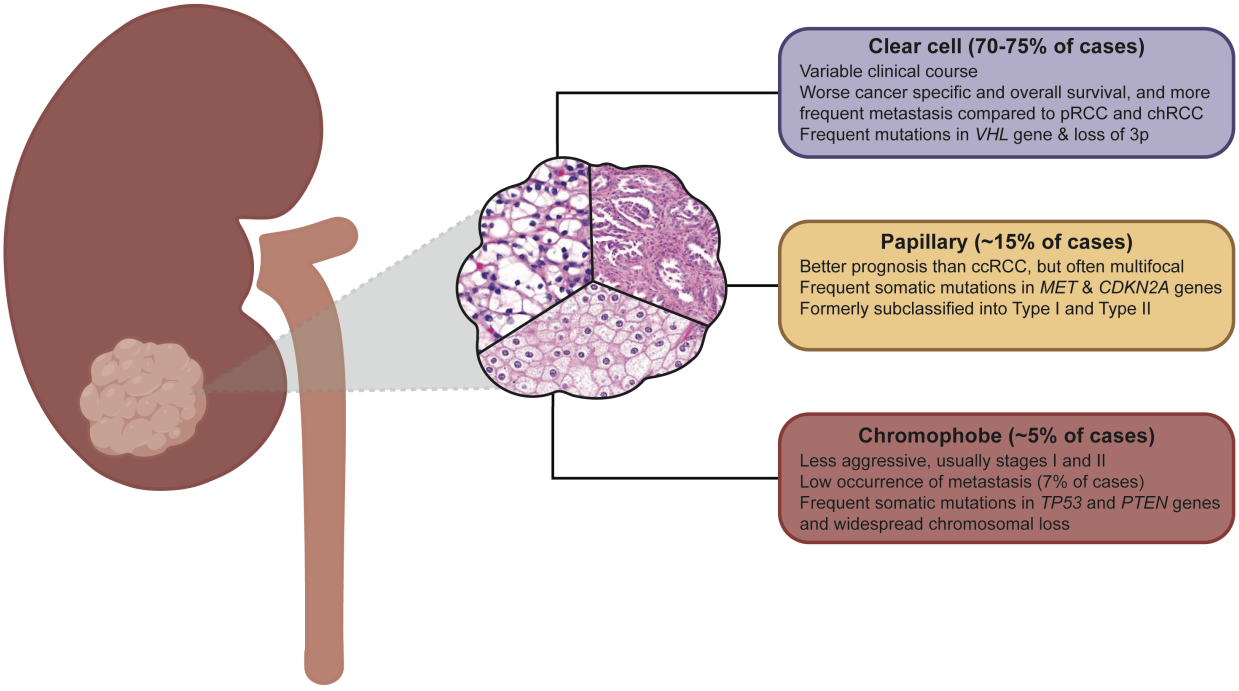


Figure 1. Major histological subtypes of renal cell carcinoma. Clear cell, papillary, and chromophobe subtypes of RCC are primarily distinguished by their histologic architecture, however, also have distinctly different prognoses, genomic drivers, and clinical courses.

1.1.3 Hereditary RCC Syndromes

In comparison to sporadic RCCs, 5% of cases[31] are attributed to familial syndromes with associations to germline mutations in specific genes. Common cancer syndromes that present with renal tumors include von Hippel-Lindau disease (*VHL* gene), hereditary papillary RCC (*MET* gene), hereditary leiomyomatosis (*FH* gene), and Birt-Hogg-Dubé syndrome (*FLCN* gene)[31]. Hereditary kidney cancer syndromes are typically associated with young age of disease onset (<45 years), and multifocal or bilateral tumors, features which guide most genetic screening criteria for renal cancer[32,33].

Von Hippel-Lindau disease (VHL) is an autosomal-dominant ccRCC-predisposing syndrome cause by germline mutations in the *VHL* gene[31,34]. VHL is characterized by the development of a spectrum of tumors in multiple organs, including the kidneys, the brain, spinal cord, and retinas. Individuals with VHL often develop bilateral ccRCC, sometimes having hundreds of kidney lesions. There is also evidence that the type of mutation present in *VHL* is correlated to disease phenotype, with truncating and missense variants being associated with varied likelihood of developing RCC or pheochromocytoma (benign tumors of the adrenal gland). Frameshift and nonsense mutations in *VHL* are highly penetrant for ccRCC, while some missense mutations do not show association to ccRCC[35]. Treatment plans for individuals with VHL disease include routine imaging to monitor the development and growth of lesions, and aggressive surgical approaches.

Hereditary papillary renal cell carcinoma (HPRCC) and Hereditary leiomyomatosis and kidney cell cancer (HLRCC) are autosomal hereditary syndromes associated to papillary subtypes of RCC. HPRCC is a rare syndrome caused by germline mutations in the *MET* gene, and is characterized by the development of bilateral and multifocal pRCCs with Type I (basophilic) features[36,37]. HLRCC, caused by mutations in the *FH* gene, is characterized by the development of kidney cancer along with cutaneous and uterine leiomyomas, and is most frequently associated with papillary RCCs with histologically Type II (eosinophilic) features[38,39]. Patients with HLRCC often have an early age-of-onset, high grade tumors, and an aggressive disease course.

Birt-Hogg-Dubé syndrome (BHD), caused by mutations in the *FLCN* gene, has a variable and often mild presentation, but is characterized by the development of fibrofolliculomas, lung cysts, spontaneous pneumothorax and kidney cancer[40,41]. Patients with BHD can develop various subtypes of renal cancers, however most frequently are hybrid oncocytic tumors that exhibit

chromophobe and oncocytoma features, and are often multifocal. Due to the mild and variable presentation of BHD, it is greatly underdiagnosed, and surveillance and treatment strategies are dependent on the specific symptoms.

Several additional hereditary syndromes are associated with an increased risk of developing RCC (**Table 1**), including Cowden Disease (caused by mutations in the *PTEN* gene), SDH-RCC (mutations in SDH genes), TSC-associated RCC (mutations in *TSC1/2* genes), and MIT family translocation RCC, though they are less common and not often included in genetic screening for renal cancers[31].

1.1.4 Treatment

Though there have been rapid advancements in the diagnosis, treatment, and management of renal cancer in the previous decades, RCC still represents a considerable clinical challenge. Early diagnosis of RCC remains difficult, as there are few symptoms present at low stages; in fact, RCCs are frequently discovered as an incidental finding during other medical procedures[42,43]. After diagnosis, the primary treatment for localized RCC (stages I-III) is surgery, where a partial or radical nephrectomy to remove the tumor is considered curative; however, 30-40% of patients will develop relapse after the surgery[44,45]. For advanced disease, nephrectomy can improve overall survival, and in some cases surgical resection of metastatic sites (metastectomy) is also feasible and effective[44].

Metastatic RCC (mRCC) is resistant to chemo- and radiotherapies, making management of aggressive disease after surgery a challenge. Prior to the 2000s, systemic treatment of metastatic or advanced RCC was limited to traditional immunotherapy approaches (cytokine therapies), which were largely ineffective in managing RCC. Interferon-α (IFNα) had response rates of only 15%, and increased overall survival (OS) by 6 months[46]. Treatment with high doses of interleukin-

2 (IL-2) saw complete remission in a small number of patients (7-10%); however was associated with high toxicity[47]. The past two decades have seen the transition from cytokine approaches to targeted therapies, followed by immunotherapy agents, which have seen higher rates of response and longer progression-free survival (**Figure 2**).

Improved understanding of the biology of ccRCC, including the upregulation of hypoxia-responsive genes (described in Chapter 1.2.1), has enabled the development of therapies targeting molecules that are overexpressed in ccRCC. A major drug target is vascular endothelial growth factor (VEGF) and its receptors (VEGFR), which is produced in high levels in *VHL*-inactivated tumors[48]. The use of tyrosine kinase inhibitors (TKIs) targeting VEGF, sorafenib and sunitinib, has improved prognosis compared to cytokine therapies, however many tumors are unresponsive and tumor re-growth occurs after some time[49]. TKIs with increased sensitivity, along with multi-kinase inhibitors targeting other receptors in addition to VEGFR have also been approved for treatment of advanced and metastatic RCC. Still, complete response is rare, and most tumors eventually acquire resistance[50]. Additional inhibitors targeting the mTOR pathway, rather than VEGF, are also approved for RCC though they have very low response rates[51].

The advent of immune checkpoint inhibitors (ICIs) has largely improved therapeutic approaches for mRCC. Immune-checkpoint molecules including programmed-cell-death protein 1 (PD-1) and cytotoxic T lymphocyte-associated protein 4 (CTLA-4) act to suppress the activity of T cells, resulting in cancer cells becoming immunotolerant[52]. By targeting PD-1, it's ligand (PD-L1) and CTLA-4, ICIs can increase the tumor immune response[53]. Nivolumab, an anti PD-1 agent, was the first immunotherapy drug approved for RCC, and showed improved survival outcomes for patients with advanced RCC who had previously received treatment[54]. Several ICIs,

including pembrolizumab (anti PD-L1) and ipilumab (anti-CTLA-4) have also shown improved survival outcome.

Despite the success of TKIs and ICIs in improving overall survival outcomes for advanced and metastatic RCC, many patients will still develop resistance to individual therapies. Combination therapy regimens have become the standard of care, with strategies involving treatment with both TKIs targeting and ICIs[55]. These combinations have more favourable outcomes than monotherapy approaches, with improved overall survival outcomes and strong responses[55]. Nonetheless, many patients will still develop resistance to current treatment options, and identification of patients who will benefit from adjuvant therapies remains difficult.

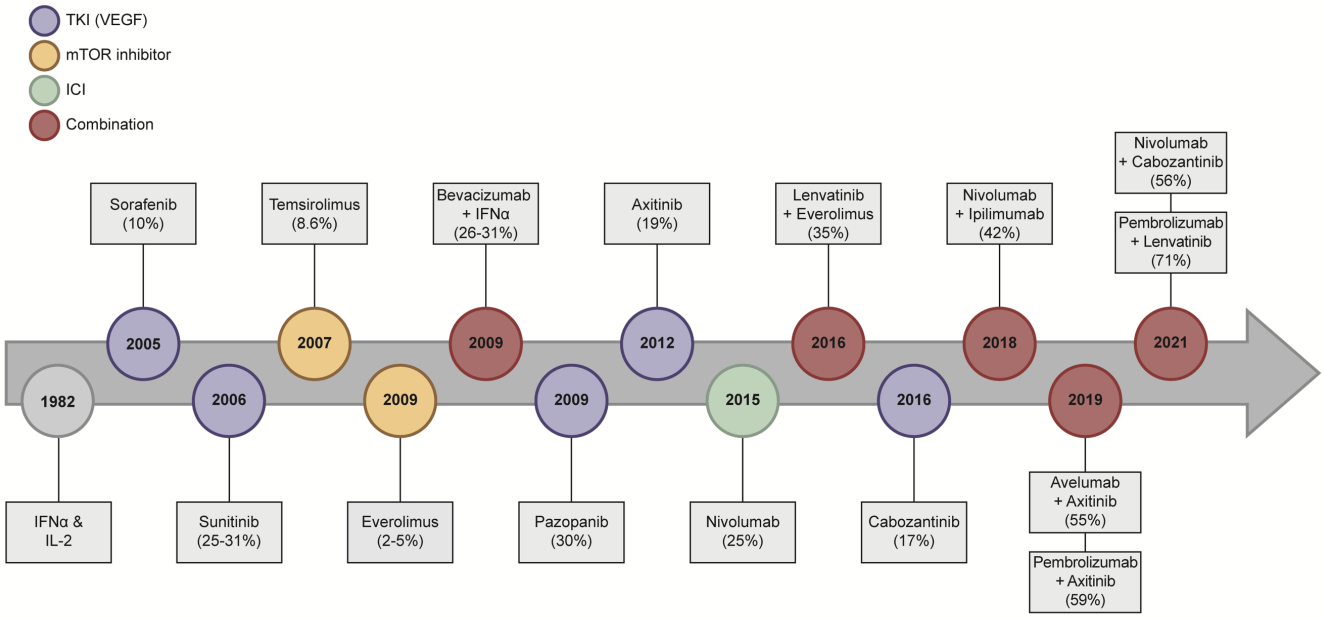


Figure 2. Timeline of key advancements in the treatment of RCC. Response rates are indicated for each agent[55].

1.1.5 Predicting Disease Recurrence

Tumor histology alone is not a good predictor of disease recurrence, and current prognostic algorithms and clinical nomograms that also consider characteristics of the tumor (stage, grade,

and size), and symptoms have all been developed retrospectively[56-62]. Evaluation of existing prognostic tools indicate that they are underperforming and inconsistent in identifying RCC patients at high risk of disease recurrence[63].

For decades, the TNM (tumor-node-metastasis) classification system played the central role in risk-prediction for RCC, however has limited ability to predict individual outcomes. Advanced statistical modeling led to the development of several predictive models, with the goal of improving individualized risk prediction. Various models for predicting overall survival (OS), cancer specific survival (CSS) and recurrence free survival (RFS) have been developed for RCC, each based on various clinical and pathologic variables. Commonly used predictive models that integrate additional clinic-pathological parameters include SSIGN (Stage, Size, Grade, and Necrosis) score[61], Leibovich score[62], Kattan score[58], and UISS (UCLA Integrated Staging System)[64]. Though these models are marginally improved for predicting risk compared to TNM staging system, they demonstrate notable variability over time and prospective validation of these models shows a substantial decrease in their predictive abilities compared to their original estimates[63].

Though we can divide RCC tumors into distinct groups based on histological examinations, this grouping is considerably limited in informing the management and treatment of patients, and when it comes to identifying the mechanisms underlying these differences. Even within the clear cell subtype, there is a considerable tumor heterogeneity, and aggressive tumors are likely a result of distinct molecular processes that can only be identified through characterization of the molecular genetics of RCC.

1.2 MOLECULAR GENETICS OF CCRCC

Efforts to sequence large cohorts of RCCs within the last decade have substantially advanced our understanding of the genomic landscape of RCC and recurrently mutated genes among RCC subtypes. Several genes have been identified as commonly altered in RCC, for both hereditary and sporadic forms (described below and summarized in **Table 1**). Familial RCC syndromes are typically associated with a germline mutation within a single gene responsible for increasing an individuals' risk of developing renal cancer, most of which have been identified and are well-characterized. For sporadic cases, investigations of somatic mutations, those present only within the tumor and not in the germline cells, provide insight into the molecular processes driving tumorigenesis. The landscape of somatic variations in RCC includes patterns of genes harbouring single nucleotide variations (SNVs), small insertions and deletions (indels), and patterns of recurrent copy number alterations (CNAs), large chromosomal losses and amplifications, that have been identified within each RCC subtype[65-67]. Though advances have been made in identifying recurrently mutated genes within RCC subtypes, the immense genetic heterogeneity of RCC tumors, even within a subtype, and the interactions among driver genes are not well understood. Studies of tumor evolution are promising in identifying molecular underpinnings driving RCC progression, particularly for ccRCCs, where the distinct patterns are poorly understood.

Table 1. Major genes commonly implicated or genomically altered in RCC

Gene	Genomic Location	Hereditary disease		Frequency in sporadic disease			Association to prognosis
		Syndrome	Subtype	Clear Cell	Papillary	Chromophobe	
VHL[67,68]	3p25.3	von Hippel Lindau disease	clear cell	>70%	1.1%	1.4%	
PBRM1[20,69,70]	3p21.1			35-45%	4.5%	-	
BAP1[20,71-73]	3p21.1			5-16%	5.6%	-	Metastatic disease
SETD2[74,75]	3p21.31			13-30%	6.4%	2.7%	Higher frequency in metastatic disease
ELOC[76] (also known as *TCEB1)*	8q21.11			1-4%	-	-	
MTOR[20]	1p36.22			6.7%	0.4%	2.7%	
PIK3CA[20]	3q26.32			2.6%	1.5%	1.4%	
KDM5C[20]	Xp11.22			6.9%	1.9%	1.4%	
TSC1[77]	9q34.13	TSC-associated PRCC	angiomyolipoma	0.6%	1.1%	2.7%	
TSC2[77]	16p13.3	TSC-associated PRCC	angiomyolipoma	1.1%	2.3%	2.7%	
MET[25]	7q31.2	HPRC	Type I papillary	0.6%	13-15%	-	
CDKN2A[20,25]	9p21.3			-	13-25%	-	
NF2[20]	22q12.2			0.9%	3.8%	-	Poor prognosis
TP53[28]	17p13.1			2.6%	1.5%	~30%	
PTEN[20,78]	10q23.31	Cowden	often clear cell	4.5%	3.4%	8.1%	
FH[25,79]	1q43	HLRCC	Type II papillary	-	-	-	
FLCN[80]	17p11.2	Birt–Hogg-Dubé	hybrid oncocytomas, chromophobe, papillary, clear cell	-	-	-	Syndrome is very aggressive
SDHB, SDHD[80]	1p36.13; 11q23.1	SDH-RCC	clear cell, chromophobe, oncocytoma	-	-	-	Syndrome is aggressive
TFE3, TFEB, MITF[80-82]	Xp11.23; 6p21.1; 3p13	MiT family translocation RCC	mixed, often overlap of clear cell and papillary	-	-	-	Syndrome is very aggressive

*Not an exhaustive list, there are many more genes that are commonly altered or implicated in RCC, however this list highlights those that are most frequent, associated to hereditary syndromes, or demonstrating prognostic value.

1.2.1 VHL

The defining feature of ccRCCs is inactivation of the von Hippel-Lindau (*VHL*) gene, which is present in >80% of ccRCCs and considered to be the initiating event[83]. *VHL* was initially identified as the gene responsible for VHL disease, a ccRCC predisposing syndrome caused by germline mutations in *VHL*[84]. Subsequently, somatic mutations in *VHL* were identified in sporadic ccRCCs, and *VHL* has since been characterized as a classic two-hit tumor suppressor gene, following Knudson's two-hit model of tumorigenesis[23,85]. Inactivation of *VHL* occurs through the loss of heterozygosity (LOH), along with inactivation of the second allele through intragenic mutations or epigenetic silencing[68].

The product of *VHL*, VHL tumor suppressor protein (pVHL), is a key regulator of cellular response to hypoxia, regulating hypoxia-inducible factors 1α (HIF1α) and 2α (HIF2α)[23,86]. pVHL, along with elongin C and elongin B, both transcription elongation factors, and Cul2 and Rbx1 form the VCB-CR complex[23]. Under normoxic conditions, HIF1α and HIF2α are hydroxylated, allowing recognition by the VCB-CR complex, and subsequent ubiquitylation of HIF-α to mark it for proteasomal degradation. Alternatively, in hypoxic conditions, the VCB-CR complex does not recognize HIF-α due to the lack of hydroxylation. In turn, this enables the stabilization and accumulation of HIF-α and subsequent dimerization with HIFβ. HIFα-HIFβ heterodimers translocate to the nucleus, and induce expression of HIF target genes by binding to hypoxia-response elements. The inactivation of *VHL* results in a pseudo hypoxic state in ccRCCs, and constitutive stabilization of HIFα results in the upregulation of HIF target genes (**Figure 3**). There are over 800 HIF target genes[87], including those that promote cell proliferation, angiogenesis, and glycolysis, whose upregulation contributes to tumorigenesis.

Recently, mutations in the gene encoding elongin C, *ELOC*, have also been described in ccRCC[23]. Mutations in *ELOC* are present in <5% of ccRCCs, and are almost always accompanied by the loss of the second allele on chromosome 8[88]. Notably, mutations in *VHL* and *ELOC* are mutually exclusive, though they both inactivate the VCB-CR complex.

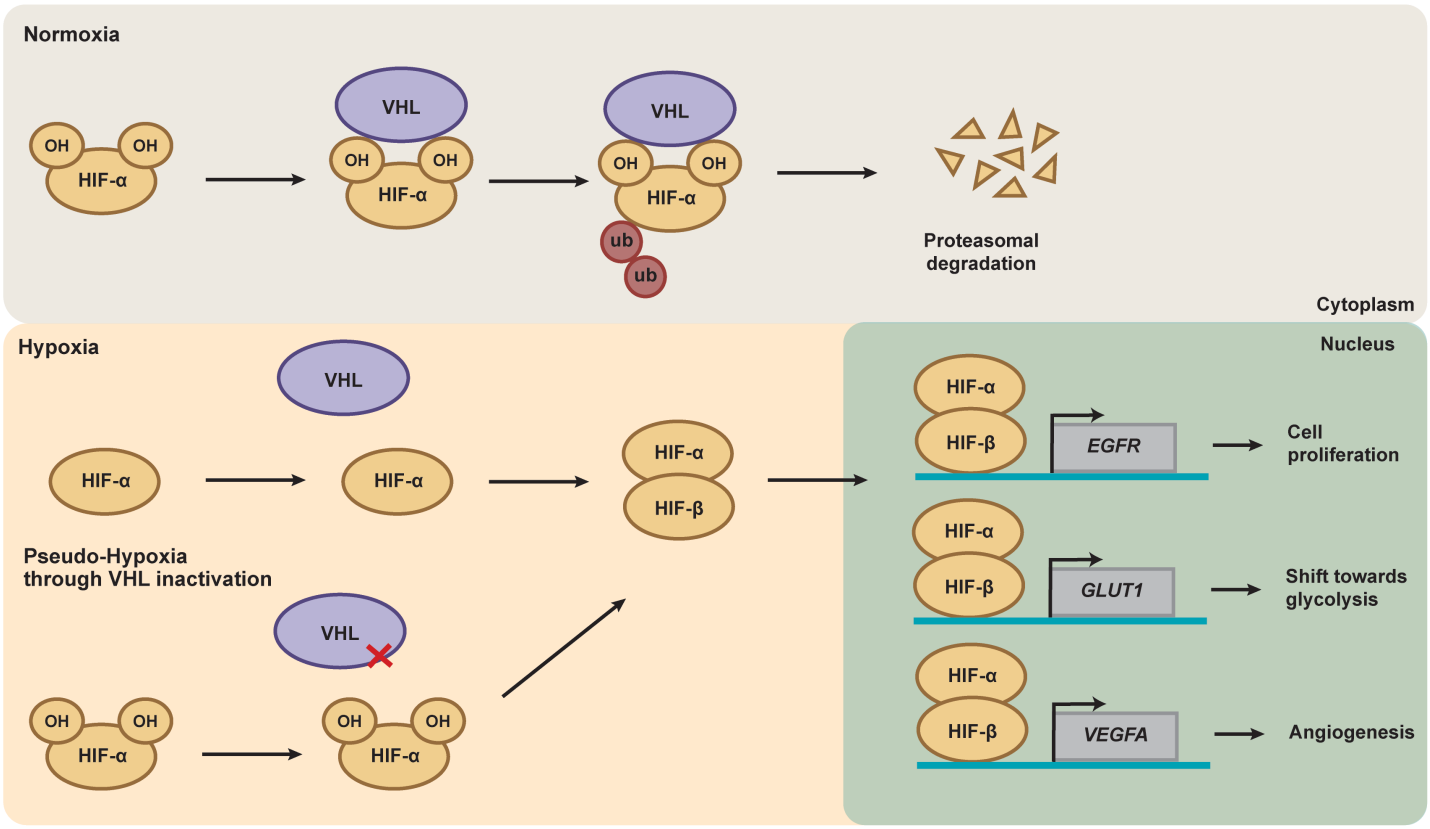


Figure 3. The VHL-HIF pathway under normoxic and hypoxic conditions. Pseudo-hypoxia induced by the loss or inactivation of VHL results in upregulation of HIF target genes that promote tumor growth. (This figure was adapted from Riazalhosseini & Lathrop, 2016[2] and Gossage et al. 2015[23]).

1.2.2 3p Tumor Suppressor Genes: PBRM1, BAP1, SETD2

VHL inactivation alone is not sufficient to cause ccRCC, though it has been the primary focus of ccRCC research in the past decades. With improvements in sequencing technologies, additional driver genes have been identified and implicated in ccRCC. *VHL* is located on 3p chromosome arm (**Figure 4**), which is also home to additional tumor suppressor genes that are commonly implicated in ccRCC – *PBRM1*, *SETD2*, and *BAP1*. The second allele of these genes is often co-deleted along with *VHL*, with the loss of 3p. Along with the two-hit hypothesis for tumor suppressor genes in this region, an additional mutation in the remaining allele can result in reduced or complete loss of protein function, and promote tumorigenesis.

The second most commonly mutated gene after *VHL* is Polybromo 1 (*PBRM1*), mutated in 35-45% of all ccRCCs. *PBRM1* encodes BAF180, a subunit of a SWI/SNF nucleosome remodeling complex that is frequently disrupted in cancer. Mutations in *PBRM1* are typically truncating, resulting in a loss-of-function[89]; however, how the loss of *PBRM1* promotes ccRCC development is not yet fully understood.

Mutations in the *SETD2* gene (encoding SET domain containing protein 2) are found in 10-15% of ccRCCs. SETD2 is a histone H3 lysine 36 (H3K36) trimethyltransferase that plays roles in both transcriptional regulation and DNA damage response (DDR), though the tumor suppressor role of SETD2 in ccRCC is not fully characterized[90]. Recent models demonstrate that loss of SETD2 increases chromatin accessibility and activates enhancers, creating an epigenetic landscape that allows for increased transcriptional output of oncogenic drivers and overall promotes metastasis[90]. Likewise, tumors with mutations in *SETD2* are associated with poorer disease-free survival and increased risk of disease recurrence in comparison to tumors without

SETD2 mutations[20]. Interestingly, mutations in *SETD2* are often observed as co-occurring with *PBRM1* mutations; however, the molecular basis and whether they cooperate is unknown.

The *BAP1* gene (BRCA1 associated protein-1) is mutated in 10-15% of sporadic ccRCCs, and germline mutations in *BAP1* have also been associated with familial ccRCC[91]. Like *SETD2*, *BAP1* mutations are also associated to high stage tumors, and worse prognosis, which suggests a role in disease progression[92,93]. BAP1 is a de-ubiquitinating enzyme involved in regulating several cellular processes, including DNA repair, replication, apoptosis, and maintaining genome stability[94]. However, the mechanisms explaining how *BAP1* mutations specifically drive renal carcinogenesis are not well understood.

While co-occurrence of *PBRM1* and *SETD2* mutations is common[95], possibly indicating their cooperation in tumorigenesis or progression, mutations of *BAP1* and *PBRM1* are mutually exclusive[95,96]. Additionally, mutations in each gene are associated to RCC with differing pathological features, including gene expression profiles, and patient outcomes. Though uncommon, tumors that do harbour co-occurring *BAP1* and *PBRM1* mutations are extremely aggressive and are associated with poor outcome[97].

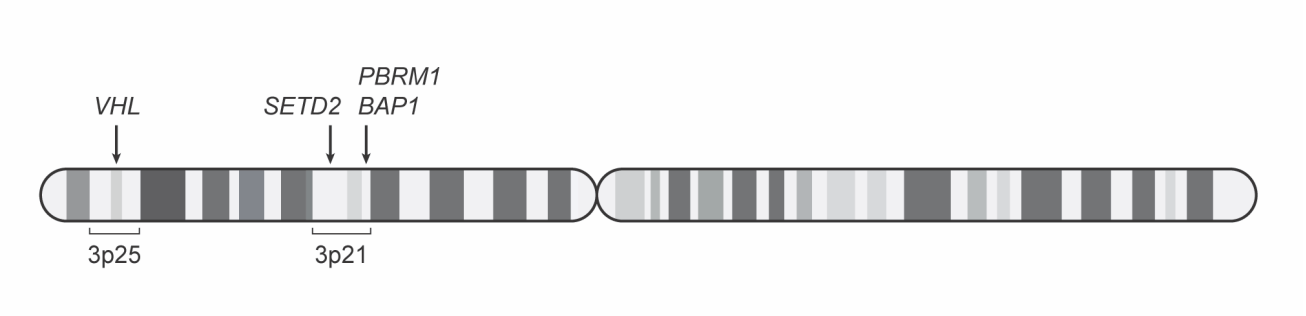


Figure 4. Tumor suppressor genes located on chromosome 3p in the vicinity of *VHL*. The loss of chromosome 3p is a defining feature of ccRCC, along with the inactivation of *VHL* (located at 3p25). Additional tumor suppressor genes *SETD2*, *BAP1*, and *PBRM1* (all located in the 3p21

region) in this region are also frequent targets for inactivation. (This figure was adapted from Dizman et al, 2020[48]).

1.2.3 PI3K/AKT/mTOR Pathway Genes

The PI3K/AKT/mTOR signalling pathway is also implicated in ccRCC, with ~28% of tumors harbouring alterations in genes encoding proteins involved in the pathway[68,76]. Playing an important role in cell survival and growth, dysregulation of this pathway is frequent in cancer, and presents a promising potential therapeutic target. mTOR inhibitor drugs have shown some success in managing metastatic and locally advanced RCC, however their efficacy is limited[98]. The somatic mutation frequency of individual genes involved in PI3K/AKT/mTOR signaling has not reached statistical significance in many studies, though the genes within the pathway are affected by somatic mutations in 28% of ccRCCs[68,76]. In ccRCC, somatic mutations have been identified in *PTEN* (1-5%), *PIK3CA* (2-5%), *TSC1*/*TSC2* (4%), and *MTOR* (5%) genes[76]. While *PTEN*, *TSC1* and *TSC2* are known tumor suppressor genes, the others are suggested to act as oncogenes, and demonstrate mutual exclusivity[76].

Interplay between VHL and PI3K/AKT/mTOR pathways may also play a role in ccRCC, as they are linked by a negative feedback loop[99,100] (**Figure 5**). In response to hypoxia, mTORC1, one of the two complexes nucleated by mTOR, is inhibited by the REDD1 protein. With *VHL* disruption, REDD1 is upregulated[99], which leads to a downregulation of mTORC1. Interestingly, overexpression of genes within the mTOR signaling pathway is observed in 60-85% of ccRCCs, indicating that pathway is activated[20,100], which may be due to mutations in additional pathway genes that are required for REDD1 signaling, such as *TSC1* and *PTEN*.

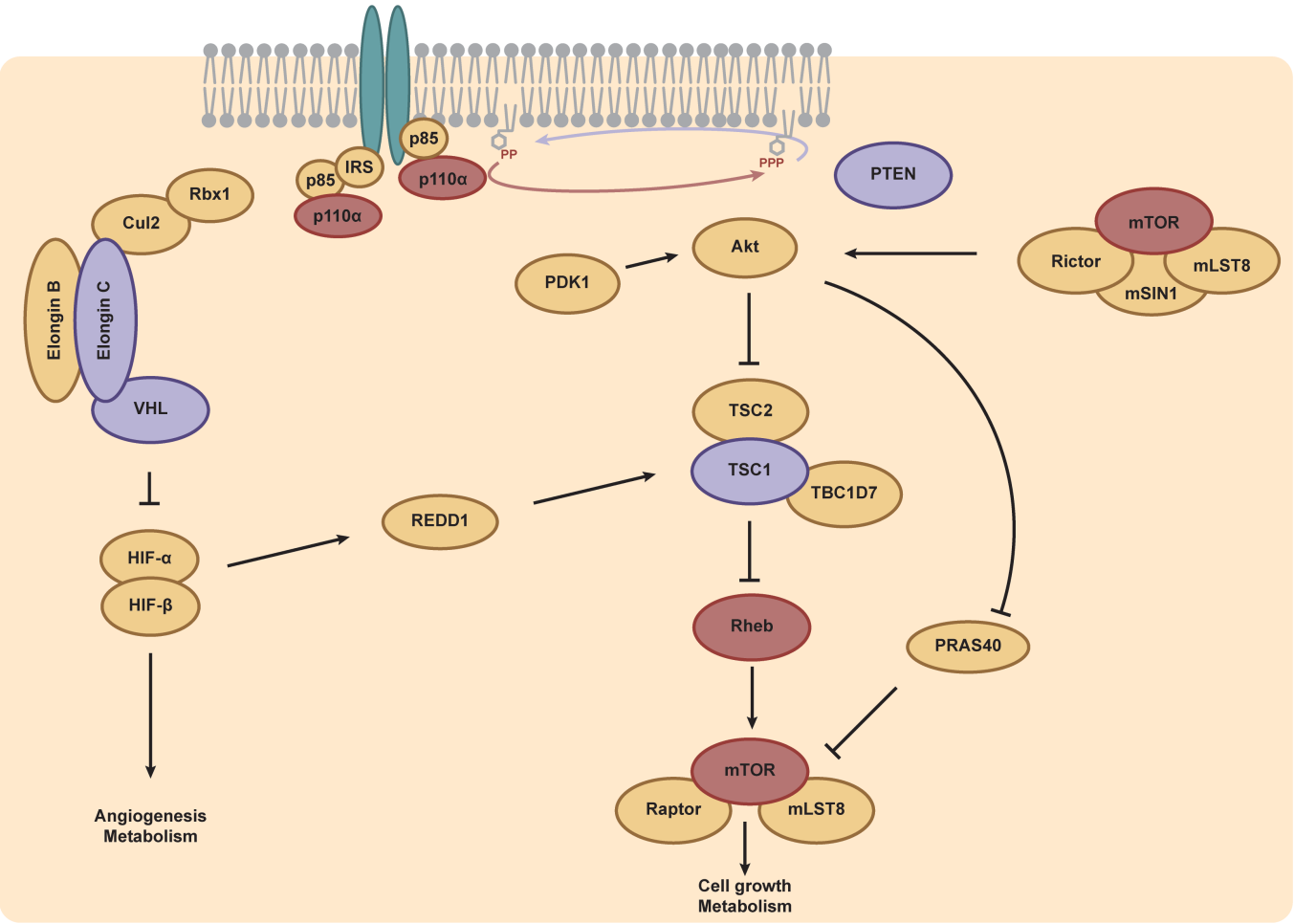


Figure 5. Interplay between VHL and MTOR pathways. The PI3K/AKT/mTOR pathway plays a major role in cell growth and proliferation and is frequently disrupted in cancer. Activating mutations in *PI3KCA* (encoding p110α) and deactivating mutations in *PTEN* increase PIP3 levels, leading to recruitment of AKT. Phosphorylation of TSC2 by AKT releases inhibition of Rheb by the TSC1/2 complex, which in turn activates MTORC1. Deactivating mutations in *TSC1* or *TSC2* can also act to increase the activity of the pathway. In response to hypoxia or pseudo-hypoxia due to *VHL* inactivation (described in Figure 3), REDD1 is transcriptionally activated by HIF-1 and HIF-2. REDD1 inhibits MTORC1 via the TSC1/2 complex. (This figure was adapted from Brugarolas, 2014[100]).

1.2.4 Genetic Evolution of ccRCC

Large scale genomic studies have highlighted the intra-tumoral heterogeneity of ccRCC, however only recently have they begun to identify patterns in the ordering of driver mutations and conserved features of tumor. Initial studies propose a model of ccRCC development in which the initiation event is the loss of chromosome arm 3p, followed by an intragenic mutation of *VHL*, which inactivates the remaining allele. This leaves other genes within 3p, including *PBRM1* and *SETD2* vulnerable to inactivation (**Figure 6)**. One mechanism driving the frequent loss of chromosome 3p for an estimated 30-40% of ccRCCs is a chromothripsis event also involving chromosome 5q; breakage in multiple chromosomes simultaneously, followed by attempted repair, results in the loss of one copy of 3p and the gain of an extra copy of 5q[101,102]. There is also evidence of chromothripsis events leading to concurrent loss of 3p and 6q, and observed unbalanced translocations with other chromosomes, along with cases with loss of the entire chromosome 3[102].

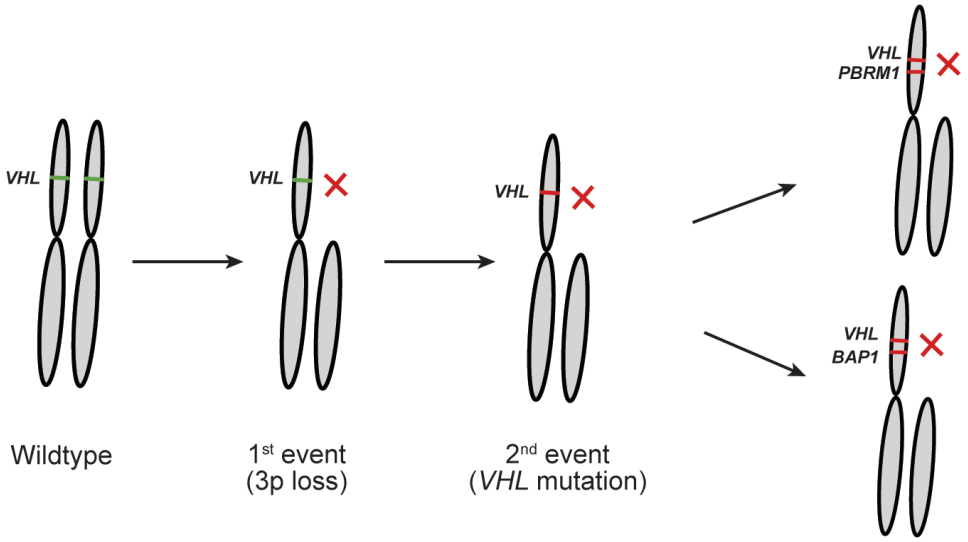


Figure 6. Model of ccRCC initiation. Following an initial loss of chromosome 3p (often due to a chromothripsis event) an intragenic mutation of *VHL* inactivates the remaining allele. Additional 3p tumor suppressor genes, such as *PBRM1* and *BAP1* are left susceptible to inactivation. (This figure was adapted Brugarolas, 2014[100]).

Though this model describes the initiation of ccRCC, it does not explain the extreme visible intratumoral heterogeneity. ccRCCs are known to have extreme subclonal diversity, with 73-75% of somatic mutations being subclonal[103]. Studies investigating the somatic landscape of RCC have also revealed branched evolutionary tumor growth, with key tumor-suppressor genes having distinct and separate inactivating mutations within a single tumor[103,104]. The phylogenetic reconstruction of tumors has helped to resolve the subclonal architecture of ccRCCs, identifying alterations in *VHL* as the only consistently truncal mutations[103]. In some cases, *PBRM1* inactivation occurs early, but all other driver mutations were subclonal[103].

1.2.5 Genetic-Based Risk Assessment

Increased knowledge of genes frequently mutated in ccRCC has helped us to gain insight into the molecular processes driving tumorigenesis, however analyses of individual genes has not shown to be successful in identifying robust markers of prognosis for RCC. Mutations in certain genes have demonstrated association to differing outcomes, such as mutations in *BAP1* and *PBRM1* being associated with differing outcomes (described in Chapter 1.2.2), however there is no single genomic marker of prognosis or response to treatment. There is also extensive intratumoral heterogeneity within ccRCC tumors, of which the patterns have not yet been fully characterized. Intratumoral heterogeneity, and the corresponding heterogeneous protein function may also interfere with targeted therapies and resistance[104]. Currently, there are no prognostic biomarkers for RCC that consider the genetic profile of the tumor, highlighting the need for large-scale genomic studies investigating genetic patterns underlying RCC subtypes and how they can inform outcome.

More recently, systematic studies focusing on the subclonal diversity of ccRCC have identified recurrent patterns of driver event ordering, co-occurrence, and mutual exclusivity. From

these patterns, ccRCC tumors could be grouped into 7 distinct evolutionary subtypes, each with correlated phenotypes[105]. These subtypes include tumors with multiple clonal drivers, *VHL*-monodrivers, those that are BAP1-driven, tumors with *PBRM1* mutations preceding *PI3K* pathway mutations, *SETD2*, and driver SCNA events (**Table 2**). These proposed evolutionary trajectories for ccRCC are linked to distinct clinical outcomes; presenting a novel and promising opportunity for genomic-based personalized management of ccRCC.

Table 2. Proposed evolutionary subtypes of ccRCC*

Evolutionary Subtype	Mutational characteristics	Associations
Multiple clonal drivers	*VHL*, followed by mutations in ≥2 of *BAP1*, *PBRM1*, *SETD2*, or *PTEN*	High levels of wGII, late stage disease, smaller number of clones, high level of %MVI/%G4/%Ki67
BAP1 driven	*VHL*, *BAP1* as lone clonal drivers	Elevated wGII, high tumor grade, small number of clones
VHL wildtype	Lack of driver mutations	Elevated wGII, frequent presence of sarcomatoid differentiation, high Ki67%
PBRM1 → *SETD2* driven	Early *PBRM1* mutation, followed by *SETD2* mutations	Advanced stage, high ITH, highly branched trees (many clones), frequent parallel evolution events, advanced disease stage
PBRM1 → *PI3K* driven	Early *PBRM1* mutation, mutational activation of PI3K/AKT/mTOR pathway	Lower grade tumors
PBRM1 → SCNA driven	Early *PBRM1* mutation, subclonal SCNAs	Lower grade tumors
VHL mono driver	*VHL*, no additional driver mutations	Low wGII, early stage

*Summarized from Turajlic, et al. 2018[105]
wGII: weighted genome instability index; estimates the fraction of the genome affected by somatic CNAs
ITH: intratumor heterogeneity, measured as the ratio of subclonal to clonal drivers
%MVI: Percentage of tumors with microvascular invasion
%G4: Percentage of tumors that are Fuhrman grade 4
%Ki67: Percentage of cells staining positive for Ki67

1.3 NGS TECHNOLOGIES IN CANCER GENOMICS

An aspect that has revolutionized the molecular characterization of many cancers, including RCC, is the advances in next-generation sequencing (NGS) technologies over recent decades which have allowed for massively parallel sequencing of tumors. NGS technologies have become less expensive, while also becoming faster, more sensitive, and higher throughput. This has improved the ability to generate large genomic datasets for individual cancers, enabling the identification of recurrent mutations, and increasing the statistical power for investigations of associations to prognosis. The ability to sequence at high depths of coverage also enables the sensitive detection of mutations present at low allele frequencies, facilitating deeper investigations into the subclonal architecture of tumors. Simultaneously, advances in molecular biology approaches including DNA isolation, library preparation, and target enrichment methods, have enabled tailored NGS workflows for 'difficult' sample types, such as Formalin-Fixed Paraffin-Embedded (FFPE) tissue and liquid biopsies. Bioinformatics tools are also becoming increasingly more powerful for analyzing NGS data, and computational approaches more sensitive for identifying genomic aberrations. With NGS technologies becoming more powerful, less expensive, it is becoming increasingly feasible to bring genome-based personalized medicine to clinical practice.

Traditionally, whole genome (WGS) and whole exome sequencing (WES) approaches have been applied for the analysis of tumor DNA. Comparing WGS or WES of matched tumor and normal samples from an individual patient allows for simultaneous screening of germline and somatic CNAs, SNVs, and small indels, supporting in depth genetic profiling of each individual tumor sample. While these allow for a more exploratory approach for gene discovery, recently targeted sequencing panels have been developed to allow for deeper sequencing at a lower costs

by focusing on specific genes of interest. Targeted sequencing panels offer the advantage of being able to obtain a high depth of coverage across genes of interest, without 'wasting' sequencing on other genomic regions. At the same time, targeted sequencing approaches are more affordable and readily applied to the clinic.

As of yet, there is no convenient NGS approach tailored to renal cancer. The sensitivity of a targeted approach is essential to capture the considerable genetic heterogeneity of RCC, where subclonal mutations may be present at extremely low mutation frequencies. However, commercially available gene panels for cancer are missing key RCC genes.

1.4 LIQUID BIOPSY

Liquid biopsies offer a promising opportunity for identifying biomarkers in cancer, as they are minimally invasive compared to tissue biopsies, provide real time access to diagnostic/actionable mutations, and can be a valuable resource for management of patients with cancer. There are considerable limitations to tissue biopsy in cancer - in addition to difficulties with obtaining tissue biopsies, or tumors that are inaccessible, a single biopsy is not informative of the entire tumor. The extensive intratumoral heterogeneity of RCC dictates that not all subclones are present within a single region from a tumor[104], which means that characterizing the entire genetic landscape of the tumor may not be possible from a single tissue biopsy. Alternatively, liquid biopsies may give a better representation of a tumors' heterogeneity, as they are not limited to one spatial region, and are capable of capturing information from the entire tumor[106,107].

Various liquid biopsy components have demonstrated diagnostic, prognostic, and potentially therapeutic utility for solid tumors. Studies of biological fluids have identified circulating tumor cells (CTCs), cell free DNA (cfDNA) and RNA (cfRNA), extracellular vesicles (EVs) including exosomes, and additional tumor derived metabolites and proteins that are

informative for cancer. Several of these analytes have also been evaluated in the context of RCC, however each has their respective challenges hindering implementation in the clinic (summarized in **Table 3**). In particular, circulating tumor DNA (ctDNA) found within cfDNA, has allowed for the identification of tumor-associated genetic alterations in the blood. Analysis of ctDNA in RCC is not yet implemented in the clinic, as analyses are lacking in sensitivity, however technical advancements in NGS are increasing the sensitivity of ctDNA analysis and there has been success in other cancer types, representing a promising future for RCC.

Table 3. Liquid biopsy components demonstrating clinical utility for renal cancer and their challenges

Analyte	Source	Application	Analytical Approaches	Challenges
Cell free DNA[108-110]	Plasma Urine	Prognostic and Diagnostic	Assays consider several characteristics of cfDNA, including: quantity, promoter site methylation, somatic mutations and fragmentation NGS Assays or RT-qPCR to detect genomic alterations qMSP assays to evaluate methylation	Quantification and fragmentation assays have low sensitivity (high false-negative rate), likely due to the low abundance of ctDNA released by RCCs cfDNA methylation assays are costly and timely NGS assays interrogating genomic alterations have low sensitivity, require matched tumor and normal biopsies for validation Heterogeneity of RCC subtypes makes biomarker selection challenging
Circulating tumor cells[108,111-113]	Whole-blood	Diagnostic	Antibody-based assays (targeting EPCAM or CK) Physical property-based assays (cell size, shape, protein expression)	Difficulty in distinguishing CTCs from leukocytes Assays have low diagnostic sensitivity due to low abundance of CTCs in early stages and low expression of markers in RCC used to identify CTCs
Extracellular vesicles[108,114,115]	Plasma Serum Urine	Diagnostic	Precipitation enrichment via ultracentrifugation Enrichment via size filtration (typically for urine) Total EV cargo evaluation as biomarkers (RNA, proteins, DNA, lipids) Identification of protein markers via mass spectrometry (proteome, lipidome), and validation through western blot or ELISA Identification of nucleotide markers via NGS (mRNA, miRNA, DNA) and validation through RT-qPCR	Lack of standardization for enrichment protocols and analyses hinders assay reproducibility Distinction between benign and cancer EVs Small size of EVs limits number of simultaneously bound antibodies for enrichment Contamination with non EV-encapsulated molecules

1.4.1 Circulating Tumor DNA

Cell free DNA, originally identified as 'non cell bound DNA' in the blood, represents fragments of DNA that have been released by cells into the blood and other biological fluids such as saliva, urine, and cerebrospinal fluid. Tumor-derived DNA (ctDNA) is also released into the circulation, and makes up a small fraction of all cfDNA. The exact mechanisms responsible for releasing ctDNA into the blood are not completely understood, however there is likely a combination of passive and active mechanisms. Passive release of cfDNA likely occurs during processes of cellular destruction – apoptosis and necrosis[116]. Evidence of spontaneous, active release of DNA into circulation has been reported in cell lines, and through secretion from extracellular vesicles[116,117]. It is hypothesized that active release of ctDNA by cancer cells may serve a purpose to promote metastasis at distant sites, affecting the transformation of susceptible cells[118,119].

ctDNA can provide valuable information about the primary tumor, and has been established as a promising diagnostic marker for molecular profiling of tumors with utility for early diagnosis, and a role in monitoring treatment response in real-time. Initial studies of ctDNA in several cancer types have demonstrated its potential for monitoring response to treatment, detecting minimal residual disease (MRD)[120], and monitoring the emergence of drug resistance. While tissue biopsies are limited to the spatial region of the tumor that was captured, ctDNA is believed to be representative of the entire tumor, making it a better representation of the subclonal evolution occurring in the tumor. Serial liquid biopsies have demonstrated that they can be more informative than tissue biopsies for evaluating global tumor heterogeneity[121]. Additionally, as the half-life of ctDNA in the blood is under 2 hours[122], it represents a real-time representation of the

tumor profile. Likewise, the burden of ctDNA is increased with tumor burden; late stage tumors are associated with higher loads of ctDNA in the plasma[116].

1.4.2 Clinical Utility of Liquid Biopsies in Cancer

Various molecular features of ctDNA, including methylation patterns, fragment size, and somatic mutations, have demonstrated utility for longitudinal clinical surveillance to predict risk of disease recurrence, or detect relapse and resistance to therapies. Quantifying the amount of ctDNA and tracking known tumor-associated genetic aberrations are informative for monitoring minimal residual disease (MRD)[120]. Studies of longitudinal surveillance demonstrate the association between increasing amounts of ctDNA and disease recurrence[123]. Additionally, longitudinal surveillance of somatic aberrations can track tumor and clonal evolution over time. For patients undergoing drug treatments, the emergence of tumor-associated genomic aberrations can indicate the emergence of drug resistant cancer cells, which can be detected by ctDNA analysis months before when conventional methods are able to detect it[124]. On the other hand, in cases where a tumor biopsy is not available, somatic mutations in relevant genes identified within ctDNA can inform treatment selection, prognosis, and risk of disease recurrence through molecular profiling[125,126]. There is also evidence that ctDNA analysis is becoming effectively sensitive for screening and early diagnosis for some cancer types[127-129].

Common methods of analyzing ctDNA include digital PCR, genome-wide NGS approaches, and targeted NGS approaches. Digital droplet PCR (ddPCR) allows for the highly sensitive, targeted amplification of a mutant allele, and is a valuable tool for monitoring ctDNA when prior knowledge of the genetic aberration of interest[130]. Analysis of shallow WGS (sWGS) from ctDNA is capable of detecting copy number aberrations of the tumor, and is often used for estimating the proportion of cfDNA that is tumor-derived. More recently, analysis of ctDNA

fragment sizes from sWGS has identified fragment size differences between healthy cfDNA and ctDNA that may be informative[131]. The study of fragmentation patterns ('fragmentomics') is a newly active area of liquid biopsy research that identified patterns indicative of tissue-of-origin, and offers potential for cancer screening[132]. Lastly, targeted NGS approaches involving commercial or custom targeted sequencing panels are capable of detecting SNVs and small indels at low allele frequencies, beneficial for investigating the somatic landscape of ctDNA.

1.4.3 Challenges for ctDNA in RCC

RCC presents a particular challenge for ctDNA analysis due to a combination of low amounts of ctDNA[106] and low mutational burden[109]. Some studies have indicated that less than 50% of patients with RCC have detectable ctDNA[106,133]. Additionally, analysis of cfDNA fragmentation features showed that cfDNA fragment sizes in the plasma of patients with late stage RCC were much longer, similar to healthy individuals, than in other late-stage cancers[131]. The relatively low tumor mutational burden (TMB) of RCC also means that there are few somatic mutations to identify within the already low amounts of ctDNA. RCCs harbour ~1.1 mutation/Mb[131], meaning that many tumor-derived fragments will not have mutations, and be indistinguishable from healthy cfDNA fragments. Some studies have also demonstrated a discordance between tumor and ctDNA[120], however this could be due to low sensitivity of NGS techniques used of analysis, long timing between the tumor and liquid biopsies, or clonal and spatial heterogeneity that was not captured within the tumor biopsy.

1.5 RATIONALE, HYPOTHESIS AND OBJECTIVES

Hypothesis: We hypothesize that analyzing the mutational status of RCC-relevant genes in large and clinically annotated cohorts is required to establish genomic markers of prognosis. Furthermore, establishing liquid biopsy assays capable of capturing genomic biomarkers will pave the way for clinical application of prognostic markers in RCC. Lastly, a deeper understanding of the germline predisposition to sporadic RCC will help to identify at-risk patients and improve precision prevention strategies in RCC.

I address this hypothesis through the following three aims:

Aim 1: Developing and optimizing a renal-cancer-appropriate NGS assay and tailored bioinformatic analysis solutions for tumor and liquid biopsies.

Aim 2: Investigating the genetic evolution of renal cell carcinoma and the clinical consequences.

Aim 3: Investigating germline predisposition to RCC, and implications for genetic screening.

PREFACE TO CHAPTER 2.

In Chapter 1, we highlighted the need to develop better biomarkers for risk stratification for patients with ccRCC, as it remains difficult to identify patients who are at high-risk of developing relapse or metastasis after curative surgery. Increasing our understanding of the genetic heterogeneity of ccRCCs, by investigating patterns of recurrently mutated genes and their associations to prognosis, may help to identify genomic markers of prognosis. To investigate the genetic landscape of ccRCCs, large-scale studies are needed to fully capture their considerable inter- and intra-tumoral heterogeneity. However, there is a lack of sequencing assays appropriate for renal cancers, as many of the commercial panels used in research are missing key RCC genes, and whole genome or exome approaches are expensive and may not be feasible at such large scale. Limiting investigations to smaller panels of genes also enables more practical application in the clinic. Thus, the first aim of this thesis was to develop a targeted sequencing approach, specific to renal cancer, to enable deeper investigations into the somatic landscape of RCC.

Chapter 2 consists of a manuscript describing the development and optimization of a renal cancer appropriate targeted sequencing panel for both tissue and liquid biopsies. The panel contains genes relevant to clear cell and non-clear cell subtypes that may have diagnostic and prognostic value. In this chapter, we demonstrate the assays capability to detect potentially actionable mutations within matched tissue and liquid biopsies from patients with ccRCC. Notably, we evaluate the ability of both cfDNA and evDNA (exosomal DNA) to capture information from the primary tumor. This NGS assay provides a resource to facilitate genomic investigations of RCC tumors, as well as a platform for establishing robust liquid biopsy strategies for RCC.

CHAPTER 2. RATIONAL DEVELOPMENT OF LIQUID BIOPSY ANALYSIS IN RENAL CELL CARCINOMA

Kate I. Glennon[1,2], Mahafarin Maralani[1,2,3], Narges Abdian[3], Antoine Paccard[1], Laura Montermini[3], Alice Jisoo Nam[1], Madeleine Arseneault[1], Alfredo Staffa[1], Pouria Jandaghi[1,2], Brian Meehan[3], Fadi Brimo[4], Simon Tanguay[5], Janusz Rak[3] and Yasser Riazalhosseini[1,2,*]

[1] McGill Genome Centre, McGill University, 740 Doctor Penfield Avenue, Montreal, QC H3A 0G1, Canada; kate.glennon@mail.mcgill.ca (K.I.G.); mah_afarin@hotmail.com (M.M.); antoine.paccard@mcgill.ca (A.P.); ji.nam@mail.mcgill.ca (A.J.N.); madeleine.arseneault@mcgill.ca (M.A.); astaffa@gmail.com (A.S.); pouria.jandaghi@mail.mcgill.ca (P.J.)

[2] Department of Human Genetics, McGill University, 1205 Doctor Penfield Avenue, Montreal, QC H3A 1B1, Canada

[3] The Research Institute of the McGill University Health Centre, Montreal, QC H4A 3J1, Canada; nargesabdian@gmail.com (N.A.); laura-m@sympatico.ca (L.M.); bmeehan82@hotmail.com (B.M.); janusz.rak@mcgill.ca (J.R.)

[4] Department of Pathology, McGill University, Montreal, QC H3A 2B4, Canada; fadi.brimo@mcgill.ca

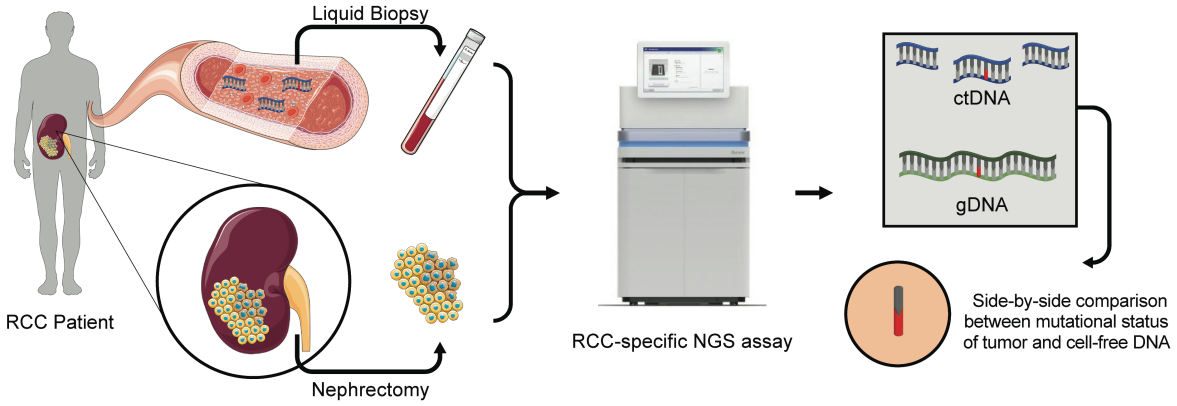
[5] Division of Urology, McGill University, Montreal, QC H4A 3J1, Canada; simon.tanguay@mcgill.ca

[*] Correspondence: yasser.riazalhosseini@mcgill.ca

Originally published in Cancers:
Glennon, K.I.; Maralani, M.; Abdian, N.; Paccard, A.; Montermini, L.; Nam, A.J.; Arseneault, M.; Staffa, A.; Jandaghi, P.; Meehan, B.; et al. Rational Development of Liquid Biopsy Analysis in Renal Cell Carcinoma. *Cancers* 2021, 13, 5825. https://doi.org/10.3390/cancers 13225825

2.1 SIMPLE SUMMARY

Among patients affected by renal cell carcinoma (RCC), the most common type of kidney cancer, it remains difficult to identify those who are at high risk for relapse or metastasis. This is in part due to the absence of reliable clinical biomarkers and robust methods to capture them. The aim of our study was to develop an improved assay to capture prognostic genomic biomarkers in circulating tumor DNA (ctDNA) in RCC. For this purpose, we first established a next generation sequencing (NGS) assay, targeting genes that are tailored for RCC and that are largely excluded from commercially available assays. Next, we showed the reliable performance of this assay to detect prognostic gene mutations in tumor DNA isolated from plasma, and from extracellular vesicles. Thus, our study provides a resource to facilitate ctDNA analysis for precision medicine in RCC.



Graphical Abstract

2.2 ABSTRACT

Renal cell carcinoma (RCC) is known for its variable clinical behavior and outcome, including heterogeneity in developing relapse or metastasis. Recent data highlighted the potential of somatic mutations as promising biomarkers for risk stratification in RCC. Likewise, the analysis of circulating tumor DNA (ctDNA) for such informative somatic mutations (liquid biopsy) is considered an important advance for precision oncology in RCC, allowing to monitor molecular disease evolution in real time. However, our knowledge about the utility of ctDNA analysis in RCC is limited, in part due to the lack of RCC-appropriate assays for ctDNA analysis. Here, by interrogating different blood compartments in xenograft models, we identified plasma cell-free (cf) DNA and extracellular vesicles (ev) DNA enriched for RCC-associated ctDNA. Additionally, we developed sensitive targeted sequencing and bioinformatics workflows capable of detecting somatic mutations in RCC-relevant genes with allele frequencies ≥ 0.5%. Applying this assay to patient-matched tumor and liquid biopsies, we captured tumor mutations in cf- and ev-DNA fractions isolated from the blood, highlighting the potentials of both fractions for ctDNA analysis. Overall, our study presents an RCC-appropriate sequencing assay and workflow for ctDNA analysis and provides a proof of principle as to the feasibility of detecting tumor-specific mutations in liquid biopsy in RCC patients.

2.3 INTRODUCTION

Mutational analysis of plasma circulating tumor DNA (ctDNA) for precision oncology has attracted considerable attention over the past decades [**1**,**2**,**3**]. This approach, often referred to as 'liquid biopsy', is of interest due to the fact that it can potentially offer a real time access to diagnostic and actionable mutations regardless of the accessibility and number of lesions present in a patient [**2**]. Therefore, liquid biopsy analysis is believed to be a powerful resource in the

management of patients with cancer [**1**]. Whereas ctDNA analysis is producing promising results in colorectal and other cancers [**1**,**2**,**3**], there has not been much success with liquid biopsy-based analysis of tumor mutations in renal cell carcinoma (RCC), the most common form of kidney cancers, in spite of the hypervascular nature of these tumors. A plausible reason for this is the absence of RCC-relevant genes in commercially available ctDNA analysis assays, which have been used in the previous studies. For example, two recent large-scale (>200 cases) liquid biopsy studies in RCC [**4**,**5**] have deployed assays that do not include commonly mutated genes in RCC, including PBRM1, SETD2, BAP1, and KDM5C, whose mutations are associated with clinical outcomes [**6**]. Thus, an RCC-appropriate liquid biopsy assay, beyond the commercially available platforms, needs to be developed and optimized to enable the interrogation of RCC-relevant genes. Furthermore, previous studies in other cancers have shown that in addition to soluble plasma, the ctDNA-enriched analytes may include circulating extracellular vesicles (EVs) [**7**,**8**,**9**,**10**], platelets [**11**] and leukocytes known to contain tumor DNA [**12**]. These observations highlight the fact that ctDNA analysis requires robust validation in several technical aspects, which need to be tailored to a particular tumor site due to differences in amenable biofluids, abundance and carriers of genomic sequences released from cancer cells. Such clinical grade information is lacking for RCC.

Among other challenges associated with liquid biopsy analysis in RCC is the low concentration of cell-free DNA (cfDNA) in the blood stream as well as the low proportion of ctDNA present within the cfDNA [**13**]. Somatic mutations of tumors are often present at very low frequencies (<3%) in cfDNA samples [**14**] and conventional next-generation sequencing (NGS) approaches are not optimized for the detection of variants with allele frequency below 5% [**15**]. The implementation of DNA barcoding methods, such as unique molecular identifiers (UMI), coupled with deep-sequencing has improved sensitivity for ctDNA detection [**14**]. However, the

optimization of NGS library preparation and bioinformatics pipelines for ctDNA analysis is a prerequisite of success in this setting. All together, these factors compound the uncertainties about whether liquid biopsy analysis can reflect on status of actionable mutations in RCC tumors.

In this study, we used animal models of RCC to investigate various compartments of blood stream for the enrichment of RCC-associated ctDNA to guide pre-analytical sample preparation for liquid biopsy analysis. Furthermore, we developed and optimized an RCC-specific targeted NGS assay for parallel mutational analyses of tumor tissue-derived DNA and cfDNA to enable a comparison between the status of somatic mutations in tumors as well as in liquid biopsy analytes. Finally, we applied our assay to matched tumor, cfDNA, and evDNA trios from eleven RCC patients to assess the feasibility of liquid biopsy analysis for capturing information of potentially actionable somatic mutations.

2.4 MATERIALS AND METHODS

2.4.1 Cell Culture

The established renal cell cancer cell line 786-O was purchased from the American Type Culture Collection (ATCC; Rockville, MD, USA), and was cultured according to the ATCC recommendations at 37 °C in humidified air with 5% CO_2. Cells were transfected with pLenti CMV V5-LUC Blast (w567-1) (addgene #21474, Watertown, MA, USA) using Lipofectamine 3000 (Invitrogen, Waltham, MA, USA) following the manufacturer's instructions. Stably tagged cells were selected following incubation in medium supplemented with 8 μg/mL blasticidin (Sigma-Aldrich, St. Louis, MI, USA) for 15 days.

2.4.2 Animal Models of RCC

We established orthotopic models of ccRCC by injecting labelled 786-O cells into the subrenal capsule of immune-deficient mice using methods described by Tracz et al. [**16**]. Briefly,

female YFP-SCID mice [**17**] aged six to eight weeks were anesthetized with isoflurane, and a small incision was made between the last rib and the hip joint of a mouse positioned in right lateral recumbency. After popping up the kidney, an ultra-fine needle was inserted into the lower pole of the kidney and advanced until the needle's point reached just below the renal subcapsule. One million viable cells mixed with 63atrigel were slowly injected (volume: 10 μL). After injection, the abdominal wall was closed with a re-absorbable suture and the skin secured with surgical staples. Tumor growth and metastatic disease progression was monitored weekly through luminescence as described previously [**18**]. The mice were sacrificed after development of metastasis and primary tumors were collected and stored at −80 °C. Blood samples were taken via the inferior vena cava (IVC) using 3.8% sodium citrate as anticoagulant, and were centrifuged to separate plasma and buffy coat samples. For EV preparation blood was centrifuged at 200× *g* for 20 min to sediment blood cells, while the upper portion was transferred to another tube and centrifuged at 1500× *g* for 20 min to remove platelets (platelet-poor plasma) before being passed through a 0.45 μm filter, following by ultracentrifugation as described below. All in vivo experiments were performed according to the Animal Use Protocol (AUP) approved by the Institutional Animal Facility Care Committee and following Guidelines of the Canadian Council of Animal Care (CCAC).

2.4.3 Collection of Blood Samples

Patient blood samples were drawn directly prior to surgery into K2 EDTA (BD, Franklin Lakes, NJ, USA) (cfDNA) and Citrate (BD, Franklin Lakes, NJ, USA) (evDNA) tubes. The tubes were inverted to mix and stored at 4 °C until centrifugation. The blood samples were centrifuged within 60 min of collection at 2000 RCF for 15 min at 4 °C to separate plasma from buffy coat

and erythrocyte layers. Plasma and buffy coat fractions were stored in 2 mL cryovials at −80 °C until DNA isolation.

2.4.4 Isolation of EV DNA from Blood Samples

Plasma prepared from mouse or patient blood samples was used for isolation of extracellular vesicles using ultracentrifugation. Briefly, platelet-poor plasma samples (~500 µL) were centrifuged at 110,000× *g* for 70 min at 4 °C. The resulting pellet was washed with PBS and was centrifuged at 110,000× *g* for 70 min at 4 °C for a second time to precipitate EVs. DNA was extracted from EV pellets using the QIAamp DNA Micro kit (Qiagen, Hilden, Germany), following the manufacturer's instructions.

2.4.5 Digital Droplet PCR (ddPCR)

Digital droplet PCR assays were established as described earlier [**12**] for the following specific *VHL* mutation that is present in 786-O cells in consultation with IDT (Integrated DNA Technologies, Coralville, IA, USA): *VHL*, c.311delG, p.G105fsX55. Mutation-specific primers, gblocks and probes (**Table S2**) for these mutations were designed and purchased from IDT. DNA samples were subjected to ddPCR for detection of the *VHL* mutation according to instructions provided by BioRad. Annealing temperature and cycling conditions were optimized, LOD and assay sensitivity were determined using serially diluted gBlocks. Data analysis was performed using QuantaSoft software following the manufacturer's instructions. 900 nM probes and 250 nM primers were mixed with 2× Droplet PCR Supermix (Bio-Rad Laboratories, Hercules, CA, USA), 6 ng of template DNA, and H2O to generate 20 µL for each reaction. The reaction mixture was placed into the sample well of an DG8 cartridge (Bio-Rad, Hercules, CA, USA). 70 µL of droplet-generation oil was loaded into the oil well, and droplets were formed in the droplet generator (BioRad). After processing, the droplets were transferred to a 96-well PCR plate (Eppendorf,

Hamburg, Germany). The PCR amplification was carried out on C1000 TouchTM Thermal Cycler (Bio-Rad) with the following thermal profile: hold at 95 °C for 10 min, 40 cycles of 95 °C 30 s, 55 °C 1 min (ramp 2 °C/s), and 72 °C 30 s, and 1 cycle at 98 °C for 10 min, and ending at 4 °C. After amplification, the plate was loaded on the droplet reader (Bio-Rad) and the droplets from each well of the plate were read automatically. QuantaSoft software was used to count the PCR-positive (FAM channel) and PCR-negative (HEX channel) droplets to provide absolute quantification of target DNA.

2.4.6 Isolation of Genomic and Soluble Cell-Free DNA

Buffy coat, tumor tissues, and plasma samples for 11 RCC patients were provided by McGill RCC biobank (**Table 1**). All samples were received following obtaining written consents from the patients and after approval of the study by McGill University Health Centre Research Ethics Board (MUHC REB). Genomic DNA was isolated from buffy coats (control) and frozen tumor tissue using Dneasy Blood and Tissue kit (Qiagen, Hilden, Germany). We used the same kit for isolation of DNA from mouse buffy coat and tumor samples. Soluble cell free DNA was isolated from 4 mL and 500 uL of patient and mouse plasma samples, respectively, using the QIAseq cfDNA All-in-One kit (Qiagen), following manufacturer's instructions. EV DNA was isolated from plasma sample as described above. Isolated DNA was quantified using Quant-iT PicoGreen dsDNA assay.

Table 1. Information about 11 enrolled patients in this study. Clinical features of tumors as well as genes affected by somatic mutations in each tumor are provided (see **Table S1** for details of somatic mutations).

Patient	Sex	Age	RCC Subtype	Pathological Tumor Stage	Pathological Tumor Grade	Mutated Genes
P1	Male	63	ccRCC	T1b	4/4	VHL, PBRM1
P2	Male	57	ccRCC	T1a	¾	**VHL**, PBRM1
P3	Female	62	ccRCC	T3a	¾	VHL, **SETD2**, PBRM1, MET
P4	Female	78	ccRCC	T1a	¾	VHL, PBRM1
P5	Female	58	ccRCC	T1a	2/4	VHL, COL11A1, SETD2, PBRM1, TRRAP, ATM
P6	Male	66	ccRCC	T1a	¾	VHL, PBRM1, KDM5C
P7	Female	58	ccRCC	T1a	2/4	VHL
P8	Male	77	ccRCC	T3b	4/4	**COL11A1**, **BAP1**, PBRM1
P9	Female	61	ccRCC	T1a	¼	VHL
P10	Female	57	Unclassified RCC	T4	4/4	**KDM5C** *, **SETD2**, **NF2** *
P11	Male	43	ccRCC	T4	4/4	**VHL**, **PBRM1** *, **SETD2** *

Note: * only detected in liquid biopsies. Genes whose tumor-specific somatic mutations were detected in liquid biopsies are indicated in bold.

2.4.7 Targeted Sequencing

Prior to library preparation, gDNA was sheared by the Covaris ultrasonicator to an average peak size of 350 bp. Genomic DNA libraries were generated using the Lucigen NxSeq AmpFree library preparation kit, with eight PCR cycles added according to the manufacturer's guidelines for optional PCR amplification. xGen Dual Index UMI Adapters were added during the library preparation. cfDNA and evDNA libraries were generated using the xGen Prism DNA library prep kit (IDT, Coralville, IA, USA), following the manufacturers guidelines and using the included adapters. Libraries were quantified by qPCR, and the average size fragment was determined using a LabChip GX (PerkinElmer, Waltham, MA, USA). Target enrichment was performed using the xGen Hybridization and Wash Kit (IDT, Coralville, IA, USA) using a custom hybridization panel for RCC (IDT). The enriched libraries were sequenced on a NovaSeq 6000 (paired-end 150).

2.4.8 Synthetic cfDNA Library Preparation

Synthetic liquid biopsy samples were generated using the Seraseq ctDNA Mutation Mix (v2, AF2%) spiked into the Seraseq ctDNA Mutation Mix WT at known allele frequencies of 0.1%, 0.5%, and 1%. Libraries were prepared using the Lucigen NxSeq AmpFree, xGen Prism, and QIAseq cfDNA library kits, following the manufacturers guidelines for each kit. A PCR module was added to amplify libraries generated from the Lucigen NxSeq AmpFree kit. Libraries were quantified by qPCR, and the average size fragment was determined using a LabChip GX (PerkinElmer, Waltham, MA, USA). The synthetic libraries generated by the xGen Prism kit were pooled for hybridization capture with a custom hybridization panel using the xGen Hybridization and Wash Kit. The hybridization capture was quantified by qPCR, average size fragment was determined using the LabChip GX, before sequencing on a NovaSeq 6000 (paired-end 150).

2.4.9 Bioinformatic Analysis

Sequencing reads were processed using the GenPipes DNA-Seq High Coverage pipeline [**19**], with adaptations made for UMI-handling and generation of consensus sequences. Adapters and low-quality reads were removed by Trimmomatic [**20**], and reads were aligned using bwa-mem2 [**21**] to the human genome build GRCh37. UMIs were processed using fgbio [**22**] following the analysis guidelines for xGen Dual Index UMI adapters (IDT) to generate consensus reads. Indel realignment and mate-pair fixing was performed using GATK [**23**] and Picard [**24**]. Somatic calls were generated using VarScan2 [**25**] as well as VarDict [**26**]. Functional annotation of the somatic calls was added by snpEff [**27**], and genomic annotation by Gemini [**28**]. Matched patient normal (buffy coat) samples were used to eliminate germline variants. Non-silent somatic calls underwent manual validation in integrative genomics viewer (IGV) [**29**] to identify somatic

variants present in tumor tissue, circulating cell free DNA, and cell free DNA isolated from extracellular vesicles.

2.4.10 Statistical Analysis

Pearson's correlations were used to assess relationships between gene-specific proportions of sequencing read in different sample types. Differences in library yields were evaluated using Welch's *t*-tests.

2.5 RESULTS

Tumor DNA may be present in several biofluid fractions such as liquid phase (e.g., plasma), EVs (including exosomes), and cells (platelets and leukocytes). To establish a liquid biopsy assay appropriate for ctDNA analysis in RCC, we sought to first identify the most informative biofluid compartment for ctDNA analysis in RCC, and then optimize an RCC-appropriate NGS approach for the detection of somatic mutations in tissue and liquid biopsy samples.

2.5.1 Characteristics of the ctDNA Repertoire in RCC Xenografts

In RCC patients ctDNA represents a modest fraction of cfDNA in blood [**13**]. We questioned whether specific compartments of blood may be enriched for RCC ctDNA, and thereby be more appropriate for liquid biopsy analysis. To minimize technical caveats that originate from the presence of wild-type (background) cfDNA, released by non-cancer cells, we developed orthotopic xenograft models of ccRCC (n = 5 animals), which served as a tool for an unambiguous detection of tumor (human) DNA in all fractions of mouse blood (**Figure 1A**). These models were developed using luminescently-tagged 786-O cancer cells with known RCC-specific mutations, including *VHL* c.311delG. Following the development of metastatic RCC lesions, blood was collected and subjected to fractionation to isolate blood cells (including leukocytes/WBCs), EVs

and soluble cfDNA (**Figure 1B**). To examine these liquid biopsy fractions for the presence of ctDNA, we used digital droplet PCR (ddPCR) to interrogate them for the aforementioned *VHL* mutation. This analysis revealed the presence of the mutated DNA in all of the examined blood fractions (**Figure 1C**). However, while at least 75% of the examined EV samples (four out of five animals) and soluble cfDNA (three out of four animals) fractions were positive for the *VHL* mutation, we detected this mutation in only 50% (two out of four animals) of tested blood cell fractions. Therefore, we focused on soluble cfDNA and EVs fractions for the analysis of patients' liquid biopsy material.

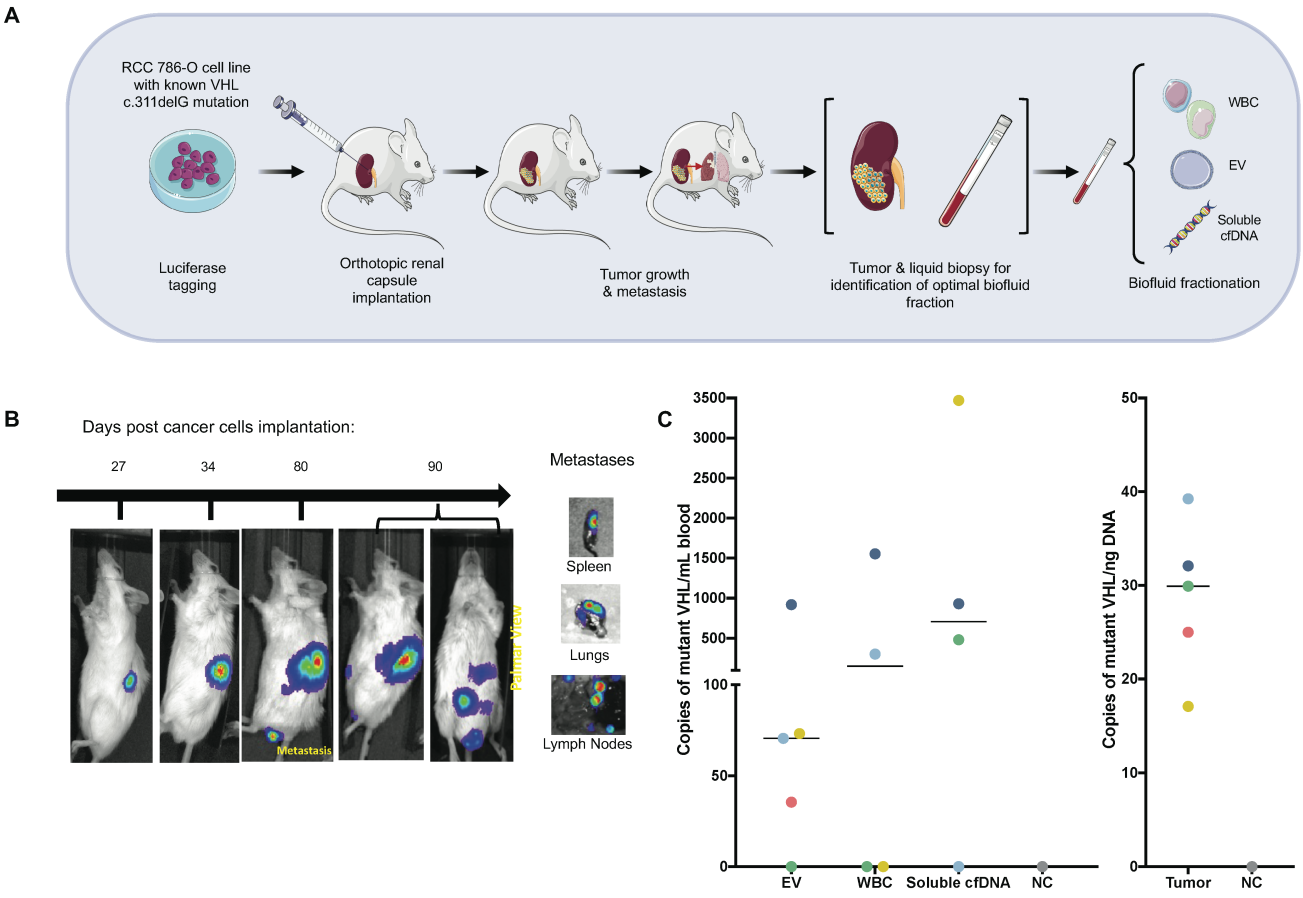


Figure 1. Characterizing the repertoire of ctDNA in RCC. (**A**) Schematic presentation of the in vivo experiments to optimize sample processing for liquid biopsy analysis in RCC. (**B**) Tumor developments in nude mice after orthotropic implantation of 786-O cells. Five mice were implanted, and one animal is shown as an example. Blood samples were drawn after the development of metastatic tumors. (**C**) The presence of the *VHL* c.311delG mutation in human DNA was interrogated in different blood compartments (left) and in tumor tissues (right) using ddPCR. Each dot represents copy number of mutant allele per mL of blood used for DNA isolation from each animal, or per ng of DNA isolated from tumor tissue. Dot colors represent individual animals. NC: negative control.

2.5.2 Development of the RCC-Appropriate Targeted NGS Assay

To enable comparison between tumor tissue and liquid biopsies for the status of potentially actionable somatic mutations in RCC, we sought to develop an NGS assay that is compatible with both genomic DNA (gDNA) and cfDNA samples. Therefore, using custom IDT xGen Lockdown Probes, we designed a targeted NGS panel to capture the entire coding regions and exon-intron boundaries of a gene panel, including *VHL*, *PBRM1*, *SETD2*, *BAP1*, *TP53*, *ATM*, *KDM5C*, *DMD*, *CDKN2A*, *MET*, *NF2*, *KDM6A*, *NFE2L3*, *PTK7*, *TRRAP*, *ATP9B*, and *COL11A1*. These genes are commonly mutated in RCC tumors (e.g., *VHL* in ccRCC or *MET* in papillary RCC), and some of them possess prognostic potential based on the previous large-scale genomic studies [**6**,**30**,**31**] (**Figure 2A**). Thus, this panel can serve for both diagnostic and prognostic purposes. First, we evaluated enrichment efficacy of the panel for the target genes by generating NGS libraries from high-quality gDNA samples using Lucigen NxSeq AmpFree assay, and subjecting them to the capture panel, followed by sequencing (average depth 1699×). These DNA samples were isolated from buffy-coat (control) or fresh-frozen RCC tumors procured from patients, enrolled in the McGill RCC biobank projects (**Table 1**). Sequencing results confirmed the average on-target rates of 87.8% across all samples within a capture (**Figure 2B**), demonstrating the reliable performance of the capture panel to enrich for the desired genes, with no significant difference in sequencing coverage of tumor and normal samples ($p = 0.255$) (**Figure 2C**). Likewise, matched tumor-normal pairs exhibited high correlations ($r > 0.97$) for gene-specific proportion of sequencing reads (**Figure 2D**), indicating that the capture performance is not biased toward sample type and maintains a stable performance across multiple samples. Next, we identified somatic mutations within the gene panel by comparing mutation profiles of tumors to those of their matched blood-driven germline DNA samples (**Table S1**). Our analysis revealed high prevalence of *VHL* (82%,

9/11), *PBRM1* (73%, 8/11) and *SETD2* (36%, 4/11) non-silent mutations in our samples, in line with previous reports [**30**,**31**]. These observations demonstrated the capability of the assay to detect somatic mutations in RCC-relevant genes.

A

Gene	Subtype	References (PMID)
VHL	Clear cell	23792563
NF2	Papillary	20054297, 5401301
MET	Papillary	15579033, 23333114
NFE2L3	Clear cell	25351205
KDM6A	Clear cell	20054297, 22138691, 19330029
TP53	Chromophobe	25401301, 23792563
PBRM1	Clear cell	21248752, 23333114
KDM5C	Clear cell	20054297, 23792563, 22138691
CDKN2A	Papillary	26536169, 23792563
PTK7	Papillary, Clear cell	31361072
BAP1	Clear cell	22683710, 23333114
DMD	Clear cell	
ATM	Clear cell	
SETD2	Clear cell	20054297, 23620406, 23792563
TRRAP	Clear cell	25351205
ATP9B	Clear cell	
COL11A1	Clear cell	25351205

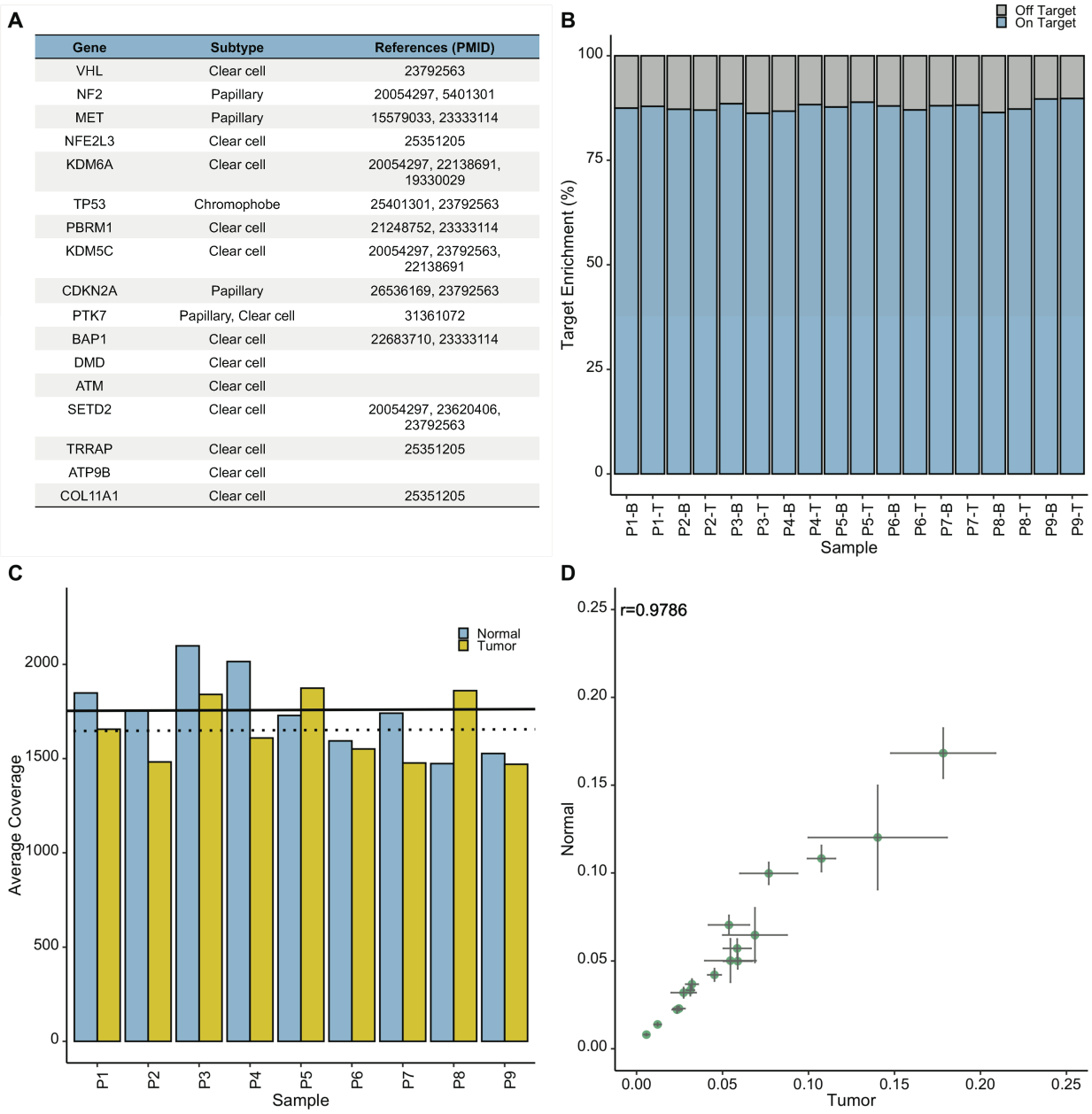


Figure 2. Development and evaluation of the RCC-appropriate NGS assay. (**A**) The RCC assay was designed to target genes that are commonly mutated in RCC tumors, have demonstrated prognostic value in RCC, or have shown association to treatment response in previous large-scale studies. (**B**) The assay showed high efficacy for capturing the targeted regions, with average on-target rates of 87.8% (SD of 0.01) across 18 samples, pooled together prior to the capture. (**C**) Average sequencing coverage is shown for the same 18 samples (9 tumor-normal pairs), included

in a given hybridization capture, with average sequencing coverage depths of 1646× and 1753× of the targeted regions for tumor (dotted line) and normal (solid line) DNA samples, respectively. (**D**) High Pearson's correlations ($r > 0.97, p < 0.0001$) between gene-specific proportions of sequencing reads from tumor and normal samples demonstrates that assay performance is not biased by sample type. Error bars show SD within sample types. SD: standard deviation.

2.5.3 Optimization of the NGS Assay for ctDNA Analysis

Next, we sought to optimize the workflow of our RCC-specific targeted assay for the reliable analysis of cfDNA. Given the low abundance of ctDNA in the limited amount of cfDNA, which can be isolated from plasma of RCC patients, we focused our efforts on two aspects: (1) identifying an effective library preparation approach for cfDNA analysis, and (2) improving detection sensitivity by ultra-deep sequencing coupled with the implementation of unique molecular identifiers (UMIs) in the library preparation workflow to correct for sequencing errors. To this end, we compared the efficacy of the Lucigen NxSeq AmpFree, Qiagen QIAseq cfDNA, and IDT PRISM methods for generating NGS libraries from synthetic cfDNA control samples, which are commercially available. Furthermore, to assess the sensitivity for mutation detection we extended our analysis by including control cfDNA templates with known variant allele frequencies (VAFs) (0.1%, 0.5%, and 1%) for five cancer-associated *TP53* mutations (see the 'Methods' section for details). Among the examined library preparation methods, the IDT PRISM resulted in the greatest library yield in all replicate samples (**Figure 3A**). In addition, the analysis of library profiles confirmed the high quality of the libraries generated by the IDT PRISM approach. Therefore, we subjected these libraries to our targeted capture panel, followed by deep-sequencing (>70 million reads per sample) in order to assess the capture efficacy and stability across multiple samples. This was assessed by evaluating the number of sequencing reads attributed to each target gene, across five replicate samples. While the reads per kilobase of target region, per million mapped reads (RPKM) values showed variable sequencing depths for individual target genes, it maintained a consistent trend across all replicates (**Figure 3B**). This confirmed the ability of the enrichment panel to capture all RCC-genes from control cfDNA samples, and the stability of assay performance.

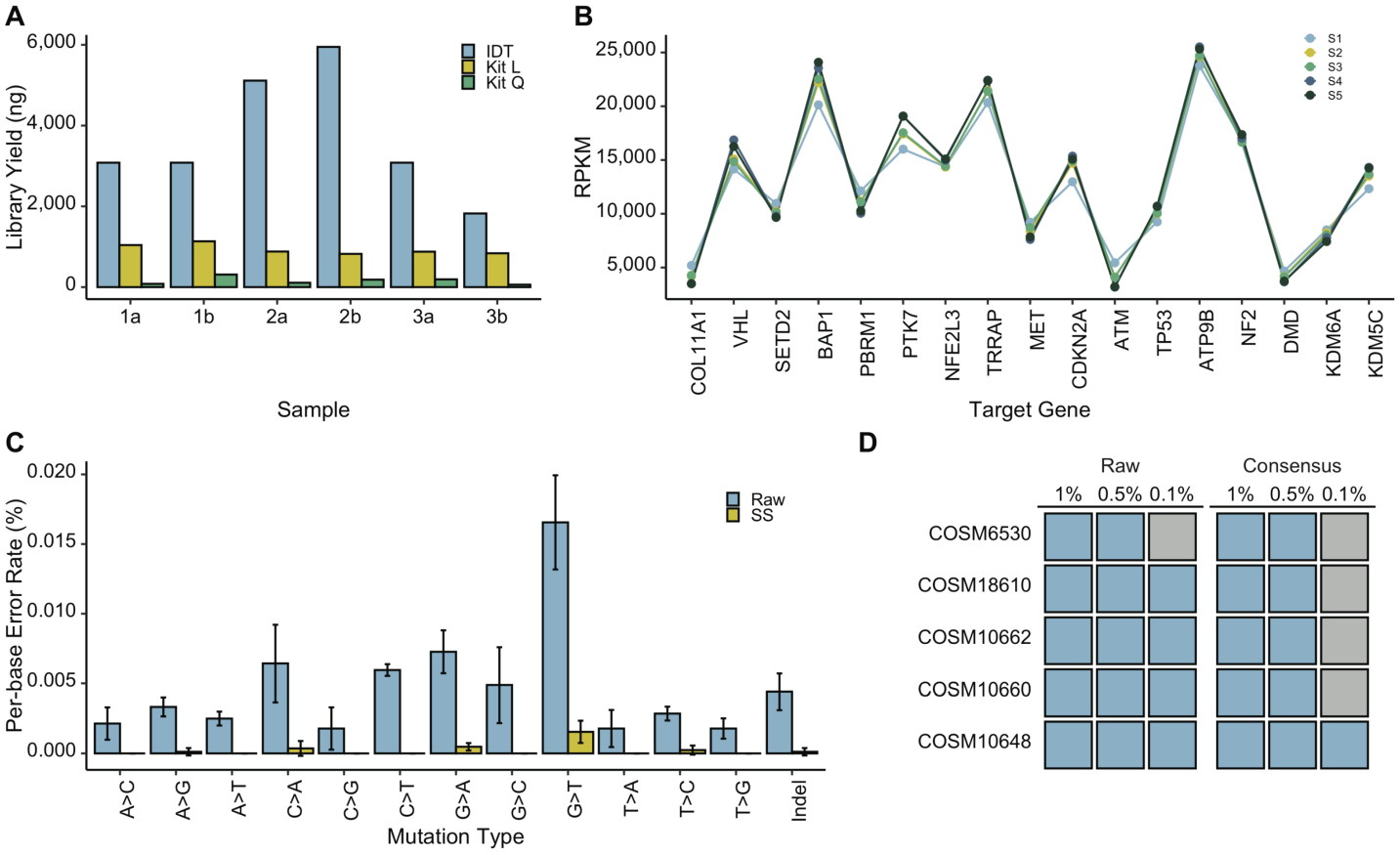


Figure 3. Optimization of the RCC NGS assay for liquid biopsy analysis. (**A**) The comparison between yields of NGS libraries generated by three different methods using 10 ng of the control cfDNA as the starting material is shown. Three individual cfDNA samples tested are identified by numbers (1–3), and technical replicates are marked by letters (a and b). The IDT Prism method resulted in far greater library yields compared to Lucigen NxSeq AmpFree (Kit L) (p = 0.007) and Qiagen QIAseq (Kit Q) (p = 0.002) methods. (**B**) Reproducible performance of the capture panel was examined by comparing gene-specific number of reads per kilobase per million reads (RPKM) values between five replicates (S1–S5) generated from control cfDNA sample using the IDT Prism method. (**C**) The effect of unique molecular identifiers (UMIs) on reducing sequencing errors is shown. False positive rates were assessed with (Consensus) and without (Raw) using UMIs (n =

5). Implementation of UMIs decreased false positive rates across all substitution mutation classes and in indels. (**D**) The assay limit of detection (LOD) was determined by assessing variant drop-out in control samples with known VAFs of 1%, 0.5% and 0.1%. Variant drop-out was observed for allele frequency of 0.1% in both raw and consensus reads. Blue and grey indicate the presence and absence of a given true mutation, respectively.

To optimize our bioinformatics pipeline for mutation detection and establishing the sensitivity of our liquid biopsy assay, we first examined the utility of the UMIs for reducing sequencing errors. For this, we focused on the *TP53* gene, for which we knew the exact location and type of five somatic mutations in the synthetic cfDNA controls, and therefore were able to distinguish them from sequencing errors. Per-base error rates were generated across *TP53* for each substitution mutation class and total mutations including indels from sequencing data, processed once without the implementation of UMIs (raw), and another time with UMIs to generate error-free consensus sequences (SS). Comparisons between the raw and consensus sequences revealed a substantial reduction in per-base error rates in all substitutions classes as well as in indels (**Figure 3C**).

Although the implementation of UMIs vastly decreases the rate of false positives, it can also cause variant drop-out at very low allele frequencies due to the greater stringency. To determine the limit of detection (LOD) for our assay, we investigated variant dropout of the known *TP53* mutations in synthetic cfDNA samples with known VAFs for these mutations ranging from 0.1–1.0%. By obtaining 70 million reads per sample, we were able to detect the *TP53* mutations with VAFs of 0.5 and 1%, whereas we observed dropout of the same mutations at VAF of 0.1%. (**Figure 3D**). Therefore, at VAFs ≥ 0.5% we were able to detect all true *TP53* variants in the synthetic cfDNA controls and minimize the number of false positives by implementing UMIs.

2.5.4 Assay Performance in RCC Liquid Biopsies

Following assay development and optimization using synthetic cfDNA controls, we extended our study to examine the performance of the assay in capturing somatic mutations in liquid biopsy samples from our patient cohort of eleven RCC patients (**Table 1**). The number of patients was limited so as to achieve assay validation (present study) before a larger clinical cohort

could be rigorously powered and examined. Therefore, we sequenced captured targets in cfDNA isolated from plasma and circulating EVs (aiming at 100 M reads/per sample, resulted in more than 5000× depth of on-target coverage) from each patient in order to enable a comparison between liquid biopsy fractions as well as between them and the tumor. For this purpose, we first compared capture efficacy between gDNA and cfDNA fractions for each patient by analyzing proportions of sequencing reads that mapped to each gene. We observed high correlations ($r > 0.95$) between liquid biopsy and tumor DNA samples for gene-specific proportion of sequencing reads (examples are shown in **Figure 4A**), confirming that the performance of the gene-enrichment assay is not dependent on the sample type, and that the assay can be used to compare genetic data between tumor and liquid biopsy samples. Next, we detected somatic mutations in liquid biopsy DNA samples by comparing them to germline DNA isolated from buffy coat samples.

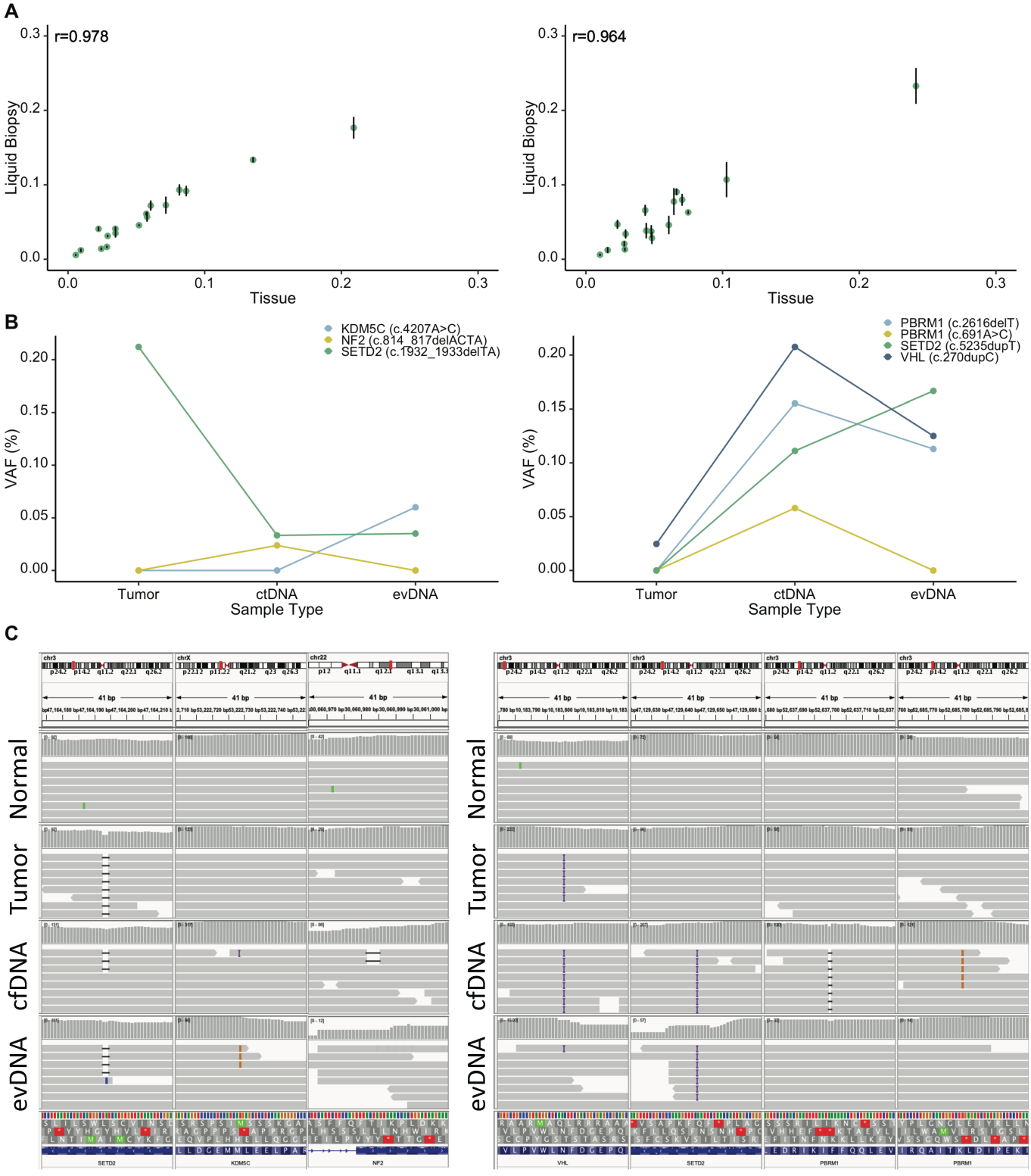

A
r=0.978
r=0.964
Liquid Biopsy
Tissue
B
KDM5C (c.4207A>C)
NF2 (c.814_817delACTA)
SETD2 (c.1932_1933delTA)
PBRM1 (c.2616delT)
PBRM1 (c.691A>C)
SETD2 (c.5235dupT)
VHL (c.270dupC)
VAF (%)
Tumor
ctDNA
evDNA
Sample Type
C
Normal
Tumor
cfDNA
evDNA
SETD2
KDM5C
NF2
VHL
PBRM1

Figure 4. Evaluating the feasibility of detecting tumor somatic mutations in liquid biopsy in RCC. (**A**) In-patient comparisons between tumor and liquid biopsy-driven DNA samples showed high Pearson's correlation in capture distribution. Example data is shown for two patients ($r = 0.978$, $p < 0.0001$ and $r = 0.964$, $p < 0.0001$, for P10 and P11 respectively). Error bars show SD within liquid biopsies. (**B**) Somatic variants identified in both tumor tissue, and in liquid biopsy fractions for patients P10 (left) and P11 (right) at variable allele frequencies. The *SETD2* variant in patient P10 was detected at low frequencies in both ctDNA and evDNA, whereas the *KDM5C* variant was detected only in the evDNA fraction, and the *NF2* variant was detected only in the cfDNA fraction. The *VHL* variant identified in P11 tumor, showed an increased allele frequency in liquid biopsy fractions, whereas mutations in *PBRM1* and *SETD2* were only detected in liquid biopsy fractions. (**C**) Somatic variants identified in tumor tissue, ctDNA, and evDNA visualized using Integrative Genomics Viewer (IGV) for both patient P10 (left) and P11 (right). These somatic variants were not present in the normal (buffy coat) samples (top panel). SD: standard deviation.

Given the different presentation patterns of ctDNA between patients with advanced tumors and those affected with early-stage cancers [**13**,**32**,**33**], we investigated our results in these groups separately. Amongst the 11 patients included in our study, 4 were affected with advanced tumors (stages T3 and T4; P3, P8, P10 and P11). We detected at least one tumor-specific somatic mutation in liquid biopsy fractions from all of these patients (100%, 4/4; **Table 1** and **Table S1**). In patient P3 (stage T3a), we detected a frameshift variant in *SETD2* (c.913dupA) in both soluble cfDNA and evDNA. Similarly, in patient P8 (stage T3b) we detected two somatic missense mutations in *COL11A1* and *BAP1* both liquid biopsy fractions. Interestingly in liquid biopsy samples from P10 and P11, who are affected with T4 stage tumors, we captured all tumor-specific somatic mutations in both cfDNA and evDNA fractions (**Table 1** and **Table S1**, **Figure 4B**). An interesting observation was about a frameshift mutation, c.270dupC, in *VHL* in patient P11, where allele frequency of this mutation was much higher in both liquid biopsy fractions compared to that of tumor DNA (3%, 20%, and 21.5% in tumor, cfDNA, and evDNA, respectively (**Figure 4B**).

Furthermore, we observed appearance of novel mutations, which were not present in the tumor tissues of patients P10 and P11 in their liquid biopsies. For P10, these were a missense mutation, c.4207A > C, in *KDM5C* in the evDNA, and a frameshift mutation in *NF2*, c.814_817delACTA in the cfDNA that were not captured in tumor sequencing data. Likewise, we observed frameshift mutations in *PBRM1* (c.2616delT) and *SETD2* (c.5235dupT), and a missense mutation in *PBRM1* (c.691A > C) in liquid biopsy fractions of P11, that were not present in the sequencing data of the primary tumor (**Figure 4C**).

In patients with low stage RCC (stages T1–T2), most somatic mutations captured in the tumor tissue were not detectable in either cfDNA or evDNA; however, we did detect a somatic stop-gain mutation of *VHL* (c.481C > T) in cfDNA sample of patient P2 who was affected by stage

T1a tumor. These results suggest that optimized liquid biopsy protocol is suitable for interrogating RCC progression in patients with high stage cancer. However, even in a limited number of cases examined the differential abilities of liquid biopsy analytes to carry mutant signatures (cfDNA, evDNA) are readily observed and this factor should be considered in designing future clinical studies.

2.6 DISCUSSION

The utility of ctDNA analysis in the management of kidney cancers has not yet been deeply explored, in part due to the lack of appropriate platforms that enable side-by-side interrogation of somatic mutations in RCC-relevant genes in tumors and in liquid biopsies. In this study, we developed an RCC-focused NGS assay, and optimized it for parallel tissue and liquid biopsy analyses of RCC-relevant mutations. It has been suggested that the most promising use of ctDNA analysis in RCC is as a surveillance biomarker for metastases and to determine the risk of disease recurrence [**34**]. Accordingly, the ability to capture the mutational status of RCC-relevant genes, including *VHL*, *BAP1* and *PBRM1*, is critical for the clinical utility of liquid biopsy analysis in RCC, as the presence of somatic mutations in these particular genes alone or in combination with each other are indicative of distinct disease outcomes [**35**].

In addition, we provided proof-of-principle evidence on the feasibility of capturing tumor-specific diagnostic and prognostic genomic biomarkers in blood-based liquid biopsies in RCC using our assay. Therefore, our assay provides a reliable platform to address key questions that should be investigated in order to establish robust liquid biopsy strategies for RCC. One of such questions is the interpretation of discordance between tumor DNA and ctDNA analysis results. The discordance between somatic alterations detected in RCC tumor tissues and those detected in ctDNA has been suggested to stem from RCC clonal and spatial heterogeneity, long time intervals

between tissue and liquid biopsy sampling, or simply low sensitivity of NGS approaches used for ctDNA analysis [**34**]. The latter is particularly plausible when a somatic mutation is present in tumor DNA but not detected in the liquid biopsy. An explanation for this is that somatic mutations can be present at extremely low allele frequencies in liquid biopsies for different reasons, including the low fraction of ctDNA within cfDNA [**14**]. Indeed, a recent study has shown that the abundance of ctDNA in RCC is very low, as compared to other cancers [**32**] and often ctDNA is detectable in less than 50% of RCC patients [**13**,**33**]. These findings highlight the need for the amplification of starting cfDNA material and applying ultra-deep sequencing of generated NGS libraries, both also known to induce errors in sequencing results [**36**]. It is reasonable to suggest that poor ctDNA detection is due to methodological factors as there are no compelling biological reasons why highly vascularized RCC lesions would not release DNA sequences into blood, passively or actively and in various forms.

Therefore, while exploiting strong amplification and ultra-deep sequencing, we thoroughly investigated the associated technical errors, and corrected for them by implementing molecular tagging through the use of UMIs in library preparation and bioinformatics workflows. Furthermore, the patients included in our study had blood drawn directly prior to surgery, ensuring that the liquid biopsy and tissue sampling are representative of the same time point in tumor evolution. We optimized sample workflow to establish comparable performances in patient matched tumor-normal and liquid biopsy samples, ensuring that potential differences in mutational profiles between sample types are not due to shortcomings in experimental procedures, and rather are reflecting true differences between these sample types. As such, we showed that the assay generates comparable results when applied to tumor DNA and cfDNA from RCC patients with advanced disease (stages T3 or T4 tumor). Notably, at least one somatic variant identified in the

primary tumors of patients with advanced RCC was also captured by our assay in the liquid biopsy fractions. Although very preliminary, due to the limited number of examined samples thus far, this result is promising, as it indicates that our assay does not suffer from major caveats that may result in false-negative observations in liquid biopsy analysis.

Strikingly, we also captured somatic mutations in liquid biopsy that were not present in the tumor DNA. The observation that these mutations were present in both patient-matched ctDNA and evDNA fractions of liquid biopsy argues against the possibility that these are false-positive calls. We believe that these mutations are true somatic mutations that were not captured by tumor DNA analysis. In fact, it has been suggested that cfDNA is a better representation of the primary tumor heterogeneity [**33**,**37**,**38**], as circulating tumor DNA sequences are believed to be shed from the entire tumor, while DNA isolated from tumor cells may be spatially biased by the sampling process and limited by availability of tissue material. This is of paramount importance when considering the clinical utility of somatic mutational analysis, given the spatial heterogeneity that is a hallmark of RCC tumors [**35**,**39**]. Indeed, ctDNA was shown to be a better predictor of drug resistance than tumor tissue in other cancer contexts [**38**]. Moreover, liquid biopsy offers an opportunity to collect multiple longitudinal samples in real time and in a non-invasive manner. As such, liquid-biopsy approaches, as compared to direct tumor biopsy, may possess considerable advantages in developing genomic-based precision medicine in RCC. However, this possibility needs further investigation through parallel analyses of patient-matched tumor DNA and cfDNA samples in large sample sets and using assays that generate comparable results from these sample types, such as the method that we presented in this study. There is also a need to further optimize the detection of ctDNA in patients with low stage RCCs. Previous studies have shown that tumor-fraction in plasma can be enhanced by size-selection of DNA fragments, thus increasing the

sensitivity to detect somatic mutations in renal cancer and other cancers with low amounts of ctDNA [**40**]. Future studies are warranted to examine whether this approach can improve the detection of ctDNA in RCC patients with early-stage tumors.

By emphasizing the assay performance and sensitivity, we explored solutions for some of the major caveats of applying liquid biopsy to renal cancer. Notably, and of equal importance, we also showed that clinically informative somatic mutations in RCC may be present not only in form of soluble ctDNA but also be encapsulated in EVs, suggesting that the analysis of both fractions may provide complementary or confirmatory results for liquid biopsy analysis in RCC. Indeed, previous studies have shown that in some cancers, tumor-derived extracellular vesicles are enriched in tumor DNA [**41**]. EVs serve as carriers of important clinical information, including driver mutations, drug resistance markers, and determinants of immunoregulation [**42**]. Additionally, they may have advantages for biomarker analysis as they protect their cargo from degradation [**42**]; however, harnessing information from EVs for liquid biopsy requires sensitive assays given the low abundance of circulating cancer-related EVs. Our results from RCC animal studies, as well as those on patient material supported the evidence from previous studies [**4**,**5**,**32**] that soluble ctDNA is appropriate for interrogating prognostic biomarkers in RCC; however, also indicated that evDNA is a strong candidate and should be considered in future investigations and with room for refinements (multiplexing, selective capture, others). Taken together, the current study provides a robust workflow and rationale for larger future studies to investigate the utility of ctDNA and evDNA for capturing diagnostic and prognostic biomarkers in RCC.

2.7 CONCLUSIONS

In this study, we developed and optimized an RCC-appropriate NGS assay applicable to both tumor tissue and liquid biopsy fractions. The assay showed consistent performance in all

sample types originating from the same patient (buffy coat, tumor tissue, cfDNA, and evDNA), as well as consistent performance within each sample type. We successfully applied the assay to matched samples from RCC patients with variable clinical features, and captured relevant somatic variants present in primary tumors in both ctDNA and evDNA of patients with advanced tumors. Notably, we demonstrated that ctDNA encapsulated in EVs may contain clinically-relevant mutations in RCC. Furthermore, our assay is the first NGS assay tailored specifically to renal cell carcinoma, including a panel of genes with both diagnostic and prognostic values. This study serves as a demonstration of the capabilities of ctDNA in capturing relevant biomarkers, and lays groundwork for larger studies to further refine the utility of liquid biopsy for enhancing personalized care in RCC.

2.8 AUTHOR CONTRIBUTIONS

Y.R. and J.R. conceived the study. K.I.G. developed and optimized the NGS assay with assistance from A.P. and A.S., established the bioinformatics methods and analyzed the NGS data. P.J. and B.M. generated xenograft models, and collected blood and tissue from the animals. M.M. and L.M. isolated EVs, established and performed ddPCR experiments. S.T. and F.B. provided clinical samples and information, including histological examinations. N.A. collected tissue and liquid biopsy samples from patients and isolated nucleic acids with assistance from M.A. A.J.N. and M.A. validated somatic mutations. K.I.G. and Y.R. prepared the initial draft of the manuscript, and all authors participated in the review and editing. All authors have read and agreed to the published version of the manuscript.

2.9 FUNDING

This work was supported by funds from Kidney Foundation of Canada (2020KHRG-673291 to Y.R.), and the M Elhilali & D Azrieli Chair in Urologic Sciences (S.T.). Y.R. is a research scholar of Fonds de recherche du Québec—Santé (FRQS). J.R. was supported by operating grants from the Cancer Research Society (#25029) and Fondation Charles Bruneau (Liquid Biopsy Program).

2.10 INSTITUTIONAL REVIEW BOARD STATEMENT

All samples were received following obtaining written consents from the patients and after approval of the study by McGill University Health Centre Research Ethics Board (MUHC REB) (2015-1257, 14-388).

2.11 INFORMED CONSENT STATEMENT

Informed consent was obtained from all subjects involved in the study.

2.12 DATA AVAILABILITY STATEMENT

Details about somatic mutations identified in tumors or liquid biopsy samples are available in **Table S1**. Raw NGS data is openly available in the European Genome-Phenome Archive (EGA) (Accession code can be obtained from the authors).

2.13 ACKNOWLEDGEMENTS

The schematic images used in the graphical abstract, and in **Figure 1A** were modified from Servier Medical art (http://smart.servier.com/ accessed on 11 November 2020), licensed under a Creative Commons Attribution 3.0 Unported License.

2.14 CONFLICTS OF INTEREST

Authors declare no conflict of interest.

2.15 SUPPLEMENTARY MATERIALS

Table S1: Detailed somatic variant information, including variant annotation and predicted impact for patients P1–P11.

Available online at: https://www.mdpi.com/article/10.3390/cancers13225825/s1

Table S2: Mutation-specific probes and primers designed for ddPCR analysis.

Probe/Primer	Sequence	Fluorescence label
VHL_DelA_INDEL_probe	/56-FAM/AGA+C+C+CCC+AAA/3IABkFQ/	FAM
VHL_DelA_WT_probe	/5HEX/AG+AC+C+A+CC+CAA/3IABkFQ/	HEX
Forward Primer	5′ACATCGTCAGGTCGCTCTA	
Reverse Primer	5′CCATCCGTTGATGTGCAATG	

2.16 REFERENCES

1. Crowley, E.; Di Nicolantonio, F.; Loupakis, F.; Bardelli, A. Liquid biopsy: Monitoring cancer-genetics in the blood. *Nat. Rev. Clin. Oncol.* **2013**, *10*, 472–484.
2. Siravegna, G.; Marsoni, S.; Siena, S.; Bardelli, A. Integrating liquid biopsies into the management of cancer. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 531–548.
3. Siena, S.; Sartore-Bianchi, A.; Garcia-Carbonero, R.; Karthaus, M.; Smith, D.; Tabernero, J.; Van Cutsem, E.; Guan, X.; Boedigheimer, M.; Ang, A.; et al. Dynamic molecular analysis and clinical correlates of tumor evolution within a phase II trial of panitumumab-based therapy in metastatic colorectal cancer. *Ann. Oncol.* **2018**, *29*, 119–126.
4. Pal, S.K.; Sonpavde, G.; Agarwal, N.; Vogelzang, N.J.; Srinivas, S.; Haas, N.B.; Signoretti, S.; McGregor, B.A.; Jones, J.; Lanman, R.B.; et al. Evolution of Circulating Tumor DNA Profile from First-line to Subsequent Therapy in Metastatic Renal Cell Carcinoma. *Eur. Urol.* **2017**, *72*, 557–564.
5. Hahn, A.W.; Gill, D.M.; Maughan, B.; Agarwal, A.; Arjyal, L.; Gupta, S.; Streeter, J.; Bailey, E.; Pal, S.K.; Agarwal, N. Correlation of genomic alterations assessed by next-generation sequencing (NGS) of tumor tissue DNA and circulating tumor DNA (ctDNA) in metastatic renal cell carcinoma (mRCC): Potential clinical implications. *Oncotarget* **2017**, *8*, 33614–33620.
6. Riazalhosseini, Y.; Lathrop, M. Precision medicine from the renal cancer genome. *Nat. Rev. Nephrol.* **2016**, *12*, 655–666.
7. Rak, J. Extracellular vesicles—Biomarkers and effectors of the cellular interactome in cancer. *Front Pharmacol.* **2013**, *4*, 21.

8. Lee, T.H.; Chennakrishnaiah, S.; Audemard, E.; Montermini, L.; Meehan, B.; Rak, J. Oncogenic ras-driven cancer cell vesiculation leads to emission of double-stranded DNA capable of interacting with target cells. *Biochem. Biophys. Res. Commun.* **2014**, *451*, 295–301.

9. Lázaro-Ibáñez, E.; Sanz-Garcia, A.; Visakorpi, T.; Escobedo-Lucea, C.; Siljander, P.; Ayuso-Sacido, A.; Yliperttula, M. Different gDNA content in the subpopulations of prostate cancer extracellular vesicles: Apoptotic bodies, microvesicles, and exosomes. *Prostate* **2014**, *74*, 1379–1390.

10. Kahlert, C.; Melo, S.A.; Protopopov, A.; Tang, J.; Seth, S.; Koch, M.; Zhang, J.; Weitz, J.; Chin, L.; Futreal, A.; et al. Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. *J. Biol. Chem.* **2014**, *289*, 3869–3875.

11. In ’t Veld, S.; Wurdinger, T. Tumor-educated platelets. *Blood* **2019**, *133*, 2359–2364.

12. Chennakrishnaiah, S.; Meehan, B.; D’Asti, E.; Montermini, L.; Lee, T.H.; Karatzas, N.; Buchanan, M.; Tawil, N.; Choi, D.; Divangahi, M.; et al. Leukocytes as a reservoir of circulating oncogenic DNA and regulatory targets of tumor-derived extracellular vesicles. *J. Thromb. Haemost.* **2018**, *16*, 1800–1813.

13. Yamamoto, Y.; Uemura, M.; Fujita, M.; Maejima, K.; Koh, Y.; Matsushita, M.; Nakano, K.; Hayashi, Y.; Wang, C.; Ishizuya, Y.; et al. Clinical significance of the mutational landscape and fragmentation of circulating tumor DNA in renal cell carcinoma. *Cancer Sci.* **2019**, *110*, 617–628.

14. Shin, H.T.; Choi, Y.L.; Yun, J.W.; Kim, N.K.D.; Kim, S.Y.; Jeon, H.J.; Nam, J.Y.; Lee, C.; Ryu, D.; Kim, S.C.; et al. Prevalence and detection of low-allele-fraction variants in clinical cancer samples. *Nat. Commun.* **2017**, *8*, 1377.

15. Spencer, D.H.; Tyagi, M.; Vallania, F.; Bredemeyer, A.J.; Pfeifer, J.D.; Mitra, R.D.; Duncavage, E.J. Performance of common analysis methods for detecting low-frequency single nucleotide variants in targeted next-generation sequence data. *J. Mol. Diagn.* **2014**, *16*, 75–88.

16. Tracz, A.; Mastri, M.; Lee, C.R.; Pili, R.; Ebos, J.M. Modeling spontaneous metastatic renal cell carcinoma (mRCC) in mice following nephrectomy. *J. Vis. Exp.* **2014**, 1485.

17. Tait, L.R.; Pauley, R.J.; Santner, S.J.; Heppner, G.H.; Heng, H.H.; Rak, J.W.; Miller, F.R. Dynamic stromal-epithelial interactions during progression of MCF10DCIS.com xenografts. *Int. J. Cancer* **2007**, *120*, 2127–2134.

18. Jedeszko, C.; Paez-Ribes, M.; Di Desidero, T.; Man, S.; Lee, C.R.; Xu, P.; Bjarnason, G.A.; Bocci, G.; Kerbel, R.S. Postsurgical adjuvant or metastatic renal cell carcinoma therapy models reveal potent antitumor activity of metronomic oral topotecan with pazopanib. *Sci. Transl. Med.* **2015**, *7*, 282.

19. Bourgey, M.; Dali, R.; Eveleigh, R.; Chen, K.C.; Letourneau, L.; Fillon, J.; Michaud, M.; Caron, M.; Sandoval, J.; Lefebvre, F.; et al. GenPipes: An open-source framework for distributed and scalable genomic analyses. *Gigascience* **2019**, *8*.

20. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120.

21. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760.

22. FulcrumGenomics. Available online: https://github.com/fulcrumgenomics/fgbio (accessed on 1 September 2020).

23. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome. Res.* **2010**, *20*, 1297–1303.

24. Broad Institute. Picard Toolkit. Available online: http://broadinstitute.github.io/picard/ (accessed on 1 September 2020).

25. Koboldt, D.C.; Zhang, Q.; Larson, D.E.; Shen, D.; McLellan, M.D.; Lin, L.; Miller, C.A.; Mardis, E.R.; Ding, L.; Wilson, R.K. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome. Res.* **2012**, *22*, 568–576.

26. Lai, Z.; Markovets, A.; Ahdesmaki, M.; Chapman, B.; Hofmann, O.; McEwen, R.; Johnson, J.; Dougherty, B.; Barrett, J.C.; Dry, J.R. VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* **2016**, *44*, e108.

27. Cingolani, P.; Platts, A.; Wang le, L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **2012**, *6*, 80–92.

28. Paila, U.; Chapman, B.A.; Kirchner, R.; Quinlan, A.R. GEMINI: Integrative exploration of genetic variation and genome annotations. *PloS Comput. Biol.* **2013**, *9*, e1003153.

29. Thorvaldsdóttir, H.; Robinson, J.T.; Mesirov, J.P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform.* **2013**, *14*, 178–192.

30. Scelo, G.; Riazalhosseini, Y.; Greger, L.; Letourneau, L.; Gonzàlez-Porta, M.; Wozniak, M.; Lathrop, M. Variation in genomic landscape of clear cell renal cell carcinoma across Europe. *Nat. Commun.* **2014**, *5*, 5135.

31. Ricketts, C.J.; De Cubas, A.A.; Fan, H.; Smith, C.C.; Lang, M.; Reznik, E.; Bowlby, R.; Gibb, E.A.; Akbani, R.; Beroukhim, R.; et al. The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep.* **2018**, *23*, 313–326.

32. Smith, C.G.; Moser, T.; Mouliere, F.; Field-Rayner, J.; Eldridge, M.; Riediger, A.L.; Chandrananda, D.; Heider, K.; Wan, J.C.M.; Warren, A.Y.; et al. Comprehensive characterization of cell-free tumor DNA in plasma and urine of patients with renal tumors. *Genome. Med.* **2020**, *12*, 23.

33. Bettegowda, C.; Sausen, M.; Leary, R.J.; Kinde, I.; Wang, Y.; Agrawal, N.; Bartlett, B.R.; Wang, H.; Luber, B.; Alani, R.M.; et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **2014**, *6*, 224.

34. Cimadamore, A.; Massari, F.; Santoni, M.; Mollica, V.; Di Nunno, V.; Cheng, L.; Lopez-Beltran, A.; Scarpelli, M.; Montironi, R.; Moch, H. Molecular characterization and diagnostic criteria of renal cell carcinoma with emphasis on liquid biopsies. *Expert. Rev. Mol. Diagn.* **2020**, *20*, 141–150.

35. Turajlic, S.; Xu, H.; Rowan, A.; Chambers, T.E.A. Tracking cancer evolution reveals constrained routes to metastases: TRACERx Renal. *Cell* **2018**, *173*, 581–594.

36. Perakis, S.; Speicher, M.R. Emerging concepts in liquid biopsies. *BMC Med.* **2017**, *15*, 75.

37. Russo, M.; Siravegna, G.; Blaszkowsky, L.S.; Corti, G.; Crisafulli, G.; Ahronian, L.G.; Mussolin, B.; Kwak, E.L.; Buscarino, M.; Lazzari, L.; et al. Tumor Heterogeneity and Lesion-Specific Response to Targeted Therapy in Colorectal Cancer. *Cancer Discov.* **2016**, *6*, 147–153.

38. Parikh, A.R.; Leshchiner, I.; Elagina, L.; Goyal, L.; Levovitz, C.; Siravegna, G.; Livitz, D.; Rhrissorrakrai, K.; Martin, E.E.; Van Seventer, E.E.; et al. Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat. Med.* **2019**, *25*, 1415–1421.

39. Gerlinger, M.; Rowan, A.J.; Horswell, S.; Math, M.; Larkin, J.; Endesfelder, D.; Gronroos, E.; Martinez, P.; Matthews, N.; Stewart, A.; et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **2012**, *366*, 883–892.

40. Mouliere, F.; Chandrananda, D.; Piskorz, A.M.; Moore, E.K.; Morris, J.; Ahlborn, L.B.; Mair, R.; Goranova, T.; Marass, F.; Heider, K.; et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci. Transl. Med.* **2018**, *10*.

41. Vagner, T.; Spinelli, C.; Minciacchi, V.R.; Balaj, L.; Zandian, M.; Conley, A.; Zijlstra, A.; Freeman, M.R.; Demichelis, F.; De, S.; et al. Large extracellular vesicles carry most of the tumour DNA circulating in prostate cancer patient plasma. *J. Extracell Vesicles* **2018**, *7*, 1505403.

42. Chennakrishnaiah, S.; Tsering, T.; Aprikian, S.; Rak, J. Leukobiopsy—A Possible New Liquid Biopsy Platform for Detecting Oncogenic Mutations. *Front Pharmacol.* **2019**, *10*, 1608.

BRIDGING STATEMENT TO CHAPTER 3

Current prognostic tools and clinical nomograms for ccRCC are reliant on clinicopathological information, and are inconsistent in identifying patients at risk of developing relapse or metastasis. Given that up to 40% of patients with localized disease will experience relapse or metastasis post-nephrectomy, identifying high-risk patients who will benefit from systemic therapies is critical for 'optimal clinical management'. However, the lack of routine prognostic biomarkers makes risk-stratification for ccRCC difficult.

There are currently no prognostic biomarkers that consider the genetic profile of ccRCC tumors. Individual genes such as *VHL*, *PBRM1*, *BAP1*, and *SETD2* have been identified as driver genes, but the mutation status of any individual gene has not demonstrated substantial prognostic value when evaluated in isolation. It is more likely that investigating the subclonal diversity of ccRCCs within large cohorts, considering combinations of recurrently mutated genes, will more effectively drive biomarker development.

Previous large studies have been limited by a focus on individual genes, methods that may not be sensitive for capturing subclonal mutations at low allele frequencies, or insufficient sample sizes for capturing the heterogeneous mutational landscape of RCC. We speculate that approaches considering evolutionary patterns in ccRCC may be more successful in identifying robust markers of disease prognosis than investigations focusing on individual genes. Recently proposed evolutionary trajectories of ccRCC, described in Chapter 1.2.4, appear to have correlated clinical phenotypes, but have not been examined at large-scale.

Chapter 3 comprises a manuscript investigating the associations between the genetic evolution and the repertoire of common somatic genetic alterations in ccRCC, and examining their potentials for prognosis in ccRCC. Leveraging the RCC-relevant assay described in Chapter 2, we characterized the landscape of somatic mutations in a large clinically annotated cohort of ccRCCs,

with the aim of identifying biomarkers of prognosis. We define a genomic classifier, consistent with evolutionary subtypes of ccRCC, that can identify groups of patients with diverging risks of disease recurrence.

CHAPTER 3. APPLICATION OF GENOMIC SEQUENCING TO REFINE PATIENT STRATIFICATION FOR ADJUVANT THERAPY IN RENAL CELL CARCINOMA

Naveen S. Vasudev[1*%], Ghislaine Scelo[2*$], Kate I. Glennon[3,4*], Michelle Wilson[1], Louis Letourneau[3], Robert Eveleigh[3], Nazanin Nourbehesht[3,4], Madeleine Arseneault[3], Antoine Paccard[3], Lars Egevad[5], Juris Viksna[6], Edgars Celms[6], Sharon M. Jackson[1], Behnoush Abedi-Ardekani[2], Anne Y. Warren[7], Peter J. Selby[1], Sebastian Trainor[1], Michael Kimuli[8], Jon Cartledge[8], Naeem Soomro[9], Adebanji Adeyoju[10], Poulam M. Patel[11], Magdalena B. Wozniak[2] , Ivana Holcatova[12], Antonin Brisuda[13], Vladimir Janout[14], Estelle Chanudet[2], David Zaridze[15], Anush Moukeria[15], Oxana Shangina[15], Lenka Foretova[16], Marie Navratilova[16], Dana Mates[17], Viorel Jinga[18], Ljiljana Bogdanovic[19], Bozidar Kovacevic[20], Anne Cambon-Thomsen[21], Guillaume Bourque[3,4], Alvis Brazma[22], Jörg Tost[23], Paul Brennan[2], Mark Lathrop[3,4], Yasser Riazalhosseini[3,4#%], Rosamonde E. Banks[1#]

[*]equal first author [#] equal last author [%] corresponding authors

1. Leeds Institute of Medical Research at St James's, University of Leeds, St James's University Hospital, Leeds LS9 7TF, UK
2. World Health Organisation (WHO), International Agency for Research on Cancer (IARC), The Genomic Epidemiology Branch, 150 cours Albert Thomas, 69372, Lyon, France.
3. McGill University Genome Centre, 740 Doctor Penfield Avenue, Montreal, QC, H3A 0G1, Canada
4. Department of Human Genetics, McGill University, 1205 Doctor Penfield Avenue Montreal, QC, H3A 1B1, Canada
5. Department of Oncology-Pathology, Karolinska Institutet, SE-171 77 Stockholm, Sweden
6. Institute of Mathematics and Computer Science, University of Latvia, 29 Rainis Blvd, Riga, LV-1459, Latvia
7. Department of Histopathology, Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge, CB2 0QQ, UK

8. Pyrah Department of Urology, Leeds Teaching Hospitals NHS Trust, Lincoln Wing, St James's University Hospital, Leeds LS9 7TF, UK
9. Newcastle Upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK
10. Stockport NHS Foundation Trust, Stockport, UK.
11. Division of Cancer & Stem Cells, School of Medicine, University of Nottingham, Nottingham UK
12. Charles University in Prague, First Faculty of Medicine, Institute of Hygiene and Epidemiology, Studničkova 7, Praha 2, 128 00, Prague, Czech Republic
13. University Hospital Motol, V Uvalu 84, 150 06 Prague, Prague, Czech Republic
14. Faculty of Health Sciences, Palacky University, Olomouc, Czech Republic
15. N.N. Blokhin National Medical Research Centre of Oncology, Kashirskoye shosse 24, Moscow 115478, Russian Federation
16. Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic, Zluty Kopec 7, 656 53 Brno, Czech Republic
17. National Institute of Public Health, Dr Leonte Anastasievici 1-3, sector 5, Bucuresti 050463, Romania
18. Carol Davila University of Medicine and Pharmacy, Prof. Dr. Th. Burghele Clinical Hospital, 20 Panduri street, 050659, Bucharest, Romania
19. Institute of Pathology, School of Medicine Belgrade, University of Belgrade, Serbia
20. Institute of Pathology and Forensic Medicine, Military Medical Academy, Belgrade, Serbia
21. Institut National de la Santé et de la Recherche Médicale (INSERM) and Université Toulouse III Paul Sabatier (UPS), Joint Unit UMR 1027, Faculté de médecine, 37 allées Jules Guesde, 31000 Toulouse, France
22. European Bioinformatics Institute, European Molecular Biology Laboratory, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK
23. Centre National de Recherche en Génomique Humaine, CEA - Institut de Biologie Francois Jacob, University Paris Saclay, 2 rue Gaston Crémieux, 91000 Evry, France

[$] Current affiliation: Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin, Italy

Originally published in Clinical Cancer Research:
Vasudev NS, Scelo G, Glennon KI, Wilson M, Letourneau L, Eveleigh R, Nourbehesht N, Arseneault M, Paccard A, Egevad L, Viksna J, Celms E, Jackson SM, Abedi-Ardekani B, Warren AY, Selby PJ, Trainor S, Kimuli M, Cartledge J, Soomro N, Adeyoju A, Patel PM, Wozniak MB, Holcatova I, Brisuda A, Janout V, Chanudet E, Zaridze D, Moukeria A, Shangina O, Foretova L, Navratilova M, Mates D, Jinga V, Bogdanovic L, Kovacevic B, Cambon-Thomsen A, Bourque G, Brazma A, Tost J, Brennan P, Lathrop M, Riazalhosseini Y, Banks RE. Application of Genomic Sequencing to Refine Patient Stratification for Adjuvant Therapy in Renal Cell Carcinoma. *Clin Cancer Res*. 2023 Apr 3;29(7):1220-1231. Doi: 10.1158/1078-0432.CCR-22-1936.

3.1 ABSTRACT

Purpose: Patients with resected localized clear cell renal cell carcinoma (ccRCC) remain at variable risk of recurrence. Incorporation of biomarkers may refine risk prediction and inform adjuvant treatment decisions. We explored the role of tumor genomics in this setting, leveraging the largest cohort to date of localized ccRCC tissues subjected to targeted gene sequencing.

Experimental design: The somatic mutation status of 12 genes was determined in 943 ccRCC cases from a multinational cohort of patients, and associations to outcomes were examined in a Discovery (n=469) and Validation (n=474) framework.

Results: Tumors containing a *VHL* mutation alone were associated with significantly improved outcomes in comparison to tumors containing a *VHL* plus additional mutations. Within the Discovery cohort, those with *VHL*+0, *VHL*+1, *VHL*+2 and *VHL*+≥3 tumors had DFS rates of 90.8%, 80.1%, 68.2% and 50.7% respectively, at 5 years. This trend was replicated in the Validation cohort. Notably, these genomically-defined groups were independent of tumor mutational burden. Amongst patients eligible for adjuvant therapy, those with a *VHL*+0 tumor (29%) had a 5-year DFS rate of 79.3% and could, therefore, potentially be spared further treatment. Conversely, patients with *VHL*+2 and *VHL*+≥3 tumors (32%) had equivalent DFS rates of 45.6% and 35.3%, respectively, and should be prioritized for adjuvant therapy.

Conclusions: Genomic characterization of ccRCC identified biologically distinct groups of patients with divergent relapse rates. These groups account for the ~80% of cases with *VHL* mutations and could be used to personalize adjuvant treatment discussions with patients as well as inform future adjuvant trial design.

3.2 STATEMENT OF TRANSLATIONAL RELEVANCE

Determination of recurrence risk in patients with resected localized renal cell carcinoma (RCC) remains reliant on pathologic grounds alone. Despite extensive characterization of these tumors at the genomic level, the application of such information for patient benefit has not been fully realized. We undertook targeted DNA sequencing of tumor-normal sample pairs from a large multinational cohort of patients with localized clear-cell RCC and explored the impact of these data on patient outcome. Using a 12-gene classifier, we are able to classify almost 80% of patients into one of four groups with highly divergent risk of recurrence or death from RCC, independent of tumor stage and grade or patient age. Tumors found to contain a von-Hippel Lindau mutation alone were associated with most favorable outcomes and defines a group of patients who may potentially be spared adjuvant therapy. Conversely, patients at very high risk of recurrence are identified as those who should be managed aggressively.

3.3 INTRODUCTION

Renal cell carcinoma (RCC) presents a growing global health problem, with over 400,000 new cases each year and rising incidence rates worldwide [**1, 2**]. The majority (70%–80%) of RCCs are clear-cell RCC (ccRCC), which are characterized by their highly variable clinical course. For clinicians, this heterogeneity poses significant challenges to the delivery of individual patient care.

Most patients (75%–80%) present with apparent localized disease and are offered curative intent treatment in the form of surgery or ablation, but 20% to 30% of these patients will subsequently relapse [**3**]. The estimation of likely patient outcomes underpins further decision-making, guiding patient counselling, length and intensity of follow-up, and the selection of patients for adjuvant therapy.

Immune checkpoint inhibitors (ICI) have been explored for their efficacy in the adjuvant RCC setting within large randomized trials involving thousands of patients [**4**]. Pembrolizumab, a programmed death 1—targeted agent, has recently received approval in both the US and Europe for use in patients with resected intermediate-high and high-risk RCC, based on results from the ongoing Phase III KEYNOTE-564 study. A significant disease-free survival (DFS) advantage in favor of 1 year of adjuvant pembrolizumab versus placebo was demonstrated in this trial [**5, 6**]. Determination of risk was based on tumor–node–metastasis (TNM) stage and tumor grade, with the majority (86%) of recruited patients having either pT2 (high grade) or pT3 (any grade) tumors. However, 68% of patients in the placebo arm of this study remained free of recurrence at 2 years [**5**], indicating the limitations of current risk-estimation methods. Given the subsequent recent publication of three negative adjuvant RCC studies also employing ICIs [**7–9**], and the financial cost and potentially severe immune-mediated adverse events associated with these agents, careful consideration must be given to both the risks versus benefits of such therapy. Those at highest risk of recurrence should be prioritized whilst sparing those with biologically lower-risk tumors, highlighting an urgent need for easily implemented molecular tools to improve individual patient stratification [**10, 11**].

ccRCC has been extensively characterized at the genetic, epigenetic, and transcriptomic level [**12, 13**]. Mutation or methylation of the von-Hippel Lindau (*VHL*) tumor suppressor gene occurs in the majority (≈80%) of sporadic cases [**13**]. In addition, recurrent mutations in *PBRM1* (≈40%), *SETD2* (≈20%), and *BAP1* (≈15%) are observed, along with a large number of lower frequency events. The application of genomics in ccRCC to inform clinical practice is yet to be realized. Clinical association studies to date have typically considered each gene in isolation (mutated versus non-mutated tumors), often using small cohorts and much of our current

understanding comes from a single dataset, The Cancer Genome Atlas (TCGA) study [**12, 14**]. Recently, distinct evolutionary subtypes of ccRCC have been proposed that appear biologically and clinically distinct, including subtypes that are *VHL* wild-type (WT), *VHL* monodrivers, and those that have multiple clonal drivers [**15**]. This has advanced our understanding of how genomic alterations may impact on disease progression, and has defined a new paradigm in linking genomic signatures of tumors to clinical outcome.

Our recent study "Cancer Genomics of the Kidney (CAGEKID)", as part of the International Cancer Genome Consortium, provided the first comprehensive multinational description of the molecular architecture of ccRCC [**13**]. Using a larger multinational validation cohort with extensive associated demographic, clinical, and follow-up data, we have now further explored some of the more commonly observed genetic changes in ccRCC to enable their molecular classification and demonstrate how such information may be applied in the clinic for patient benefit.

3.4 METHODS

3.4.1 Patients and samples

Patients undergoing nephrectomy for suspected renal cancer between March 1998 and February 2014 across 17 centers (detailed in **Supplementary Methods**) in the UK, Czech Republic, Romania, Russia, and Serbia donated blood and tissue samples for research following written informed consent, based on the Declaration of Helsinki principles. Ethical approvals were obtained from the Leeds (East) Local Research Ethics Committee, the International Agency for Research on Cancer Ethics Committee, as well as from local ethics committee for recruiting centers in Czech Republic, Romania, Russia, Serbia, and Bosnia & Herzegovina. Inclusion and exclusion criteria and sampling were as previously described [**13**]. All samples were subject to

panel pathology review. A single frozen tissue block was used for sequencing in the majority (86%) of cases, with a minimum cutoff of 70% (cohorts 1 and 2; C1 and C2 – combined to form the Discovery cohort) or 50% (cohort 3; C3 – the Validation cohort) viable tumor cells in sections flanking the analyzed tissue. For the remaining cases, formalin-fixed, paraffin-embedded (FFPE) tissue was used with three targeted 1-mm punches from a single block in each case.

3.4.2 Preparation of DNA

DNA was isolated from buffy coats and frozen tumor tissue using FlexiGene DNA Kit and Dneasy Blood & Tissue Kit (Qiagen, Toronto, Canada) respectively, following manufacturers' instructions. Chemagic DNA Cell Kit Special (Perkin Elmer) was used to isolate DNA from FFPE tissues, and DNA samples were quantified using Quant-iT PicoGreen dsDNA Assay (Life Technologies, Burlington, Canada).

3.4.3 Genomic Sequencing

Samples from C1 were analyzed using whole-genome sequencing (WGS) as previously reported [**13**]. Libraries were generated using the Nextera Rapid Capture Enrichment library preparation Kit (Illumina; 42-gene panel, C2) and the Lucigen AmpFree library preparation kit with xGen Dual Index UMI adapters [Integrated DNA Technologies (IDT); 12-gene panel, C3] according to the manufacturers' recommendations. Libraries were quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies) and the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (Kapa Biosystems). Average fragment size was determined using a LabChip GX (PerkinElmer) instrument. Captured libraries were sequenced on a HiSeq 2500 (2×100 cycles; C2) or on a NovaSeq 6000 (2×150 cycles; C3), and bcl2fastq (Illumina) was used to de-multiplex samples and generate fastq reads.

3.4.4 Bioinformatic Analyses

Data analysis was performed using the GenPipes DNA-Seq High coverage pipeline [**16**] with the default parameters, and an added step for generating consensus reads (described below). Adapters and low quality reads were removed using Trimmomatic (RRID: SCR_011848), and reads were aligned to the human genome build GRCh37 using bwa-mem. Mapped reads were further refined using GATK (RRID: SCR_001876) and Picard (RRID: SCR_006525) for indel realignment and verifying mate-pairs. Deduplication (Cohorts 1 and 2) was performed with Picard. UMIs (Cohort 3) were processed using fgbio [**17**] according to the IDT analysis guidelines for xGen Dual Index UMI Adapters to generate consensus reads. The consensus reads were input back into the pipeline for refining with GATK and Picard. Somatic and germline calls were generated using VarScan2 [**18**] and the identified indels as well as complex variants were re-called using VarDict [**19**] to ensure proper identification of complex variants that were not properly resolved by VarScan2. Calls were further processed with the addition of functional annotations using snpEff [**20**] and genomic annotation by Gemini [**21**] for prioritization of candidate mutations. We extracted somatic variants with predicted high or moderate functional impact and filtered the candidate somatic mutations for a minimum allele frequency of 5% in tumor samples. Among the resulting variants, those with a frequency less than 0.01% in the 1000 Genomes database were selected for manual validation in Integrative Genomics Viewer (IGV) to ensure the quality of the calls and the absence in patient-matched germline DNA.

3.4.5 Survival Definitions

DFS was defined as time from date of surgery to local or regional recurrence, metastases, contralateral kidney cancer, or death, whichever occurred first. Any patients without disease recurrence were censored at the date they were last known to be recurrence free (for patients who

died without recurrence this was date of death). RCC-specific survival was defined as time from date of surgery to death from RCC. Non-RCC–related survival was defined as time from date of surgery to death from causes other than RCC.

3.4.6 Statistical Methods

Cox proportional hazards (PH) models were used to estimate the association between gene mutation frequency and the selected endpoints. The Kaplan–Meier method was used to estimate survival functions and multivariable survival models were constructed to assess the independence of associations between gene mutation status and each endpoint.

To validate the Cox PH model constructed for comparison of genomically defined groups, patients were divided into Discovery and Validation datasets. The Kaplan–Meier curves for each dataset were used to confirm good separation between risk groups, and the HRs between risk groups were well maintained between the Discovery and Validation datasets.

Cumulative incidence functions were estimated using the cmprsk package [**22**], and cause-specific hazard models were used to assess RCC-related death in the presence of competing risks. All statistical analyses were undertaken in the R environment for statistical computing.

3.4.7 Data Availability

Clinical information and list of somatic mutations of the 12 genes are included in the supplementary data (**Supplementary Tables S1** and **S2**). Raw sequence data are available in the European Genome-phenome Archive under accession codes EGAS00001000083 (C1 and C2) and EGAS00001007004 (C3).

3.5 RESULTS

3.5.1 Patients

A total of 943 patients with ccRCC were included in this study. Patient characteristics are summarized in **Table 1** (detailed in **Supplementary Table S3**). Most patients (806; 85%) had stage I–III disease. Median follow-up was 5.7 years [interquartile range (IQR), 3.8–7.3], with 160 (17%) recurrences and 192 (20%) cancer-specific deaths recorded.

Table 1. Clinical and demographic characteristics in all patients (additional detail in Supplementary Table S1).

Characteristic		All (n=943)	Discovery (n=469)	Validation (n=474)
Age at surgery (years)	Median (range)	61 (23-86)	62 (23-86)	61 (26-85)
Sex	Female	359 (38.1)	193 (41.2)	166 (35.0)
	Male	580 (61.5)	276 (58.8)	304 (64.1)
	Missing	4 (0.4)	0 (0)	4 (0.8)
Body mass index	Median (range)	27.8 (14.9-49.2)	27.8 (14.9-49.6)	27.7 (16-49.2)
Country	Czech Republic	342 (36.3)	133 (28.4)	209 (44.1)
	UK	291 (30.9)	150 (32.0)	141 (29.7)
	Russia	197 (20.9)	129 (27.5)	68 (14.3)
	Romania	75 (8.0)	50 (10.7)	25 (5.3)
	Serbia	34 (3.6)	7 (1.5)	27 (5.7)
	Missing	4 (0.4)	0 (0)	4 (0.8)
Pathological tumor size (mm)	Median (range)	55 (13-220)	58 (13-220)	55 (12-225)
Pathological tumor stage	1a	277 (29.4)	128 (27.3)	149 (31.4)
	1b	215 (22.8)	114 (24.3)	101 (21.3)
	2	107 (11.3)	57 (12.2)	50 (10.5)
	3	288 (30.5)	155 (33.0)	133 (28.1)
	4	10 (1.1)	7 (1.5)	3 (0.6)
	Missing	46 (4.9)	8 (1.7)	38 (8.0)
Overall stage (TNM)	I	486 (51.5)	235 (50.1)	251 (53.0)
	II	95 (10.1)	47 (10.0)	48 (10.1)
	III	225 (23.9)	117 (24.9)	108 (22.8)
	IV	132 (14.0)	70 (14.9)	62 (13.1)
	Missing	5 (0.5)	0 (0)	5 (1.1)
FFPE Fuhrman grade[a]	1	122 (12.9)	66 (14.1)	56 (11.8)
	2	405 (42.9)	215 (45.8)	190 (40.1)
	3	279 (29.6)	139 (29.6)	140 (29.5)
	4	91 (9.7)	45 (9.6)	46 9.7)
	Missing	46 (4.9)	4 (0.9)	42 (8.9)
Follow-up (years)	Median (IQR)	5.7 (3.8-7.3)	6.0 (3.6-7.6)	5.5 (4.1-6.8)

Note: *n* (%) unless otherwise stated.
[a]As per original reporting pathologist at each center.

3.5.2 Overview of gene sequencing results

Patient samples were sequenced in three cohorts, as summarized in **Fig. 1A**. C1 consisted of 93 patient sample pairs (tumor and matched germline DNA) subjected to WGS, previously reported in a descriptive study [**13**]. C2 consisted of sample pairs from an additional 376 patients, of which 24 were analyzed by exome sequencing and 352 underwent targeted sequencing of 42 genes, identified as being most frequently mutated in C1 or other large-scale genomic studies [**12, 23**; Supplementary Table S4]. C3 consisted of sample pairs from an additional 474 patients, which were analyzed by targeted sequencing of 12 genes (*ATM, ATP9B, BAP1, COL11A1, DMD, KDM5C, PBRM1, PTK7, SETD2, TP53, TRRAP*, and *VHL*), included in an RCC-focused gene panel [**24**]. These genes were selected on the basis of their known role in ccRCC biology/previously reported clinical associations (*BAP1*, *KDM5C*, *PBRM1*, *SETD2*, *TP53*, *VHL*), and/or our preliminary observed associations with outcome or other clinical parameters in C1 and C2 (*ATM*, *ATP9B*, *COL11A1*, *DMD*, *PTK7*, *TRRAP*).

For the main analysis presented here, we focus on the final selected 12 genes only, irrespective of sequencing approach/panel. Amongst all cases (n = 943), the most frequently mutated genes were *VHL* (76%), *PBRM1* (39%), *SETD2* (18%), *BAP1* (14%), and *KDM5C* (8%). Our *VHL* mutation detection rate is notably higher than that reported in previous studies [**14, 25–27**] and is likely to reflect the high depth of sequencing coverage in our targeted sequencing approach (average 296X and 1475X in C2 and C3, respectively), sensitivity for detection of indel mutations, and the exclusion of non-ccRCCs. Furthermore, the similar rates of gene mutations between cohorts and irrespective of whether frozen or FFPE tissue, confirm the sensitivity of our screen, and consistency of genomic results across examined cohorts (Supplementary Table S4). Therefore, to identify reproducible and robust associations, we combined cases from C1 and C2

to form a Discovery cohort (n = 469) and considered C3 as a Validation cohort (n = 474). The Discovery and Validation cohorts showed no significant differences in mutation rates for each gene, and for relevant clinical features such as tumor stage, grade, and patient age (**Fig. 1B** and **Table 1**).

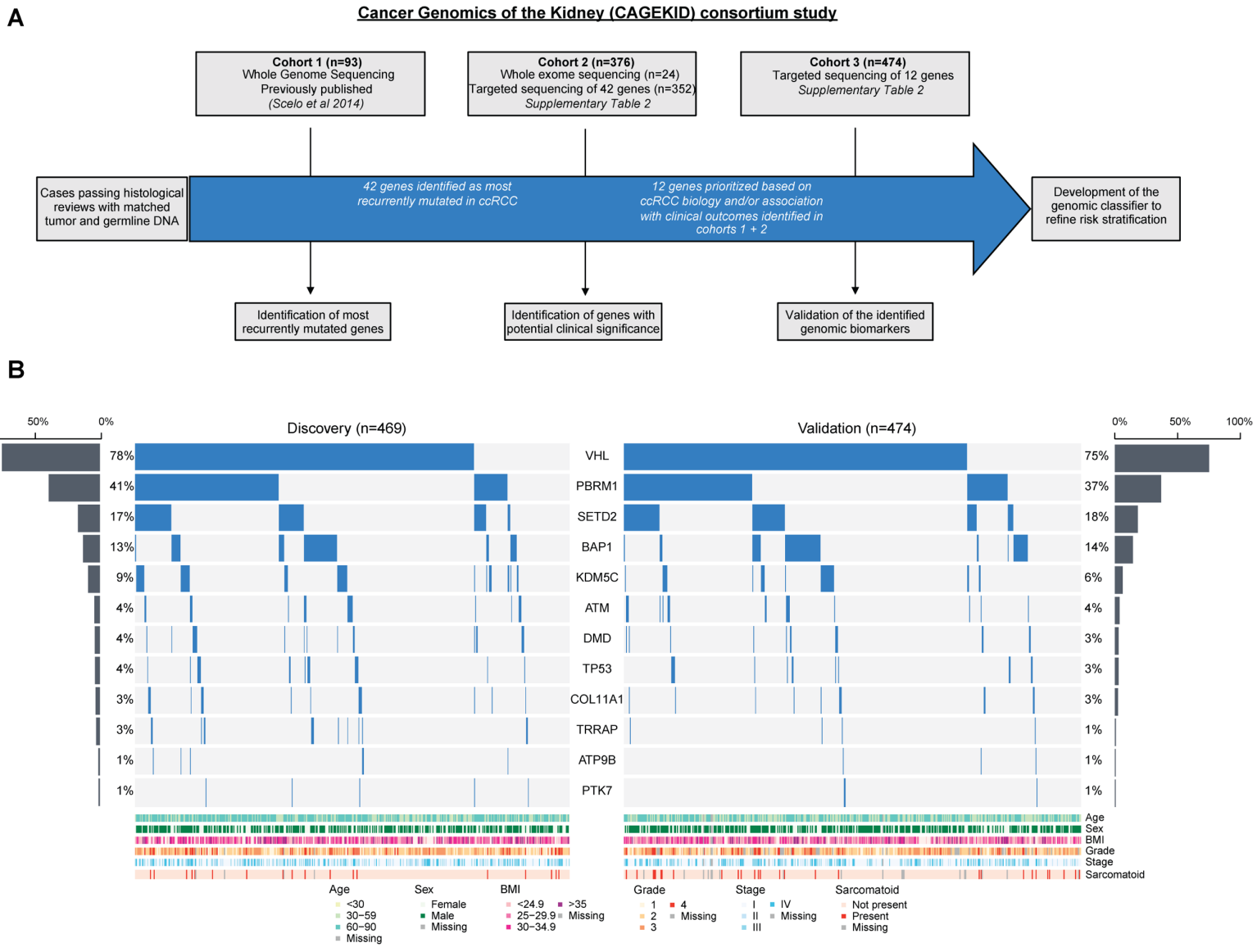


Figure 1. Study summary (A) and mutational profiling of the Discovery and Validation cohorts (B).

3.5.3 Association of gene mutation status with survival

In multivariable models stratified for TNM stage, and incorporating patient age, mutations in *BAP1* were found to be significantly associated with DFS in the Discovery cohort ($q = 0.02$). However, no genes retained significance in both cohorts once adjusted for multiple testing (**Supplementary Table S5**). When conducting a competing risks analysis using cause-specific hazards models, no genes retained significant associations with RCC-specific survival in either cohort (**Supplementary Tables S6** and **S7**).

3.5.4 Genomically-defined subgroups in ccRCC

Following the recently proposed evolutionary trajectory of RCC [**15**], we next explored the classification of tumors based on the number of identified mutated genes to create genomically defined groups. Specifically, we focused on cases with a *VHL* mutation ($n = 720$) detected in isolation (*VHL*+0; $n = 245$; 26%) versus those with a *VHL* mutation plus other driver events; *VHL*+1 ($n = 284$; 30%), *VHL*+2 ($n = 148$; 16%), and *VHL*+≥3 ($n = 43$; 5%). *VHL* WT cases ($n = 218$; 23%) were excluded from the analysis due to the potential differences in biological drivers and associated clinical behavior that are suggested to exist between *VHL*-mutated and *VHL* WT ccRCCs, including the prevalence of sarcomatoid features and copy-number aberrations in tumors that are *VHL* WT [**15**].

The characteristics of tumors in each group are presented in **Table 2**, illustrating the ability to identify clinically distinct subgroups of ccRCC. Tumors with a *VHL* mutation alone (*VHL*+0) were predominantly composed of stage I and II ($n = 175$; 72%), low-grade (grade 1 or 2; $n = 154$; 63%) cancers, with none (0/245) reported as containing sarcomatoid and/or rhabdoid change. By comparison, half of tumors containing a *VHL* mutation plus 2 or more other mutations (*VHL*+2 and *VHL*+≥3) were stage III or IV or high-grade cancers ($n = 96$; 50%) and 9.9% (19/191)

contained sarcomatoid and/or rhabdoid change. Tumors containing a *VHL* mutation plus only one additional mutation were a mix of low-stage (*n* = 177; 63%) and high-stage cancers (*n* = 88; 31%), with 4.2% (12/284) reported as containing sarcomatoid and/or rhabdoid change.

Table 2. Clinical characteristics of cases by gene group.

		Genomically-defined groups				
Characteristic		**VHL+0 (n=245)**	**VHL+1 (n=284)**	**VHL+2 (n=148)**	**VHL+≥3 (n=43)**	**P[b]**
Age at surgery (years)	Median (range)	60 (23-86)	62 (26-85)	63 (38-83)	64 (43-83)	<0.001
Sex	Female	86 (35.1)	118 (41.5)	46 (31.1)	16 (35.2)	0.157
	Male	158 (64.5)	165 (58.1)	102 (68.9)	27 (64.8)	
	Missing	1 (0.4)	1 (0.4)	0 (0)	0 (0)	
Pathological tumour size (mm)	Median (range)	50 (15-160)	55 (12-170)	60 (22-225)	65 (18-165)	0.002
Fuhrman Grade	1	30 (12.2)	46 (16.2)	11 (7.4)	3 (7.0)	<0.001
	2	124 (50.6)	116 (40.8)	61 (41.2)	12 (27.9)	
	3	75 (30.6)	80 (28.2)	42 (28.4)	18 (41.9)	
	4	8 (3.3)	25 (8.8)	27 (18.2)	9 (20.9)	
	Missing	8 (3.3)	17 (6.0)	7 (4.7)	1 (2.3)	
Overall Stage (TNM)	I	147 (60.0)	148 (52.1)	66 (44.6)	16 (37.2)	<0.001
	II	28 (11.4)	24 (8.5)	11 (7.4)	2 (4.7)	
	III	43 (17.6)	71 (25.0)	43 (29.1)	17 (39.5)	
	IV	26 (10.6)	39 (13.7)	28 (18.9)	8 (18.6)	
	Missing	1 (0.4)	2 (0.7)	0 (0)	0 (0)	
Sarcomatoid and/or rhabdoid change[a]	Present	0 (0)	12 (4.2)	15 (10.1)	4 (9.3)	<0.001
	Absent	238 (97.1)	264 (93.0)	130 (87.8)	39 (90.7)	
	Missing	7 (2.9)	8 (2.8)	3 (2.0)	0 (0)	

Note: *n* (%) unless otherwise stated.
[a]Seven tumors containing sarcomatoid and/or rhabdoid change were reported as grade 2 or grade 3 by original diagnostic pathologist.
[b]Based on comparisons using Kruskal–Wallis test.

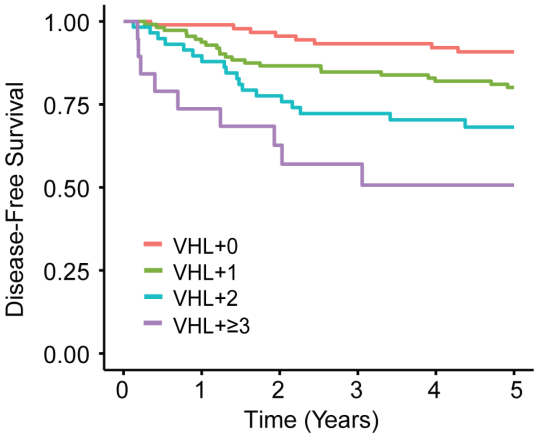
3.5.5 Survival outcomes amongst genomically-defined groups

We explored whether consideration of these gene groups would allow stratification of patients by outcome. Within both the Discovery and Validation cohorts, we observed that with an increasing number of mutations in driver genes, the risk of disease recurrence increases. In the Discovery cohort, the 5-year DFS rate was 50.7% [95% confidence interval (CI), 32–80%] for patients with *VHL*+≥3 tumors, 68.2% (95% CI, 57%–82%) for patients with *VHL*+2 tumors, and 80.1% (95% CI, 73%–88%) for patients with *VHL*+1 tumors, compared with 90.8% (95% CI, 85%–97%) for patients with only mutations in *VHL* (**Fig. 2A**). A similar trend was observed within the Validation cohort, with 5-year DFS rates of 61.5%, 73.7%, 84.7%, and 90.4% for patients with *VHL*+≥3, *VHL*+2, *VHL*+1, and *VHL*+0 tumors, respectively (**Fig. 2B**). Furthermore, this association remained independently significant when accounting for stage and patient age. We observed within both cohorts increasing risk of disease recurrence from *VHL*+1 to *VHL*+≥3 (**Figs. 2A** and **B**), and association with disease recurrence was significant for *VHL*+2 [Discovery HR = 4.3 (1.8–10.2), P = 0.000862; Validation HR = 2.5 (1.1–5.7), P = 0.02883] and *VHL*+≥3 [Discovery HR = 6.7 (2.4–18.3), P = 0.000212; Validation HR = 4.6 (1.5–13.5), P = 0.00615] groups (**Figs. 2A** and **B**). These observations were independently replicated amongst the 247 *VHL*-mutated ccRCCs from the TCGA dataset (ref. **12**; **Supplementary Fig. S1**).

We observed a similar association between these genomically defined groups and risk of RCC-related death. A competing-risks analysis showed that risk of RCC-related death increases with the number of mutated genes, whereas the cumulative incidence curves for patients with *VHL*+0 tumors were almost identical for RCC-related death and death from other causes (**Fig. 2C**). This pattern was also demonstrated in the Validation dataset (**Fig. 2D**). When stratifying for tumor stage, and considering patient age as a covariate, a cause-specific hazards model showed

increasing risk with additional driver mutations and a significant association with RCC-related death for *VHL*+2 [HR = 3.4 (1.6–7.2), $P = 0.00190$] and *VHL*+≥3 groups [HR = 4.1 (1.6–10.5), $P = 0.00286$], which was also seen in the Validation cohort (**Figs. 2C** and **D**). This trend was not observed for non-RCC–related death.

A



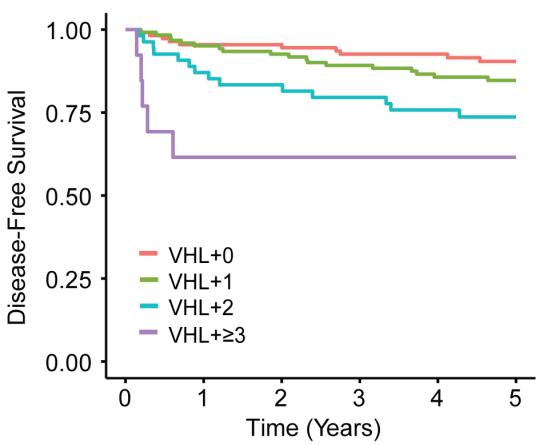
Number at risk

VHL+0	97	92	85	80	75	73
VHL+1	113	105	95	92	88	84
VHL+2	60	52	44	39	32	31
VHL+≥3	21	14	11	9	8	8

0 1 2 3 4 5
Time (Years)

Group	Hazard Ratio	95% CI	P Value
VHL+0	1	-	-
VHL+1	2.4	1.1-5.6	3.5×10^{-02}
VHL+2	4.3	1.8-10.2	8.6×10^{-04}
VHL+≥3	6.7	2.4-18.3	2.1×10^{-04}

B



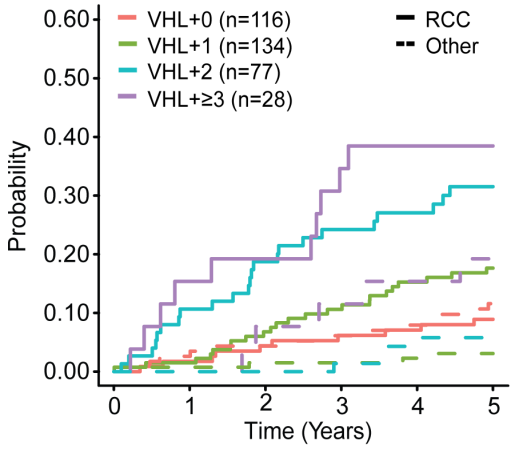
Number at risk

VHL+0	111	105	101	96	88	73
VHL+1	123	115	111	104	95	82
VHL+2	55	47	44	42	36	27
VHL+≥3	13	8	8	8	8	8

0 1 2 3 4 5
Time (Years)

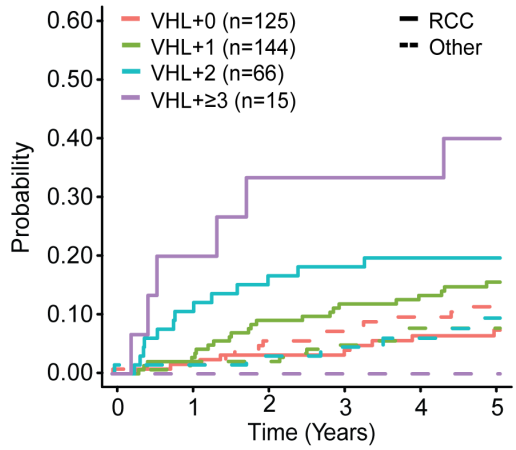
Group	Hazard Ratio	95% CI	P Value
VHL+0	1	-	-
VHL+1	1.4	0.64-3.03	4.0×10^{-01}
VHL+2	2.5	1.1-5.7	2.9×10^{-02}
VHL+≥3	4.6	1.5-13.5	6.2×10^{-03}

C



	Group	Hazard Ratio	95% CI	P Value
RCC-related	VHL+0	1	-	-
	VHL+1	1.8	0.82-3.80	1.5×10^{-01}
	VHL+2	3.4	1.6-7.2	1.9×10^{-03}
	VHL+≥3	4.1	1.6-10.5	2.9×10^{-03}
Other	VHL+0	1	-	-
	VHL+1	0.19	0.06-0.59	4.2×10^{-04}
	VHL+2	0.38	0.12-1.18	9.5×10^{-02}
	VHL+≥3	1.6	0.55-4.67	3.9×10^{-01}

D



	Group	Hazard Ratio	95% CI	P Value
RCC-related	VHL+0	1	-	-
	VHL+1	1.9	0.89-4.25	9.4×10^{-02}
	VHL+2	2.2	0.92-5.20	7.7×10^{-02}
	VHL+≥3	9.6	3.2-28.9	5.2×10^{-05}
Other	VHL+0	1	-	-
	VHL+1	0.70	0.32-1.56	3.8×10^{-01}
	VHL+2	0.85	0.32-2.24	7.4×10^{-01}
	VHL+≥3	2.4×10^{-08}	0	9.9×10^{-01}

Figure 2. DFS outcomes and Competing Risks Analysis for RCC-related death amongst patients with VHL mutations stratified into genomically defined groups. Kaplan–Meier survival curves comparing DFS amongst *VHL*+0, *VHL*+1, *VHL*+2, and *VHL*+≥3 groups within the (A) Discovery and (B) Validation cohorts. Cox PH models estimating association between genomically defined groups and 5-year DFS within the Discovery (left) and Validation (right) cohorts. Cumulative incidence functions amongst *VHL*+0, *VHL*+1, *VHL*+2, and *VHL*+≥3 groups comparing risk of death caused by RCC (solid line) compared with other causes (dotted line) within the (C) Discovery and (D) Validation cohorts. Cox PH models estimating association between genomically defined groups and 5-year RCC-related death compared with death from other causes within the Discovery (left) and Validation (right) cohorts.

Because tumors containing both a *BAP1* and *PBRM1* mutation are known to be associated with poor outcomes [**28, 29**], we examined the effects of mutations in *PBRM1*, *BAP1*, and in both genes on DFS in patients with somatic *VHL* mutations in our cohort. In line with previous reports, we observed poorer DFS when tumors harbored *PBRM1* [HR 1.72 (95% CI, 1.09–2.72); $P = 0.20$] or *BAP1* mutations [HR 2.59 (95% CI, 1.51–4.44); $P = 0.0006$], compared with those without these mutations, whereas patients with co-occurrence of *BAP1* and *PBRM1* mutations were associated with the poorest survival [HR 7.29 (95% CI, 3.03–17.54); $P < 0.0001$; **Fig. 3A**]. To assess the added value of our classifier to these known associations, we divided patients with *VHL*+2 tumors into two groups: those with *PBRM1* and *BAP1* mutations and those whose tumors harbor mutations in any two other genes from our 12-gene classifier, except both *PBRM1* and *BAP1*. We then compared DFS outcomes of these groups with that of patients with *VHL*+0 status. We observed that whilst *BAP1/PBRM1*–mutated tumors are associated with a higher risk of recurrence [HR 12.08 (95% CI, 3.95–136.95); $P < 0.0001$], remaining *VHL*+2 tumors continue to be associated with significantly poorer outcomes [HR 2.74 (95% CI, 1.49–5.01); $P = 0.001$; **Fig. 3B**; Supplementary Table S8]. Given that *VHL*+2 tumors account for approximately 20% of all classifiable tumors in our cohort, and *PBRM1*/*BAP1* co-occurrence accounts for just 7% (8/115) of these cases, our genomic classifier allows for the stratification of a greater proportion of patients by outcome. We performed similar analyses for tumors containing both a *PBRM1* and *SETD2* mutation [**30**] and observed comparable results (**Fig. 3C** and **D**; **Supplementary Table S9**).

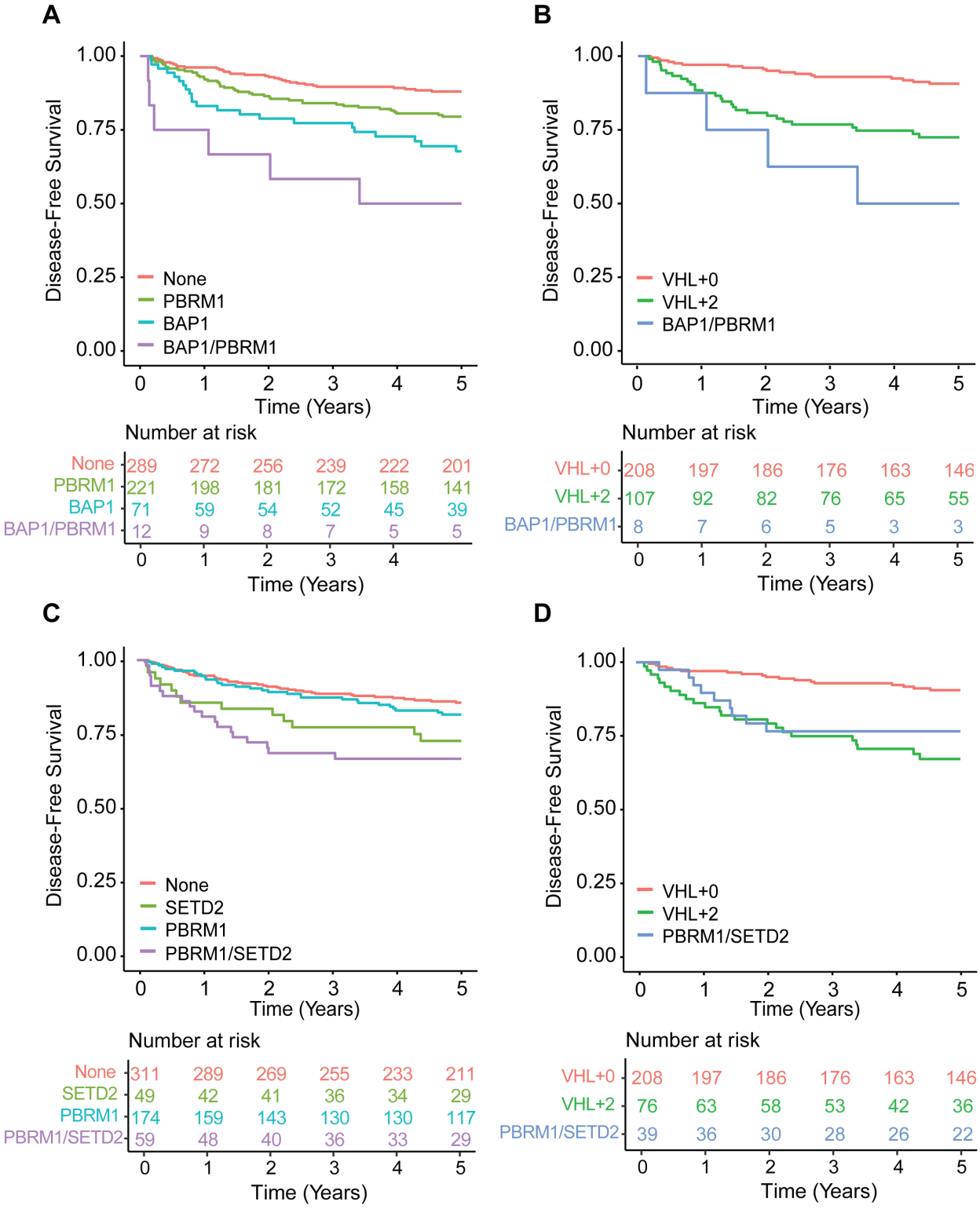

A
Disease-Free Survival
1.00
0.75
0.50
0.25
0.00
None
PBRM1
BAP1
BAP1/PBRM1
0 1 2 3 4 5
Time (Years)
Number at risk
None 289 272 256 239 222 201
PBRM1 221 198 181 172 158 141
BAP1 71 59 54 52 45 39
BAP1/PBRM1 12 9 8 7 5 5
0 1 2 3 4
Time (Years)
B
Disease-Free Survival
1.00
0.75
0.50
0.25
0.00
VHL+0
VHL+2
BAP1/PBRM1
0 1 2 3 4 5
Time (Years)
Number at risk
VHL+0 208 197 186 176 163 146
VHL+2 107 92 82 76 65 55
BAP1/PBRM1 8 7 6 5 3 3
0 1 2 3 4 5
Time (Years)
C
Disease-Free Survival
1.00
0.75
0.50
0.25
0.00
None
SETD2
PBRM1
PBRM1/SETD2
0 1 2 3 4 5
Time (Years)
Number at risk
None 311 289 269 255 233 211
SETD2 49 42 41 36 34 29
PBRM1 174 159 143 130 130 117
PBRM1/SETD2 59 48 40 36 33 29
0 1 2 3 4 5
Time (Years)
D
Disease-Free Survival
1.00
0.75
0.50
0.25
0.00
VHL+0
VHL+2
PBRM1/SETD2
0 1 2 3 4 5
Time (Years)
Number at risk
VHL+0 208 197 186 176 163 146
VHL+2 76 63 58 53 42 36
PBRM1/SETD2 39 36 30 28 26 22
0 1 2 3 4 5
Time (Years)

Figure 3. DFS outcomes by *BAP1/PBRM1* and *PBRM1/SETD2* mutation status. Kaplan–Meier survival curves based on (A) *BAP1* and *PBRM1* mutation status and (B) *VHL*+0 tumors, *VHL*+2 tumors containing both a *BAP1* and *PBRM1* mutation, and remaining *VHL*+2 tumors (i.e., those not containing both a *BAP1* and *PBRM1* mutation). Kaplan–Meier survival curves based on (C) *PBRM1* and *SETD2* mutation status and (D) *VHL*+0 tumors, *VHL*+2 tumors containing both a *PBRM1* and *SETD2* mutation, and remaining *VHL*+2 tumors (i.e., those not containing both a *PBRM1* and *SETD2* mutation).

3.5.6 Risk stratification among patients eligible for adjuvant therapy

We investigated the potential utility of the genomic classifier for selection of patients for adjuvant therapies (clinically eligible defined as pT2 grade 3–4; pT3 or pT4 (any grade); any pT, any grade, N+). DFS rate at 5 years amongst the 397 patients not considered eligible for adjuvant therapy was 90.6% (95% CI, 88%–94%) versus 63.6% (95% CI, 57%–71%) for the 196 patients eligible for adjuvant therapy (**Fig. 4A**). Patients defined as being eligible for adjuvant therapy could be further stratified by risk of relapse based on the genomic classification of their tumors. Five-year DFS rates were 79.3% (95% CI, 69%–91%) amongst the 56 (29%) patients with *VHL*+0 tumors, 69.4% (95% CI, 60%–81%) amongst the 77 (39%) patients with *VHL*+1 tumors, 45.6% (95% CI, 33%–63%) amongst the 46 (23%) patients with *VHL*+2 tumors, and 35.3% (95% CI, 19%–67%) amongst the 17 (9%) patients with *VHL*+≥3 tumors (**Fig. 4B**).

Notably, the *VHL*+2 and *VHL*+≥3 groups had significantly poorer survival compared with the *VHL*+0 group ($P = 0.00055$ and $P < 0.0001$, respectively). The potential clinical application of these findings, to inform individual patient counselling and decision making, is depicted in **Fig. 4C**. Groups of patients who may be both spared (*VHL*+0) versus prioritized (*VHL*+2 and *VHL*+≥3) for adjuvant treatment, based on risk of recurrence, are identified.

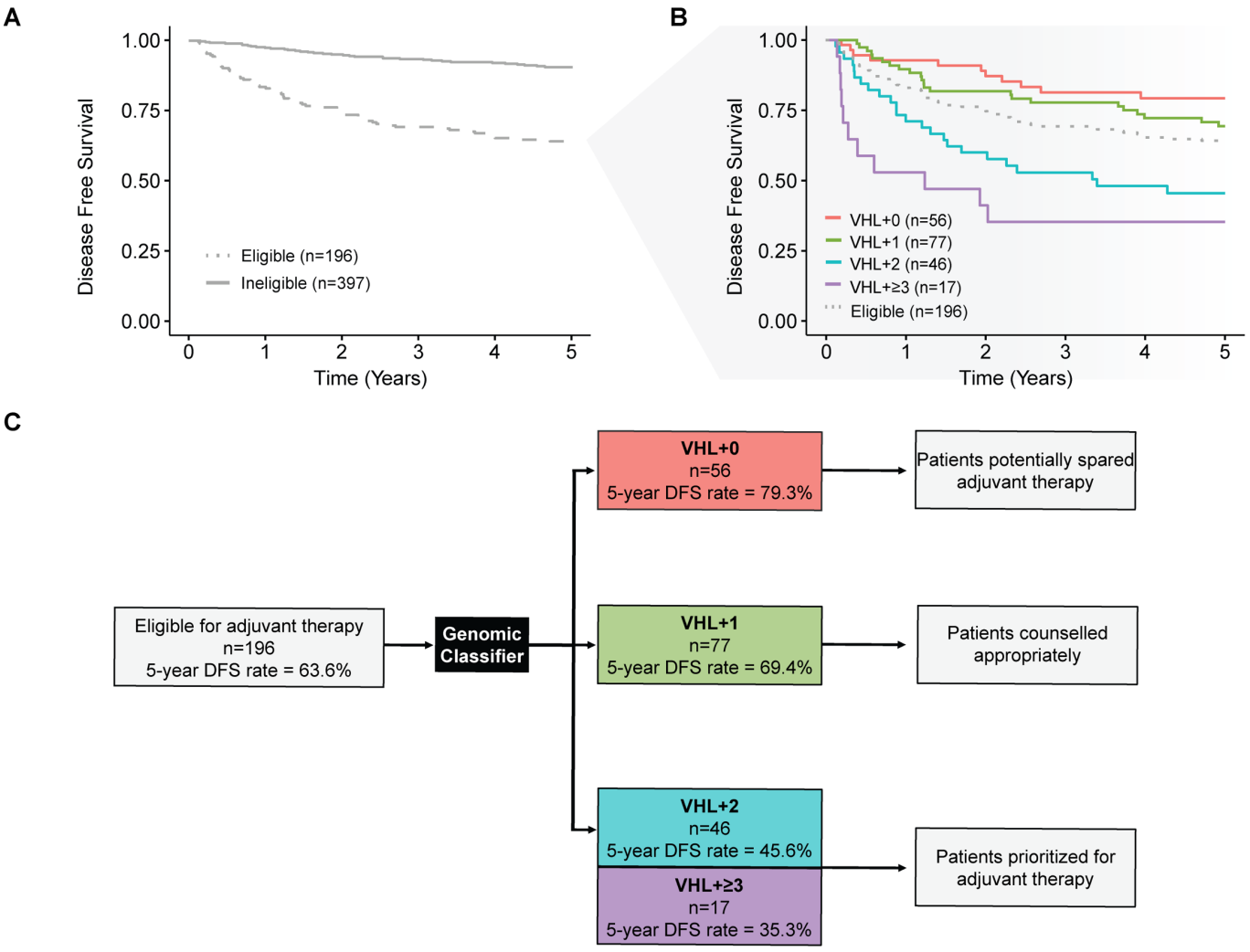


Figure 4. Patients eligible for adjuvant therapy stratified by the genomic classifier. A, DFS amongst patients considered eligible [pT2 grade 3–4; pT3 or pT4 (any grade); any pT, any grade, N+] versus ineligible [pT1 (any grade); pT2 grade 1–2] for adjuvant therapy. **B**, DFS by *VHL*+0, *VHL*+1, *VHL*+2, and *VHL*+≥3 groups amongst patients considered eligible for adjuvant therapy. **C**, Flow diagram demonstrating potential clinical application. Application of the genomic classifier to patients typically considered eligible for adjuvant therapy allows sub-stratification of patients into groups with highly divergent risk of relapse. This information could usefully inform individual patient discussions around the benefit versus risks of adjuvant therapy.

We also explored the added utility of our genomic classifier when first considering the Leibovich score, a validated prognostic nomogram that incorporates tumor stage, grade, size, necrosis, and lymph node status [**31**], available in 181/196 (92%) patients. As expected, patients with high-risk disease had significantly poorer DFS rates in comparison with those with intermediate-risk disease ($P < 0.001$; **Supplementary Fig. S2A**; **Supplementary Table S10**). We found that our genomic classifier retains the ability to meaningfully substratify within these groups. For example, amongst patients with Leibovich score defined intermediate-risk disease, *VHL*+0 patients had a 5-year DFS rate of 89% (95% CI, 80%–99%) compared with 66% (95% CI, 47%–92%) for patients with *VHL*+2 tumors ($P = 0.02$) and 25% (95% CI, 5%–100%) for *VHL*+≥3 tumors ($P < 0.0001$), although it should be noted there were only 4 patients in the latter group (**Supplementary Fig. S2B**).

Finally, we examined outcomes by our classifier amongst patients with stage I disease, who are not usually considered for adjuvant therapy. Patients with stage I *VHL*+0 tumors (140/365; 38%) were associated with excellent outcomes, with a 5-year DFS rate of 96% (95% CI, 93%–99%). By contrast, patients with a stage I *VHL*+≥3 tumor had relatively poorer outcomes, with a 5-year DFS of 78% (95% CI, 59%–100%; $P = 0.01$), although the wide CIs reflect the small number of patients in this group (16/365; 4%) and should be considered an exploratory finding. Outcomes by tumor stage stratified by our genomic classifier are shown in **Supplementary Fig. S3** and **Supplementary Table S11**.

3.5.7 Genomic classifier independence from tumor mutational burden

To investigate whether the increasing number of driver mutations in RCC-focused genes act as a classifier independent of tumor mutational burden (TMB), a known prognostic marker in some cancers, we compared TMB values between the genomically defined groups. TMB differed

only between the *VHL*+0 and other groups within the WGS samples (Cohort 1). Notably, TMB was not significantly different between *VHL*+1, *VHL*+2, and *VHL*+≥3 tumors (**Supplementary Table S12**). This analysis was replicated using the publically available TCGA dataset, which showed only marginally significant difference between the *VHL*+0 and *VHL*+2 groups, and no differences between the *VHL*+1, *VHL*+2, and *VHL*+≥3 tumors (**Supplementary Table S12**). In addition, whereas TMB alone did not show a significant association to DFS, when incorporating TMB as a covariate, *VHL*+2 and *VHL*+≥3 groups showed significantly poorer DFS in both the C1 cohort ($P = 0.025$ and $P = 0.043$ for *VHL*+2 and *VHL*+≥3, respectively) and TCGA ($P = 0.002$ and $P = 0.046$ for *VHL*+2 and *VHL*+≥3, respectively) cohorts (**Supplementary Table S13**). These results demonstrate that the performance of the classifier is independent from TMB.

3.6 DISCUSSION

The prospect of offering effective adjuvant therapy to patients with RCC represents a major step forward in the management of this disease. With this, however, comes the challenge of deciding who should or should not be treated, based on risk of cancer recurrence. In this multinational study, comprising final data from a total of 943 patients, we have examined the role of targeted gene sequencing to improve individual risk stratification. We show that by considering a panel of 12 RCC-focused genes, clinically and biologically distinct groups can be identified, accounting for 76% of cases in our cohort, that can be used to refine individual risk-estimates with potential immediate application for selection of patients for adjuvant therapy. Importantly, these findings were observed in both the Discovery and Validation cohorts, and were independent from patient age, tumor stage, and TMB.

Studies examining the clinical impact of somatic mutations in localized ccRCC have consistently revealed the association of mutations in specific genes, such as *BAP1* and *SETD2*,

with poorer outcomes [**28, 32**]. However, few studies have examined their independent prognostic value. In a pooled analysis of 1,049 patients with ccRCC, incorporating four independent cohorts (including patients from C1 of the current study), only mutations in *SETD2* remained marginally significant for recurrence-free survival in a model including TNM stage and patient age [**33**]. We found mutations in *BAP1*, but not *SETD2*, to retain significance on multivariate testing when considering DFS. However, this finding failed to replicate between our Discovery and Validation cohorts. Given these inconsistencies, the value of considering the mutation status of any single gene to inform clinical practice remains uncertain.

The existence of seven distinct, genomically defined, evolutionary subtypes of ccRCC was proposed in a study employing multi-region sampling of primary tumors, elegantly described amongst 63 cases [**15**]. These included tumors containing multiple clonal drivers (n = 12; defined as tumors with mutations in 2 or more of *BAP1*, *PBRM1*, *SETD2*, or *PTEN* clonal mutations), demonstrating limited intra-tumor heterogeneity (ITH) and associated exclusively with higher-stage disease. Another 11 tumors were defined as *VHL* monodrivers, again demonstrating limited subclonality and predominantly composed of stage I cancers. A third group (n = 6), defined as *VHL* WT, were associated with increased somatic copy-number alterations and enriched for tumors containing sarcomatoid differentiation [**15**].

These intriguing exploratory observations led us to examine our data using a similar principle, in our much larger sample set, but subject to single region sequencing. Most patients within the CAGEKID cohort demonstrated a mutation in the *VHL* gene (n = 718; 76%) and could be classified as those with tumors containing a *VHL* mutation alone (*VHL*+0), or a *VHL* mutation plus additional driver events (*VHL*+1, *VHL*+2, *VHL*+≥3), within the sequenced region and equate to the '*VHL* monodriver' and 'multiple clonal driver' subtypes proposed above. These genomic

groups showed striking divergence in terms of their clinical behavior. Importantly, whilst a preponderance of either early stage, low grade, or high stage, high grade cancers were found in *VHL*+0 and *VHL*+2/*VHL*+≥3 tumors, respectively, no exclusivity was observed and all groups were represented across the disease spectrum.

Determination of recurrence risk in localized RCC currently remains reliant on clinicopathologic criteria alone. TNM stage and tumor grade provide useful broad stratification of patients into low-, intermediate-, and high-risk groups, but ultimately fail to adequately account for individual variance in tumor biology and outcomes. Efforts to improve risk prediction has led to the development of several prognostic nomograms [**31, 34, 35**], incorporating additional tumor or clinical characteristics, which have also been widely adopted. However, their predictive accuracy appears, at best, only marginally better than TNM alone and their performance has declined over time [**36–38**].

These deficiencies lead to a significant risk that many patients may be overtreated in the adjuvant setting. The ability of the current genomic classifier to identify over a quarter of such patients (*VHL*+0) with a 5-year DFS rate approaching 80% is therefore important. Potential avoidance of adjuvant therapy carries significant benefits to both patients and healthcare systems, given the costs and toxicity of ICI-based therapies. Conversely, patients with tumors containing multiple driver mutations appear to be at extremely high risk of recurrence (5-year DFS rate 35%–46%), representing a group who may benefit most from adjuvant treatment. Our findings also carry important implications for the design, costs, and success of future adjuvant RCC trials.

This study adds to a growing literature in ccRCC demonstrating the ability to infer tumor biology from genomic data derived from a single tumor region [**33, 39–42**]. The impact of spatial ITH, well characterized in ccRCC [**43**], on clinical association studies such as ours remains poorly

understood. Whilst multi-region sequencing would almost certainly increase mutation detection [**15**], in all but the *VHL* gene, the clinical impact of a small subclonal driver event within the context of an otherwise largely clonal tumor, for example, remains unknown [**44**].

Our gene panel is not definitive and it is likely that further refinement is possible. For example, other than *COL11A1*, we did not consider genes within the PI3K/AKT/mTOR pathway. Furthermore, the significance, at an individual level, of the lower frequency events included in our panel could not be robustly established despite the size of our cohort and will require even larger studies to achieve this. Analysis of other known genomic features, such as copy-number alterations, were not undertaken and may allow further refinement of genomic groups. Whilst *VHL*+2 and *VHL*+≥3 tumors are associated with the highest risk of disease recurrence, the benefit of adjuvant ICI in these patients remains unknown and warrants investigation. The TCGA sample set represents the next largest available cohort of genetically defined ccRCC and was useful in validating our findings and in investigating associations to TMB, but ideally a larger cohort, with a similarly high *VHL* mutation rate to our study (notably the TCGA cohort reported a rate of *VHL* mutation of 52.3%; ref. **14**), would have been employed.

In conclusion, this study establishes the ability to define biologically distinct molecular subgroups of ccRCC that could be used to better inform patients and their physicians regarding individual risk of tumor recurrence following nephrectomy. These genetic groups can be defined based on the mutation status of a small panel of genes captured within a single tumor region and, therefore, readily applied to the clinic. Further prospective evaluation of these findings is warranted.

3.7 STUDY CENTERS

Patients were recruited from the following centers (with lead investigator(s) at each site): St James's University Hospital, Leeds, Dr. Naveen Vasudev; Newcastle Upon Tyne Hospitals NHS Foundation Trust, Professor Naeem Soomro; Stockport NHS Foundation Trust, Mr. Adebanji Adeyoju; Nottingham University Hospitals NHS Trust, Professor Poulam Patel; NHS Lothian, Professor Grant Stewart; Charing Cross Hospital, Mr. David Hrouda; Oxford University Hospitals NHS Foundation Trust, Mr. Mark Sullivan; Northwick Park Hospital, Mr. Jeff Webster; University Hospital Motol, Dr. Antonin Brisuda; General University Hospital, Dr. Roman Sobotka; Masaryk Memorial Cancer Institute, Dr. Lenka Foretova; Palacký University Hospital, Dr. Vladimir Janout; České Budějovice Regional Hospital, Dr. Vladimir Janout; Th. Burghele Hospital, Bucharest, Dr. Viorel Jinga; N.N. Blokhin Cancer Research Center, Dr. David Zaridze; Clinical Center of Serbia (KCS), Dr. Ljiljana Bogdanovic; Military Medical Academy, Dr. Bozidar Kovacevic.

3.8 AUTHOR DISCLOSURES

N.S. Vasudev reports grants, personal fees, and nonfinancial support from Bristol Myers Squibb; personal fees and nonfinancial support from EUSA Pharma and Ipsen; and personal fees from Eisai Ltd, Merck Serono, Pfizer, and 4D Pharma outside the submitted work. M.B. Wozniak reports personal fees from Novartis Ireland Ltd outside the submitted work. J. Tost reports grants from European Union during the conduct of the study, as well as grants from European Union, Agence National de Recherche (ANR), Fondation pour la Recherche Medicale (FRM), DBV Technologies, MSD Avenir, and Agence Nationale Sécurité Sanitaire Alimentaire Nationale (Anses) outside the submitted work. Y. Riazalhosseini reports grants from Cancer Research Society, Kidney Foundation of Canada, and Fonds de recherche du Québec – Santé (FRQS) during

the conduct of the study. R.E. Banks reports grants from EU and NIHR during the conduct of the study. No disclosures were reported by the other authors.

3.9 DISCLAIMER

The views and opinions expressed by the authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the UK Department of Health, or the International Agency for Research on Cancer/World Health Organization.

3.10 ACKNOWLEDGEMENTS

We thank all the staff at each of the recruiting centers, including the technical support provided by staff of genomic facilities of McGill University and Genome Québec Innovation Centre. We thank Drs. Patricia Harnden and Morag Seywright (both now retired) for their substantial contributions to panel pathology review. We are grateful to the patients for consenting to take part in this study.

This work was supported by the EU FP7 under grant agreement number 241669 (the CAGEKID project, **http://www.cng.fr/cagekid**) and grants from Génome Québec, le Ministère de l'Enseignement supérieur, de la Recherche, de la Science et de la Technologie (MESRST) Québec, Cancer Research Society (CRS 22592), Kidney Foundation of Canada (2020KHRG-673291), and McGill University. Y. Riazalhosseini is a research scholar of the Fonds de recherche du Québec – Santé (FRQS). The study in Brno, Czech Republic, was supported by MH CZ - DRO (MMCI, 00209805). A.Y. Warren is supported by the NIHR Cambridge Biomedical Research Centre. We also acknowledge the support of Cancer Research UK, Experimental Cancer Medicine Centre and National Institute for Health Research (NIHR) Clinical Research Facility infrastructure funding in Leeds, the Leeds Multidisciplinary and NIHR Research Tissue Banks and are grateful

to the sample processing, urology, pathology, and oncology clinical teams at St James's University Hospital, Leeds.

The publication costs of this article were defrayed in part by the payment of publication fees. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734.

3.11 SUPPLEMENTAL MATERIALS

3.11.1 Supplementary Tables

The below supplementary tables are available online at:
https://aacrjournals.org/clincancerres/article/29/7/1220/718766/Application-of-Genomic-Sequencing-to-Refine/

Supplementary Table S1. Clinical information and 12 gene mutation status for cohorts C1-C3

Supplementary Table S2. Somatic mutations detected in 12 interrogated genes for cohorts C1-C3

Supplementary Table S3. Patient and tumor characteristics

Characteristic		All n=943	Discovery	Validation
Age at surgery (years)	median (range)	61.42 (23.4-86.1)	61.7 (23.4-86.1)	61.1 (25.6-85.1)
Sex n (%)	Female	359 (38.1)	193 (41.2)	166 (35.0)
	Male	580 (61.5)	276 (58.8)	304 (64.1)
	Missing	4 (0.4)	0 (0)	4 (0.8)
Body mass index (BMI)	median (range)	27.8 (14.9)	27.8 (14.9-49.6)	27.7 (16-49.2)
BMI categorised n (%)	<24.9	222 (23.5)	102 (21.7)	120 (25.3)
	25-29.9	344 (36.5)	167 (35.6)	177 (37.3)
	30-34.9	199 (21.1)	101 (21.5)	98 (20.7)
	>35	84 (8.9)	44 (9.4)	40 (8.4)
	Missing	94 (10.0)	55 (11.7)	39 (8.2)
Tobacco exposure n (%)	Never	468 (49.6)	245 (52.2)	223 (47.0)
	Ex-smoker	258 (27.4)	115 (24.5)	143 (30.2)
	Current smoker	206 (21.8)	105 (22.4)	101 (21.3)
	Missing	11 (1.2)	4 (0.9)	7 (1.5)
Country	Czech Republic	342 (36.3)	133 (28.4)	209 (44.1)
	UK	291 (30.9)	150 (32.0)	141 (29.7)
	Russia	197 (20.9)	129 (27.5)	68 (14.3)
	Romania	75 (8.0)	50 (10.7)	25 (5.3)
	Serbia	34 (3.6)	7 (1.5)	27 (5.7)
	Missing	4 (0.4)	0 (0)	4 (0.8)
Pathological tumor size (mm)	median (range)	55 (13-220)	58 (13)	54.5 (12-225)
Pathological tumor stage	1a	277 (29.4)	128 (27.3)	149 (31.4)
	1b	215 (22.8)	114 (24.3)	101 (21.3)
	2	107 (11.3)	57 (12.2)	50 (10.5)
	3	288 (30.5)	155 (33.0)	133 (28.1)
	4	10 (1.1)	7 (1.5)	3 (0.6)
	Missing	46 (4.9)	8 (1.7)	38 (8.0)
pN stage n (%)	0/X	866 (91.8)	449 (95.7)	417 (88.0)
	1	37 (3.9)	16 (3.4)	21 (4.4)
	Missing	40 (4.2)	4 (0.9)	36 (7.6)
pM stage n (%)	0/X	823 (87.3)	430 (91.7)	393 (82.9)
	1	83 (8.8)	37 (7.9)	46 (9.7)
	Missing	37 (3.9)	2 (0.4)	35 (7.4)
Necrosis n (%)	Absent	678 (71.9)	360 (76.8)	318 (67.1)
	Present	228 (24.2)	108 (23.0)	120 (25.3)
	Missing	37 (3.9)	1 (0.2)	36 (7.6)

FFPE grade n (%)	1	122 (12.9)	66 (14.1)	56 (11.8)
	2	405 (42.9)	215 (45.8)	190 (40.1)
	3	279 (29.6)	139 (29.6)	140 (29.5)
	4	91 (9.7)	45 (9.6)	46 9.7)
	Missing	46 (4.9)	4 (0.9)	42 (8.9)
Overall TNM stage n (%)	I	486 (51.5)	235 (50.1)	251 (53.0)
	II	95 (10.1)	47 (10.0)	48 (10.1)
	III	225 (23.9)	117 (24.9)	108 (22.8)
	IV	132 (14.0)	70 (14.9)	62 (13.1)
	Missing	5 (0.5)	0 (0)	5 (1.1)
Rhabdoid or sarcomatoid change n(%)	Absent	874 (92.7)	447 (95.3)	427 (90.1)
	Present	46 (4.9)	18 (3.8)	28 (5.9)
	Missing	23 (2.4)	4 (0.9)	19 (4.0)
Laterality n (%)	Right kidney	436 (46.2)	224 (47.8)	212 (44.7)
	Left kidney	434 (46.0)	221 (47.1)	213 (44.9)
	Bilateral	2 (0.2)	0 (0)	2 (0.4)
	Missing	71 (7.5)	24 (5.1)	47 (9.9)
Nephrectomy type n (%)	Partial	116 (12.3)	46 (9.8)	70 (14.8)
	Radical	811 (86.0)	416 (88.7)	395 (83.3)
	Missing	16 (1.7)	7 (1.5)	9 (1.9)
Relapse n (%)	No	622 (66.0)	291 (62.0)	331 (69.8)
	Yes	160 (17.0)	84 (17.9)	76 (16.0)
	Missing	27 (2.9)	22 (4.7)	5 (1.1)
	Not applicable	134 (14.2)	72 (15.4)	62 (13.1)
Dead n (%)	No	609 (64.6)	286 (61.0)	323 (68.1)
	Yes	327 (34.7)	180 (38.4)	147 (31.0)
	Missing	7 (0.7)	3 (0.6)	4 (0.8)
RCC related death n (%)	No	125 (13.3)	62 (13.2)	63 (13.3)
	Yes	192 (20.4)	109 (23.2)	83 (17.5)
	Missing	17 (1.8)	12 (2.6)	5 (1.1)
	Not applicable	609 (64.6)	286 (61.0)	323 (68.1)
Total mutations* (out of 12) n (%)	0	91 (9.7)	44 (9.4)	47 (9.9)
	1	331 (35.1)	154 (32.8)	177 (37.3)
	2	319 (33.8)	155 (33.0)	164 (34.6)
	3	158 (16.8)	87 (18.6)	71 (15.0)
	4	40 (4.2)	27 (5.8)	13 (2.7)
	5	3 (0.3)	1 (0.2)	2 (0.4)
	6	1 (0.1)	1 (0.2)	0 (0)

Supplementary Table 4. Gene mutation frequencies across cohorts

Gene	Mutated (N/Y)	All n=943 n (%)	Cohort 1 WGS n=93 n (%)	Cohort 2 42 genes n=376 n (%)	Cohort 3 12 genes n=474 n (%)
ATM	No	907 (96.2)	88 (94.6)	363 (96.5)	456 (96.2)
	Yes	36 (3.8)	5 (5.4)	13 (3.5)	18 (3.8)
ATP9B	No	935 (99.2)	90 (96.8)	374 (99.5)	471 (99.4)
	Yes	8 (0.8)	3 (3.2)	2 (0.5)	3 (0.6)
BAP1	No	812 (86.1)	82 (88.2)	324 (86.2)	406 (85.7)
	Yes	131 (13.9)	11 (11.8)	52 (13.8)	68 (14.3)
COL11A1	No	915 (97.0)	87 (93.5)	366 (97.3)	462 (97.5)
	Yes	28 (3.0)	6 (6.5)	10 (2.7)	12 (2.5)
DMD	No	909 (96.4)	88 (94.6)	361 (96.0)	460 (97)
	Yes	34 (3.6)	5 (5.4)	15 (4.0)	14 (3)
KDM5C	No	869 (92.2)	86 (92.5)	339 (90.2)	444 (93.7)
	Yes	74 (7.8)	7 (7.5)	37 (9.8)	30 (6.3)
PBRM1	No	579 (61.4)	57 (61.3)	221 (58.8)	301 (63.5)
	Yes	364 (38.6)	36 (38.7)	155 (41.2)	173 (36.5)
PTK7	No	934 (99.0)	90 (96.8)	373 (99.2)	471 (99.4)
	Yes	9 (1.0)	3 (3.2)	3 (0.8)	3 (0.6)
SETD2	No	774 (82.1)	75 (80.6)	312 (83.0)	387 (81.6)
	Yes	169 (17.9)	18 (19.4)	64 (17.0)	87 (18.4)
TP53	No	910 (96.5)	89 (95.7)	361 (96.0)	460 (97)
	Yes	33 (3.5)	4 (4.3)	15 (4.0)	14 (3)
TRRAP	No	925 (98.1)	87 (93.5)	368 (97.9)	470 (99.2)
	Yes	18 (1.9)	6 (6.5)	8 (2.1)	4 (0.8)
VHL	No	220 (23.3)	14 (15.1)	86 (22.9)	120 (25.3)
	Yes	720 (76.4)	76 (81.7)	290 (77.1)	354 (74.7)
	Missing	3 (0.3)	3 (3.2)	0 (-)	0 (-)
ABCA13	No	(-)	89 (95.7)	352 (93.6)	(-)
	Yes	(-)	4 (4.3)	24 (6.4)	(-)
ANK2	No	(-)	89 (95.7)	361 (96.0)	(-)
	Yes	(-)	4 (4.3)	15 (4.0)	(-)
ANPEP	No	(-)	88 (94.6)	370 (98.4)	(-)
	Yes	(-)	5 (5.4)	6 (1.6)	(-)
ARID1A	No	(-)	88 (94.6)	363 (96.5)	(-)
	Yes	(-)	5 (5.4)	13 (3.5)	(-)

CACNA1I	No	(-)	89 (95.7)	366 (97.3)	(-)
	Yes	(-)	4 (4.3)	10 (2.7)	(-)
COL12A1	No	(-)	88 (94.6)	366 (97.3)	(-)
	Yes	(-)	5 (5.4)	10 (2.7)	(-)
CSMD3	No	(-)	85 (91.4)	364 (96.8)	(-)
	Yes	(-)	8 (8.6)	12 (3.2)	(-)
DNAH7	No	(-)	88 (94.6)	369 (98.1)	(-)
	Yes	(-)	5 (5.4)	7 (1.9)	(-)
DOCK8	No	(-)	89 (95.7)	372 (98.9)	(-)
	Yes	(-)	4 (4.3)	4 (1.1)	(-)
DST	No	(-)	89 (95.7)	363 (96.5)	(-)
	Yes	(-)	4 (4.3)	13 (3.5)	(-)
FAT1	No	(-)	88 (94.6)	367 (97.6)	(-)
	Yes	(-)	5 (5.4)	9 (2.4)	(-)
FAT3	No	(-)	86 (92.5)	355 (94.4)	(-)
	Yes	(-)	7 (7.5)	21 (5.6)	(-)
FAT4	No	(-)	88 (94.6)	365 (97.1)	(-)
	Yes	(-)	5 (5.4)	11 (2.9)	(-)
FRAS1	No	(-)	89 (95.7)	368 (97.9)	(-)
	Yes	(-)	4 (4.3)	8 (2.1)	(-)
GUCY1A2	No	(-)	90 (96.8)	375 (99.7)	(-)
	Yes	(-)	3 (3.2)	1 (0.3)	(-)
KMT2C	No	(-)	87 (93.5)	359 (95.5)	(-)
	Yes	(-)	6 (6.5)	17 (4.5)	(-)
LRP1B	No	(-)	88 (94.6)	363 (96.5)	(-)
	Yes	(-)	5 (5.4)	13 (3.5)	(-)
MDN1	No	(-)	88 (94.6)	364 (96.8)	(-)
	Yes	(-)	5 (5.4)	12 (3.2)	(-)
MTOR	No	(-)	86 (92.5)	353 (93.9)	(-)
	Yes	(-)	7 (7.5)	23 (6.1)	(-)
NRXN1	No	(-)	87 (93.5)	369 (98.1)	(-)
	Yes	(-)	6 (6.5)	7 (1.9)	(-)
PEAK1	No	(-)	90 (96.8)	372 (98.9)	(-)
	Yes	(-)	3 (3.2)	4 (1.1)	(-)
PIEZO2	No	(-)	87 (93.5)	367 (97.6)	(-)
	Yes	(-)	6 (6.5)	9 (2.4)	(-)
PRKDC	No	(-)	89 (95.7)	370 (98.4)	(-)
	Yes	(-)	4 (4.3)	6 (1.6)	(-)
PTEN	No	(-)	89 (95.7)	367 (97.6)	(-)

	Yes	(-)	4 (4.3)	9 (2.4)	(-)
TENM1	No	(-)	88 (94.6)	367 (97.6)	(-)
	Yes	(-)	5 (5.4)	9 (2.4)	(-)
USP24	No	(-)	88 (94.6)	368 (97.9)	(-)
	Yes	(-)	5 (5.4)	8 (2.1)	(-)
WDFY3	No	(-)	87 (93.5)	369 (98.1)	(-)
	Yes	(-)	6 (6.5)	7 (1.9)	(-)
ZAN	No	(-)	88 (94.6)	371 (98.7)	(-)
	Yes	(-)	5 (5.4)	5 (1.3)	(-)
ZFHX4	No	(-)	84 (90.3)	363 (96.5)	(-)
	Yes	(-)	9 (9.7)	13 (3.5)	(-)
ZNF469	No	(-)	86 (92.5)	361 (96.0)	(-)
	Yes	(-)	7 (7.5)	15 (4.0)	(-)

WGS – whole genome sequencing

Supplementary Table 5. Multivariable disease-free survival analysis

		Disease-free survival (Discovery)				Disease-free survival (Validation)			
Predictor	Levels	HR	95% CI	p-value	q-value	HR	95% CI	p-value	q-value
ATM	0	1.00	-	-	-	1.00	-	-	-
	1	1.13	[0.35;3.62]	0.838	1.000	0.98	[0.24;4.02]	0.975	1.000
Age		1.00	[0.98;1.02]	0.970	1.000	1.00	[0.98;1.03]	0.806	1.000
ATP9B	0	1.00	-	-	-	1.00	-	-	-
	1	4.06	[1.26;13.10]	0.019	0.455	(-)	(-)	(-)	(-)
Age		1.00	[0.98;1.02]	0.990	1.000	1.00	[0.98;1.03]	0.845	1.000
BAP1	0	1.00	-	-	-	1.00	-	-	-
	1	2.68	[1.50;4.78]	0.001	0.020	1.92	[1.11;3.30]	0.019	0.458
Age		1.00	[0.98;1.02]	0.870	1.000	1.00	[0.98;1.03]	0.727	1.000
COL11A1	0	1.00	-	-	-	1.00	-	-	-
	1	0.91	[0.22;3.72]	0.893	1.000	3.47	[1.08;11.19]	0.037	0.893
Age		1.00	[0.98;1.02]	0.951	1.000	1.00	[0.98;1.03]	0.866	1.000
DMD	0	1.00	-	-	-	1.00	-	-	-
	1	1.86	[0.68;5.13]	0.228	1.000	1.64	[0.59;4.55]	0.344	1.000
Age		1.00	[0.98;1.02]	0.990	1.000	1.00	[0.98;1.03]	0.742	1.000
KDM5C	0	1.00	-	-	-	1.00	-	-	-
	1	1.50	[0.76 ;2.96	0.240	1.000	0.86	[0.27;2.75]	0.803	1.000
Age		1.00	[0.98;1.02]	0.981	1.000	1.00	[0.98;1.03]	0.811	1.000
PBRM1	0	1.00	-	-	-	1.00	-	-	-
	1	1.22	[0.76;1.98]	0.414	1.000	1.26	[0.77;2.08]	0.355	1.000
Age		1.00	[0.98;1.02]	0.937	1.000	1.00	[0.98;1.03]	0.883	1.000
PTK7	0	1.00	-	-	-	1.00	-	-	-
	1	2.90	[0.70;11.97]	0.141	1.000	(-)	(-)	(-)	(-)
Age		1.00	[0.98;1.02]	0.962	1.000	1.00	[0.98;1.03]	0.768	1.000
SETD2	0	1.00	-	-	-	1.00	-	-	-
	1	1.56	[0.88;2.77]	0.128	1.000	2.23	[1.33;3.76	0.003	0.060
Age		1.00	[0.98;1.02]	0.769	1.000	1.00	[0.98;1.03]	0.781	1.000
TP53	0	1.00	-	-	-	1.00	-	-	-
	1	3.38	[1.35;8.45]	0.009	0.221	0.69	[0.17;2.84	0.605	1.000
Age		1.00	[0.98;1.02]	0.932	1.000	1.00	[0.98;1.03]	0.759	1.000
TRRAP	0	1.00	-	-	-	1.00	-	-	-
	1	2.65	[0.96;7.36]	0.061	1.000	1.81	[0.25;13.13	0.558	1.000
Age		1.00	[0.98;1.02]	0.924	1.000	1.00	[0.98;1.03]	0.798	1.000
VHL	0	1.00	-	-	-	1.00	-	-	-
	1	1.43	[0.76;2.67]	0.268	1.000	0.85	[0.50;1.46]	0.559	1.000
Age		1.00	[0.98;1.02]	0.870	1.000	1.00	[0.98;1.03]	0.787	1.000

* Cox PH model includes a stratification by tumor stage; Bonferroni adjustment was made to account for multiple testing

Supplementary Table 6. Multivariable cause-specific survival analysis (Discovery)

		RCC-related death				Non-RCC-related death			
Predictor	Levels	HR	95% CI	p-value	q-value	HR	95% CI	p-value	q-value
ATM	0	1.00	-	-	-	1.00	-	-	-
	1	2.66	[1.12 ;6.33]	0.027	0.638	(-)	(-)	(-)	1.000
Age		1.00	[0.98;1.03]	0.734	1.000	1.06	[1.02;1.09]	0.003	0.079
ATP9B	0	1.00	-	-	-	1.00	-	-	-
	1	6.21	[1.87;20.68]	0.003	0.070	2.90	[0.39;21.64]	0.299	1.000
Age		1.01	[0.98;1.03]	0.541	1.000	1.05	[1.02;1.10]	<0.005	0.111
BAP1	0	1.00	-	-	-	1.00	-	-	-
	1	1.84	[1.09;3.10]	0.022	0.534	0.72	[0.23;2.38]	0.590	1.000
Age		1.01	[0.99;1.03]	0.450	1.000	1.05	[1.02;1.09]	0.005	0.117
COL11A1	0	1.00	-	-	-	1.00	-	-	-
	1	2.29	[0.92;5.72]	0.076	1.000	1.90	[0.45;8.00]	0.381	1.000
Age		1.01	[0.98;1.03]	0.537	1.000	1.05	[1.02;1.09]	<0.005	0.112
DMD	0	1.00	-	-	-	1.00	-	-	-
	1	1.74	[0.70;4.31]	0.234	1.000	0.80	[0.11;5.94]	0.832	1.000
Age		1.01	[0.99;1.03]	0.444	1.000	1.05	[1.02;1.09]	<0.005	0.108
KDM5C	0	1.00	-	-	-	1.00	-	-	-
	1	1.23	[0.65;2.33]	0.533	1.000	1.25	[0.43;3.58]	0.682	1.000
Age		1.01	[0.98;1.03]	0.510	1.000	1.05	[1.02;1.09]	<0.005	0.118
PBRM1	0	1.00	-	-	-	1.00	-	-	-
	1	0.81	[0.51;1.29]	0.384	1.000	1.10	[0.55;2.18]	0.785	1.000
Age		1.01	[0.99;1.04]	0.404	1.000	1.05	[1.02;1.09]	0.005	0.126
PTK7	0	1.00	-	-	-	1.00	-	-	-
	1	4.07	[1.47;11.27]	0.007	0.168	(-)	(-)	(-)	1.000
Age		1.01	[0.98;1.03]	0.475	1.000	1.06	[1.02;1.09]	0.004	0.102
SETD2	0	1.00	-	-	-	1.00	-	-	-
	1	1.24	[0.72;2.13]	0.441	1.000	1.29	[0.57;2.92]	0.534	1.000
Age		1.01	[0.98;1.03]	0.582	1.000	1.05	[1.01;1.09]	0.006	0.154
TP53	0	1.00	-	-	-	1.00	-	-	-
	1	1.74	[0.83;3.64]	0.144	1.000	1.70	[0.40;7.24]	0.472	1.000
Age		1.01	[0.99;1.03]	0.451	1.000	1.06	[1.02;1.10]	0.004	0.103
TRRAP	0	1.00	-	-	-	1.00	-	-	-
	1	2.87	[1.24;6.66]	0.014	0.333	1.93	[0.46;8.10]	0.370	1.000
Age		1.01	[0.98;1.03]	0.555	1.000	1.05	[1.02;1.09]	0.005	0.120
VHL	0	1.00	-	-	-	1.00	-	-	-
	1	1.34	[0.75;2.39]	0.327	1.000	0.89	[0.40;1.97]	0.772	1.000
Age		1.01	[0.99;1.03]	0.462	1.000	1.06	[1.02;1.10]	0.004	0.104

* Cox PH model includes a stratification by tumor stage; Bonferroni adjustment was made to account for multiple testing

Supplementary Table 7. Multivariable cause-specific survival analysis (Validation)

		RCC-related death				Non-RCC-related death			
Predictor	Levels	HR	95% CI	p-value	q-value	HR	95% CI	p-value	q-value
ATM	0	1.00	-	-	-	1.00	-	-	-
	1	1.04	[0.25;4.24]	0.961	1.000	0.62	[0.09;4.51]	0.636	1.000
Age		1.02	[1.00;1.04]	0.114	1.000	1.06	[1.03;1.10]	0.00015	0.004
ATP9B	0	1.00	-	-	-	1.00	-	-	-
	1	2.42	[0.33;17.87]	0.387	1.000	(-)	(-)	(-)	(-)
Age		1.02	[1.00;1.04]	0.104	1.000	1.06	[1.03;1.10]	0.00017	0.004
BAP1	0	1.00	-	-	-	1.00	-	-	-
	1	2.22	[1.28;3.85]	0.005	0.112	0.86	[0.34;2.22]	0.762	1.000
Age		1.02	[1.00;1.05]	0.076	1.000	1.06	[1.03;1.10]	0.00015	0.004
COL11A1	0	1.00	-	-	-	1.00	-	-	-
	1	2.96	[0.91;9.63]	0.071	1.000	2.40	[0.57;10.03]	0.230	1.000
Age		1.02	[1.00;1.04]	0.092	1.000	1.06	[1.03;1.10]	0.00017	0.004
DMD	0	1.00	-	-	-	1.00	-	-	-
	1	1.55	[0.37;6.51]	0.551	1.000	2.27	[0.54;9.51]	0.263	1.000
Age		1.02	[1.00;1.04]	0.107	1.000	1.07	[1.03;1.10]	0.00013	0.003
KDM5C	0	1.00	-	-	-	1.00	-	-	-
	1	0.47	[0.17;1.30]	0.145	1.000	0.59	[0.14;2.50]	0.470	1.000
Age		1.02	[1.00;1.04]	0.091	1.000	1.06	[1.03;1.10]	0.00014	0.003
PBRM1	0	1.00	-	-	-	1.00	-	-	-
	1	0.93	[0.56;1.53]	0.766	1.000	1.01	[0.54;1.89]	0.981	1.000
Age		1.02	[1.00;1.04]	0.108	1.000	1.06	[1.03;1.10]	0.00016	0.004
PTK7	0	1.00	-	-	-	1.00	-	-	-
	1	0.83	[0.114;6.07]	0.856	1.000	5.13	[0.64;40.99]	0.123	1.000
Age		1.02	[1.00;1.04]	0.113	1.000	1.06	[1.03;1.10]	0.00020	0.005
SETD2	0	1.00	-	-	-	1.00	-	-	-
	1	1.91	[1.14;3.19]	0.014	0.327	0.56	[0.20;1.62]	0.286	1.000
Age		1.02	[1.00;1.04]	0.183	1.000	1.06	[1.03;1.10]	0.00018	0.004
TP53	0	1.00	-	-	-	1.00	-	-	-
	1	1.18	[0.37;3.82]	0.777	1.000	(-)	(-)	(-)	(-)
Age		1.02	[1.00;1.04]	0.123	1.000	1.07	[1.03;1.10]	0.00013	0.003
TRRAP	0	1.00	-	-	-	1.00	-	-	-
	1	3.77	[0.51;27.90]	0.194	1.000	(-)	(-)	(-)	1.000
Age		1.02	[1.00;1.04]	0.112	1.000	1.06	[1.03;1.10]	0.00015	0.004
VHL	0	1.00	-	-	-	1.00	-	-	-
	1	1.04	[0.60;1.81]	0.883	1.000	0.91	[0.44;1.87]	0.793	1.000
Age		1.02	[1.00;1.04]	0.114	1.000	1.06	[1.03;1.10]	0.00016	0.004

* Cox PH model includes a stratification by tumor stage; Bonferroni adjustment was made to account for multiple testing

Supplementary Table 8. Multivariable disease-free survival analysis investigating *BAP1* & *PBRM1* co-occurrence within the overall cohort and within *VHL*+2 tumors.

BAP1/PBRM1 co-occurrence					
		Disease-free survival			
Predictor	Levels	N	HR	95% CI	p-value
Mutation Status	None	289	1.00	-	-
	PBRM1 only	221	1.72	[1.09;2.72]	0.020
	BAP1 only	71	2.59	[1.51;4.44]	0.0006
	BAP1 & PBRM1	12	7.29	[3.03;17.54]	<0.0001
Age		593	1.00	[0.98;1.02]	0.874
BAP1/PBRM1 co-occurrence & other VHL+2 scenarios					
		Disease-free survival			
Predictor	Levels	N	HR	95% CI	p-value
Mutation Status	VHL+0	208	1.00	-	-
	VHL+2**	107	2.74	[1.49;5.01]	0.001
	BAP1 & PBRM1	8	12.08	[3.95;136.95]	<0.0001
Age		323	1.01	[0.98;1.037]	0.589

* Cox PH model includes a stratification by tumor stage
**VHL+2 group excludes co-occurring BAP1/PBRM1 mutations

Supplementary Table 9. Multivariable disease-free survival analysis investigating *PBRM1* & *SETD2* co-occurrence within the overall cohort and within *VHL*+2 tumors.

PBRM1/SETD2 co-occurrence					
		Disease-free survival			
Predictor	Levels	N	HR	95% CI	p-value
Mutation Status	None	311	1.00	-	-
	SETD2 only	49	1.80	[0.96;3.38]	0.067
	PBRM1 only	174	1.30	[0.81;2.08]	0.283
	SETD2 & PBRM1	59	2.68	[1.54;4.67]	0.0005
Age		593	1.00	[0.98;1.02]	0.937
PBRM1/SETD2 co-occurrence & other VHL+2 scenarios					
		Disease-free survival			
Predictor	Levels	N	HR	95% CI	p-value
Mutation Status	VHL+0	208	1.00	-	-
	VHL+2**	76	3.44	[1.84;6.44]	0.0001
	SETD2 & PBRM1	39	2.30	[1.02;5.21]	0.045
Age		323	1.01	[0.98;1.04]	0.599

* Cox PH model includes a stratification by tumor stage;
**VHL+2 group excludes co-occurring SETD2/ PBRM1 mutations

Supplementary Table 10. Multivariable disease-free survival analysis stratified by Leibovich risk group, and by genomic classifier within intermediate- and high-risk Leibovich groups

Eligible for adjuvant therapy stratified by Leibovich Risk Group					
		Disease-free survival			
Predictor	Levels	N	HR	95% CI	p-value
Leibovich	Intermediate	106	1.00	-	-
	High	75	3.19	[1.89;5.38]	<0.0001
Age		181	0.99	[0.97;1.02]	0.639
Genomic classifier within Leibovich Intermediate-Risk Group					
		Disease-free survival			
Predictor	Levels	N	HR	95% CI	p-value
Genomic Classifier	VHL+0	39	1.00	-	-
	VHL+1	44	2.64	[0.79; 8.81]	0.114
	VHL+2	19	4.72	[1.27;17.55]	0.021
	VHL+≥3	4	22.59	[4.46;114.49]	0.0002
Age		106	0.98	[0.95;1.01]	0.202
Genomic classifier within Leibovich High-Risk Group					
		Disease-free survival			
Predictor	Levels	N	HR	95% CI	p-value
Genomic Classifier	VHL+0	15	1.00	-	-
	VHL+1	12	1.14	[0.43;3.05]	0.792
	VHL+2	15	2.62	[1.00;6.83]	0.050
	VHL+≥3	7	2.80	[0.92;8.47]	0.0689
Age		49	0.99	[0.95;1.02]	0.431

Supplementary Table 11. Multivariable disease-free survival analysis stratified by genomic classifier within stage I-III tumors

Genomic classifier within Stage I tumors					
		Disease-free survival			
Predictor	Levels	N	HR	95% CI	p-value
Genomic Classifier	VHL+0	140	1.00	-	-
	VHL+1	146	2.22	[0.78;6.35]	0.137
	VHL+2	63	2.64	[0.80;8.73]	0.111
	VHL+≥3	16	6.45	[1.53;27.27]	0.011
Age		365	1.01	[0.97;1.05]	0.623
Genomic classifier within Stage II tumors					
		Disease-free survival			
Predictor	Levels	N	HR	95% CI	p-value
Genomic Classifier	VHL+0	26	1.00	-	-
	VHL+1	22	1.73	[0.56;5.34]	0.338
	VHL+2	11	1.31	[0.31;5.54]	0.710
	VHL+≥3	2	3.67	[0.42;31.78]	0.237
Age		61	0.98	[0.92;1.03]	0.390
Genomic classifier within Stage III tumors					
		Disease-free survival			
Predictor	Levels	N	HR	95% CI	p-value
Genomic Classifier	VHL+0	42	1.00	-	-
	VHL+1	68	2.07	[0.83;5.17]	0.117
	VHL+2	41	5.38	[2.19;13.25]	0.0003
	VHL+≥3	16	7.33	[2.65;20.27]	0.0001
Age		167	1.00	[0.97;1.02]	0.707

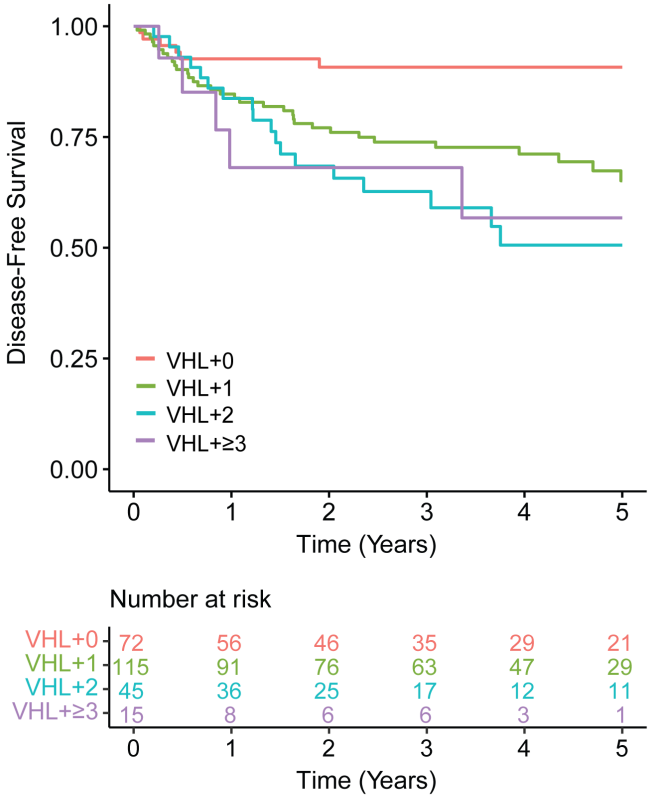
Supplementary Table 12. Wilcoxon Rank Sum Tests comparing Tumor Mutation Burden (TMB) within genomically-defined groups

Cagekid C1 Cohort				TCGA			
	VHL+0	VHL+1	VHL+2		VHL+0	VHL+1	VHL+2
VHL+1	0.0065	(-)	(-)	VHL+1	0.110	(-)	(-)
VHL+2	0.0020	0.6032	(-)	VHL+2	0.048	0.172	(-)
VHL+≥3	0.0179	0.6032	0.7451	VHL+≥3	0.110	0.290	0.798

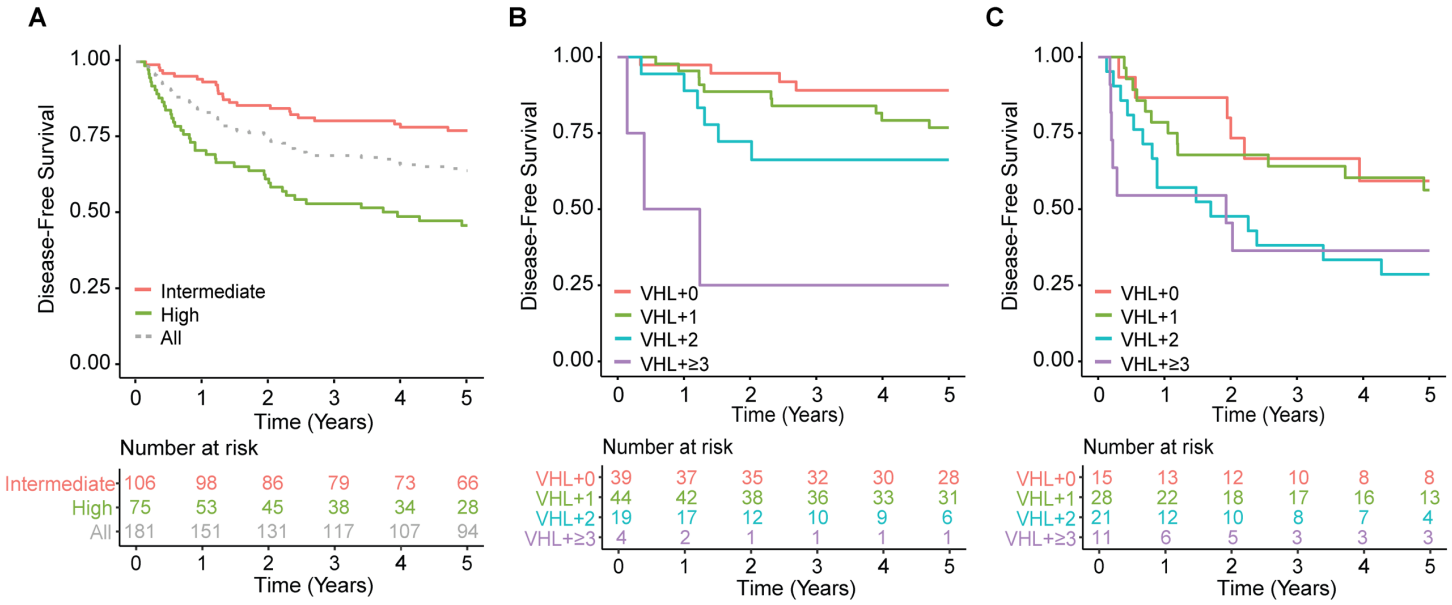
Supplementary Table 13. Multivariable disease-free survival analysis including Tumor Mutation Burden (TMB)

		Disease-free survival (C1)			Disease-free survival (TCGA)		
Predictor	Levels	HR	95% CI	p-value	HR	95% CI	p-value
VHL Category	VHL+0	1.00			1.00		
	VHL+1	1.66	[0.28;10.04]	0.579	1.70	[0.88;3.30]	0.117
	VHL+2	6.59	[1.26;34.30]	0.025	3.10	[1.50;6.40]	0.002
	VHL+≥3	8.83	[1.07;72.80]	0.043	3.00	[1.02;8.57]	0.046
TMB		0.92	[0.65;1.30]	0.643	1.05	[0.76;1.47]	0.754

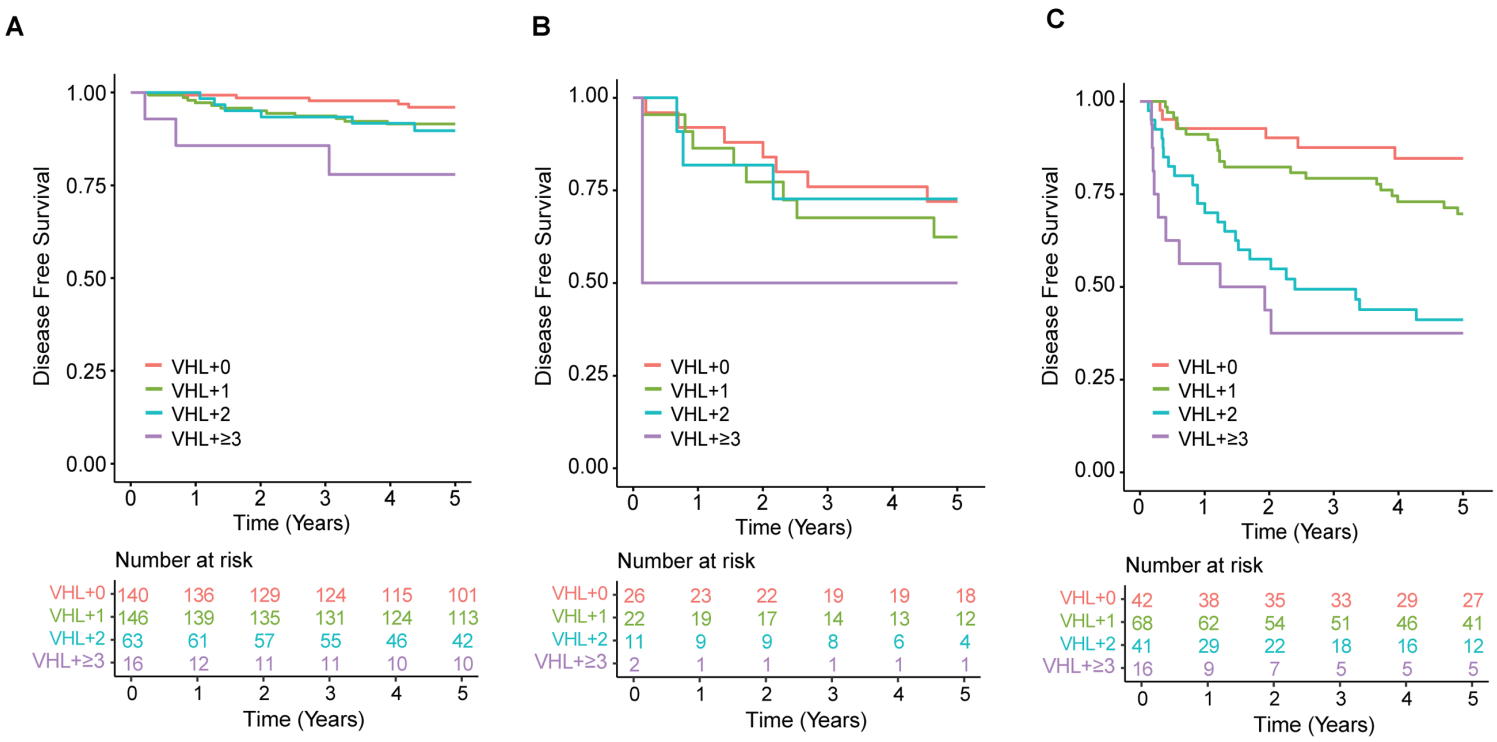
3.11.2 Supplementary Figures



Supplementary Figure S1. Disease-Free Survival outcomes amongst genomic groups when applied to the 247 *VHL* mutated ccRCCs from the TCGA dataset.



Supplementary Figure S2. Disease-Free Survival amongst Leibovich risk groups stratified by genomic classifier. Kaplan Meier survival curves amongst A) All eligible adjuvant patients by Leibovich intermediate- or high-risk group, B) Leibovich intermediate-risk patients stratified by genomic group, and C) Leibovich high-risk patients stratified by genomic group.



Supplementary Figure S3. Disease-Free Survival by tumor stage stratified by genomic classifier. Kaplan Meier survival curves for A) Stage I, B) Stage II, and C) Stage III tumors.

3.11.3 Supplementary Methods

Study Centres

Patients were recruited to the study from the following centres: St James's University Hospital, Leeds, UK; Newcastle Upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK; Stockport NHS Foundation Trust, Stockport, UK; Nottingham University Hospitals NHS Trust, Nottingham, UK; NHS Lothian, Edinburgh, UK; Charing Cross Hospital, London, UK; Oxford University Hospitals NHS Foundation Trust, Oxford UK; Northwick Park Hospital, Harrow, UK; University Hospital Motol, Prague, Czech Republic; General University Hospital, Prague, Czech Republic; Masaryk Memorial Cancer Institute, Brno, Czech Republic; Palacký University Hospital, Olomouc, Czech Republic; České Budějovice Regional Hospital, České Budějovice, Czech Republic; Th. Burghele Hospital, Bucharest, Romania; N. N. Blokhin Cancer Research Centre, Moscow, Russia; Clinical Center of Serbia (KCS), Belgrade, Serbia; Military Medical Academy, Belgrade, Serbia. All were recruited to the study after obtaining informed consent. Recruitment in Central and Eastern Europe was coordinated by the International Agency for Research on Cancer. Ethical approvals were obtained from the Leeds (East) Local Research Ethics Committee, the International Agency for Research on Cancer Ethics Committee, as well as from local ethics committee for recruiting centres in Czech Republic, Romania, Russia, Serbia and Bosnia & Herzegovina.

Study overview

Cohort 1 (C1) consisted of 93 patient samples subject to whole genome sequencing, previously reported in a descriptive study [**1**]. C2 consisted of an additional 376 patient samples, of which 24 were analysed by exome sequencing and 352 underwent targeted sequencing of 42 genes, identified as being most frequently mutated in C1 (**Supplementary Table 2**). C3 consisted

of an additional 474 patient samples, all of which were analysed by targeted sequencing of 12 genes (*ATM*, *ATP9B*, *BAP1*, *COL11A1*, *DMD*, *KDM5C*, *PBRM1*, *PTK7*, *SETD2*, *TP53*, *TRRAP*, and *VHL*), included in an RCC-focused gene panel [**2**]. These genes were selected on the basis of their known role in ccRCC biology, previously reported clinical associations, and/or our preliminary observed potential associations with outcome or other clinical parameters in C1 and C2.

References

1. Scelo G, Riazalhosseini Y, Greger L, et al. Variation in genomic landscape of clear cell renal cell carcinoma across Europe. *Nature Communications* **2014**;5: 5135.

2. Glennon KI, Maralani M, Abdian N, et al. Rational Development of Liquid Biopsy Analysis in Renal Cell Carcinoma. *Cancers (Basel)* **2021**;13:5825.

3.12 REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–49.
2. Znaor A, Lortet-Tieulent J, Laversanne M, Jemal A, Bray F . International variations and trends in renal cell carcinoma incidence and mortality. Eur Urol 2015;67:519–30.
3. Dabestani S, Beisland C, Stewart GD, Bensalah K, Gudmundsson E, Lam TB, et al. Long-term outcomes of follow-up for initially localized clear- cell renal cell carcinoma: RECUR database analysis. Eur Urol Focus 2019;5:857–66.
4. Larroquette M, Peyraud F, Domblides C, Lefort F, Bernhard JC, Ravaud A, et al. Adjuvant therapy in renal cell carcinoma: current knowledge and future perspectives. Cancer Treat Rev 2021;97:102207.
5. Choueiri TK, Tomczak P, Park SH, Venugopal B, Ferguson T, Chang YH, et al. Adjuvant pembrolizumab after nephrectomy in renal cell carcinoma. N Engl J Med 2021;385:683–94.
6. Powles T, Tomczak P, Park SH, Venugopal B, Ferguson T, Symeonides SN, et al. Pembrolizumab versus placebo as post-nephrectomy adjuvant therapy for clear-cell renal cell carcinoma (KEYNOTE-564): 30-month follow-up analysis of a multicenter, randomized, double-blind, placebo-controlled, phase III trial. Lancet Oncol 2022;23:1133–44.
7. Pal SK, Uzzo R, Karam JA, Master VA, Donskov F, Suarez C, et al. Adjuvant atezolizumab versus placebo for patients with renal cell carcinoma at increased risk of recurrence

following resection (IMmotion010): a multicenter, randomized, double-blind, phase III trial. Lancet 2022;400:1103–16.

8. Motzer RJ, Russo P, Gruenwald V, Tomita Y, Zurawski B, Parikh OA, et al. Adjuvant nivolumab plus ipilimumab (NIVO+IPI) vs placebo (PBO) for localized renal cell carcinoma (RCC) at high risk of relapse after nephrectomy: results from the randomized, Phase III CheckMate 914 trial. Ann Oncol 2022;33:S808–S69.

9. Allaf M, Kim SE, Harshman LC, McDermott DF, Master VA, Signoretti S, et al. Phase III randomized study comparing perioperative nivolumb (nivo) versus observation in patients (Pts) with renal cell carcinoma (RCC) undergoing nephrectomy (PROSPER, ECOG-ACRIN EA8143), a national clinical trials network trial. Ann Oncol 2022;33:S808–S69.

10. Figlin RA, Leibovich BC, Stewart GD, Negrier S . Adjuvant therapy in renal cell carcinoma: does higher risk for recurrence improve the chance for success? Ann Oncol 2018;29:324–31.

11. Rossi SH, Blick C, Handforth C, Brown JE, Stewart GD . Renal Cancer Gap Analysis Collaborative. Essential research priorities in renal cancer: a modified Delphi consensus statement. Eur Urol Focus 2020;6:991–98.

12. Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, Reznik E, et al. The Cancer Genome Atlas comprehensive molecular characterization of renal cell carcinoma. Cell Rep 2018;23:313–26.

13. Scelo G, Riazalhosseini Y, Greger L, Letourneau L, Gonzalez-Porta M, Wozniak MB, et al. Variation in genomic landscape of clear- cell renal cell carcinoma across Europe. Nat Commun 2014;5:5135.

14. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear- cell renal cell carcinoma. Nature 2013;499:43–9.

15. Turajlic S, Xu H, Litchfield K, Rowan A, Horswell S, Chambers T, et al. Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. Cell 2018;173:595–610.

16. Bourgey M, Dali R, Eveleigh R, Chen KC, Letourneau L, Fillon J, et al. GenPipes: an open-source framework for distributed and scalable genomic analyses. Gigascience 2019;8:giz037.

17. FulcrumGenomics. 1/09/2020. fgbio. Available from: https://github.com/fulcrumgenomics/fgbio.

18. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy-number alteration discovery in cancer by exome sequencing. Genome Res 2012;22:568–76.

19. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res 2016;44:e108.

20. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single-nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. Fly 2012;6:80–92.

21. Paila U, Chapman BA, Kirchner R, Quinlan AR . GEMINI: integrative exploration of genetic variation and genome annotations. PLoS Comput Biol 2013;9:e1003153.

22. Gray RJ . cmprsk: subdistribution analysis of competing risks 2010.

23. Riazalhosseini Y, Lathrop M . Precision medicine from the renal cancer genome. Nat Rev Nephrol 2016;12:655–66.

24. Glennon KI, Maralani M, Abdian N, Paccard A, Montermini L, Nam AJ, et al. Rational development of liquid biopsy analysis in renal cell carcinoma. Cancers 2021;13:5825.

25. Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature 2010;463:360–3.

26. Guo G, Gui Y, Gao S, Tang A, Hu X, Huang Y, et al. Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear- cell renal cell carcinoma. Nat Genet 2012;44:17–9.

27. Sato Y, Yoshizato T, Shiraishi Y, Maekawa S, Okuno Y, Kamura T, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. Nat Genet 2013;45:860–7.

28. Kapur P, Pena-Llopis S, Christie A, Zhrebker L, Pavia-Jimenez A, Rathmell WK, et al. Effects on survival of BAP1 and PBRM1 mutations in sporadic clear-cell renal cell carcinoma: a retrospective analysis with independent validation. Lancet Oncol 2013;14:159–67.

29. Joseph RW, Kapur P, Serie DJ, Parasramka M, Ho TH, Cheville JC, et al. Clear- cell renal cell carcinoma subtypes identified by BAP1 and PBRM1 expression. J Urol 2016;195:180–7.

30. Ho TH, Kapur P, Joseph RW, Serie DJ, Eckel-Passow JE, Tong P, et al. Loss of histone H3 lysine 36 trimethylation is associated with an increased risk of renal cell carcinoma-specific death. Mod Pathol 2016;29:34–42.

31. Leibovich BC, Blute ML, Cheville JC, Lohse CM, Frank I, Kwon ED, et al. Prediction of progression after radical nephrectomy for patients with clear- cell renal cell carcinoma: a stratification tool for prospective clinical trials. Cancer 2003;97:1663–71.

32. Hakimi AA, Ostrovnaya I, Reva B, Schultz N, Chen YB, Gonen M, et al. Adverse outcomes in clear- cell renal cell carcinoma with mutations of 3p21 epigenetic regulators BAP1 and SETD2: a report by MSKCC and the KIRC TCGA research network. Clin Cancer Res 2013;19:3259–67.

33. Manley BJ, Zabor EC, Casuscelli J, Tennenbaum DM, Redzematovic A, Becerra MF, et al. Integration of recurrent somatic mutations with clinical outcomes: a pooled analysis of 1,049 patients with clear- cell renal cell carcinoma. Eur Urol Focus 2017;3:421–27.

34. Frank I, Blute ML, Cheville JC, Lohse CM, Weaver AL, Zincke H . 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 2002;168:2395–400.

35. Zisman A, Pantuck AJ, Wieder J, Chao DH, Dorey F, Said JW, et al. Risk group assessment and clinical outcome algorithm to predict the natural history of patients with surgically resected renal cell carcinoma. J Clin Oncol 2002;20:4559–66.

36. Correa AF, Jegede O, Haas NB, Flaherty KT, Pins MR, Messing EM, et al. Predicting renal cancer recurrence: defining limitations of existing prognostic models with prospective trial-based validation. J Clin Oncol 2019;37:2062–71.

37. Oza B, Eisen T, Frangou E, Stewart GD, Bex A, Ritchie AWS, et al. External validation of the 2003 leibovich prognostic score in patients randomly assigned to SORCE, an

international phase III trial of adjuvant sorafenib in renal cell cancer. J Clin Oncol 2022;40:1772–82.

38. Vasudev NS, Hutchinson M, Trainor S, Ferguson R, Bhattarai S, Adeyoju A, et al. UK multicenter prospective evaluation of the Leibovich score in localized renal cell carcinoma: performance has altered over time. Urology 2020;136:162–8.

39. Rini B, Goddard A, Knezevic D, Maddala T, Zhou M, Aydin H, et al. A 16-gene assay to predict recurrence after surgery in localized renal cell carcinoma: development and validation studies. Lancet Oncol 2015;16:676–85.

40. Rini BI, Escudier B, Martini JF, Magheli A, Svedman C, Lopatin M, et al. Validation of the 16-gene recurrence score in patients with locoregional, high-risk renal cell carcinoma from a phase III trial of adjuvant sunitinib. Clin Cancer Res 2018;24:4407–15.

41. Tosoian JJ, Feldman AS, Abbott MR, Mehra R, Tiemeny P, Wolf JS Jr, et al . Biopsy cell-cycle proliferation score predicts adverse surgical pathology in localized renal cell carcinoma. Eur Urol 2020;78:657–60.

42. Motzer RJ, Banchereau R, Hamidi H, Powles T, McDermott D, Atkins MB, et al. Molecular subsets in renal cancer determine outcome to checkpoint and angiogenesis blockade. Cancer Cell 2020;38:803–17.

43. Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, et al. Genomic architecture and evolution of clear- cell renal cell carcinomas defined by multi-region sequencing. Nat Genet 2014;46:225–33.

44. Gulati S, Martinez P, Joshi T, Birkbak NJ, Santos CR, Rowan AJ, et al. Systematic evaluation of the prognostic impact and intra-tumor heterogeneity of clear- cell renal cell carcinoma biomarkers. Eur Urol 2014;66:936–48.

BRIDGING STATEMENT TO CHAPTER 4.

In the previous chapter, we focused on characterizing the somatic mutation profiles of ccRCC tumors to identify biomarkers of disease prognosis that demonstrate clinical utility. Germline mutations in cancer predisposition genes are also informative for the clinical management of RCC. Screening schedules for individuals with germline pathogenic variants (PVs) that confer a higher risk of developing renal cancer enables early diagnosis. PVs in certain genes may also be associated to specific phenotypes and clinical outcomes that can be informative for risk-stratification, with recent evidence also suggesting that PVs may be more frequent in metastatic cancers[134]. For patients with advanced RCC or at high-risk for disease recurrence, the presence of PVs in certain genes can guide systemic therapy decisions. Tumors with germline mutations in *MET* have a more active tyrosine kinase inhibitors targeting MET[135], while those with mutations in *BRCA1/2* are more vulnerable to PARP inhibitors[136]; knowledge of such mutations allows clinicians to prioritize certain systemic therapies. Identifying additional genes with PVs associated to RCC may also reveal additional therapeutic targets, and increase our understanding of RCC pathogenesis.

In Chapter 4, we investigate germline susceptibility to sporadic RCC within a large cohort of RCC patients from Canada. We identified risk-genes for clear cell and non-clear cell RCCs that are prevalent in Canada, and investigated their associations to clinical phenotypes and outcome. Additionally, to shed light on the visible global differences in RCC incidence, we compared gene burden for risk genes among large available cohorts from various countries. Lastly, we evaluated clinical guidelines for genetic screening to assess their capability to identify individuals with heritable predisposition to renal cancer.

CHAPTER 4: GERMLINE SUSCEPTIBILITY TO RENAL CELL CARCINOMA AND IMPLICATIONS FOR GENETIC SCREENING

Kate I. Glennon[1-3], Mikiko Endo[3], Yoshiaki Usui[3], Yusuke Iwasaki[3], Rodney H. Breau[4], Anil Kapoor[5], Mark Lathrop[1,2], Simon Tanguay[6], Yukihide Momozawa[3], Yasser Riazalhosseini[1,2]

[1]Department of Human Genetics, McGill University, Montreal, Canada

[2]Victor Philip Dahdaleh Institute of Genomic Medicine at McGill University, Montreal, Canada,

[3]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

[4]University of Ottawa, Ottawa, Canada

[5]Juravinski Cancer Centre, McMaster University, Hamilton, ON, Canada

[6]Department of Surgery, Division of Urology, McGill University, Montreal, Canada

Correspondence:

Dr. Yasser Riazalhosseini

Victor Philip Dahdaleh Institute of Genomic Medicine at McGill University

740 Dr. Penfield Ave,

Montreal, Quebec H3A 0G1, Canada

Email: yasser.riazalhosseini@mcgill.ca

4.1 ABSTRACT

Background: Genetic susceptibility to non-syndromic renal cell carcinoma (RCC) remains poorly understood, especially for different histological subtypes. Furthermore, no comparative analysis between genetic data from different cohorts has been reported to date.

Methods: We conducted targeted sequencing of 19 RCC-related and 27 cancer predisposition genes for 960 RCC patients from Canada, and identified genes enriched in rare germline PVs in RCC compared to cancer-free controls. We combined our results with those reported in patients from Japan, the UK, and the USA to investigate PVs variations in different patient populations. Furthermore, we evaluated the performance of referral criteria for genetic screening for including patients with rare PVs in Canada, the USA, and the UK.

Results: We identified 39 germline PVs in 56 patients (5.8%) from the Canadian cohort. Compared to cancer-free controls, PVs in *CHEK2* genes were significantly enriched in patients with clear cell RCC, whereas PVs in *FH* were enriched in patients with non-clear cell RCC. PVs in *BRCA1*, *BRCA2* and *ATM* were associated with metastasis. Comparisons of gene-burden to other large cohorts revealed an enrichment for PVs in *TP53* in patients from Japan, in *CHEK2* and *ATM* in patients from Canada, USA and the UK, and in *FH* and *BAP1* in USA.

Conclusions: *CHEK2*, *ATM*, and *FH* were identified as risk-genes for RCC within the Canadian population, while germline PVs in *BRCA1/2* and *ATM* are associated with metastasis. Globally, clinical guidelines for genetic screening in RCC fail to include over 70% of patients with rare germline PVs.

4.2 INTRODUCTION

Renal cell carcinoma (RCC) accounts for over 400,000 new cancer diagnoses, and almost 200,000 deaths each year [**1**]. RCC is characterized by heterogeneous tumor biology and clinical outcomes, with clinical characteristics varying by RCC histological subtype [**1**-**3**]. The majority of renal cortical tumors are clear cell RCC (ccRCC; ~75%), whereas non-clear cell subtypes (nccRCC) include papillary (~10%), chromophobe (~5%), and other histology. Genetic susceptibility to RCC remains poorly understood, despite evidence that having a family history of renal cancer doubles risk [**4**].

Germline genetic variation contributes strongly to individual differences in susceptibility to cancer. Approximately 5-8% of RCCs are associated to known hereditary kidney cancer syndromes, including von Hippel Lindau disease, Birt-Hogg-Dubé syndrome, and hereditary leiomyomatosis; each of which are characterized by germline mutations in known renal cancer genes [**5**,**6**]. While genetic drivers of these syndromic forms of the disease are well-established, genetic factors predisposing to sporadic RCC are less known and have only recently been investigated at large scales. Additionally, non-syndromic germline predisposition to RCC subtypes is not fully understood.

There are unexplained variations in RCC incidence across the globe, which highlight the need for RCC studies that have diverse geographic representation. Recent RCC studies limited to specific populations report contrasting results about the prevalence of germline pathogenic variants (PVs) [**7**-**9**]. Additionally, previous studies lack statistical power or are missing pathological information required to explain variations in PV incidence among patients with RCC subtypes.

Here, we investigated the prevalence of germline PVs in RCC-related and cancer-predisposing genes within a Canadian cohort and examined differences in risk-genes between RCC patients from Canada and other large population studies.

4.3 MATERIALS AND METHODS

4.3.1 Patient Cohort

Clinical information and buffy-coat driven DNA samples were obtained for 997 unrelated RCC patients (790 with ccRCC and 207 with non-ccRCC), collected through the Ontario Tumour Bank between 2005 and 2019, following informed written consent. Pathology review to confirm subtypes was performed by accredited clinical pathologists at each of the hospital sites where samples were collected. Patients with known hereditary syndromes at the time of sample collection were excluded from the study. Ethical approval was obtained from the Faculty of Medicine and Health Sciences Institutional Review Board at McGill University, the Hamilton Integrated Research Ethics Board, and the ethics committee of the RIKEN Center for Integrative Medical Sciences.

4.3.2 Targeted Sequencing and Bioinformatic Analysis

We conducted targeted sequencing of 19 RCC-related and 27 cancer predisposition genes using multiplex PCR-based amplification followed by next-generation sequencing as described previously [**8**] (**Supplementary Table 1**). Genes with established associations to cancer predisposition or to RCC were included in this study [**8**]. We sequenced all coding regions, with 2-bp flanking sequences, with a total target region of 145,459bp. Sequencing was conducted on an Illumina NovaSeq to an average sequencing depth of 902.81X. Raw sequencing data was processed using a tailored bioinformatic pipeline as previously described [**10**], to identify germline genetic variants. Variants listed as ‘Pathogenic’ or ‘Likely Pathogenic’ in ClinVar, and those with

a predicted 'HIGH' impact from SnpEff [**11**] were included as Pathogenic Variants. For the *MITF* and *HOXB13* genes, HIGH impact variants with predicted 'loss-of-function' effects were not included, as there is evidence that they do not cause cancer [**12**]. After quality control measures to ensure adequate sequencing depth and quality, data from 960 patients was included in the analysis of pathogenic germline variants. Mutations in the *MUTYH* gene were not considered in the analyses, as the gene follows an autosomal recessive inheritance pattern, and all identified germline variants in *MUTYH* were heterozygous.

4.3.3 Statistical Analyses

Patient samples were divided into ccRCC and nccRCC subtypes for the majority of statistical analyses. All histological subtypes other than traditional clear cell RCC were categorized as nccRCC, including those with notably mixed histology, such as clear cell papillary features. Gene-based association tests were conducted between patients with RCC and non-cancer control data from the gnomAD public database (gnomAD_v2.1.1 non-cancer dataset, European (Non-Finnish) population) [**13**]. Gene burden and PV burden to RCC was evaluated using Firth's logistic regression, with a Bonferroni correction applied to correct for multiple testing. Associations to clinical characteristics were assessed using Fisher's Exact tests. Adjusted p-values <0.05 were considered significant for all analyses.

4.4 RESULTS

4.4.1 Cohort Demographics and Clinical characteristics

We analyzed germline mutation status for 960 Canadian patients with renal cell carcinoma. The cohort consisted of 759 patients with ccRCC, and 201 patients with nccRCC. Patients with nccRCC more frequently had a personal history of other cancers (29.1%) compared to patients with ccRCC (21.1%) ($p=0.028$). Alternatively, ccRCC patients had a greater incidence of family

history of RCC compared to patients with nccRCC (2.3% and 1.6% respectively) (p=0.040) (**Supplementary Table 2**).

4.4.2 Identification of Germline Pathogenic Variants

We identified 39 different germline pathogenic variants (PVs) in 5.8% of studied patients (56/960), with no patients harbouring multiple germline PVs in the examined genes (**Figure 1**) (**Supplementary Table 3**). We observed no significant differences in the overall number of germline PVs between ccRCC (5.8%, 44/759) and nccRCC (6.0%, 12/201) patients (**Figure 2A**). The most frequently mutated gene was *CHEK2*, with 69% of *CHEK2* mutations being the known c.1100delC variant (11/16 patients), an established breast cancer susceptibility allele, which has also been suggested to increase the risk of developing other cancers [**14**-**18**]. Within ccRCC cases, germline PVs were most commonly found in *CHEK2* (14), *ATM* (7), *BRCA2* (6), and *MITF* (5). In patients with nccRCC, the most commonly mutated genes were *FH* (4) and *CHEK2* (2). For variants identified in multiple patients and reported in gnomAD, we compared the frequency of PVs in our cohort with that of 118,148 cancer-free controls to identify variants with association to RCC (**Supplementary Table 4**). Notably, the *CHEK2* c.1100delC variant was found in 11 RCC patients and showed to be significantly enriched in RCC compared to cancer-free controls (OR: 4.7, 95% CI: 2.4-8.1, p=0.0009). The *MITF* p.E318K variant did not show significant enrichment in RCC compared to cancer-free controls (OR: 2.9, 95% CI: 0.2-1.8, p=0.586).

Additionally, 8 RCC patients harbored low-penetrance *CHEK2* variants p.Ser428Phe (1) and p.Ile157Thr (7) (**Figure 2B**). These variants have traditionally been excluded from association analyses due to high population frequencies and conflicting interpretations regarding their pathogenicity [**19**]. However, recent studies in RCC continue to suggest an enrichment for these variants within ccRCC patients [**20**].

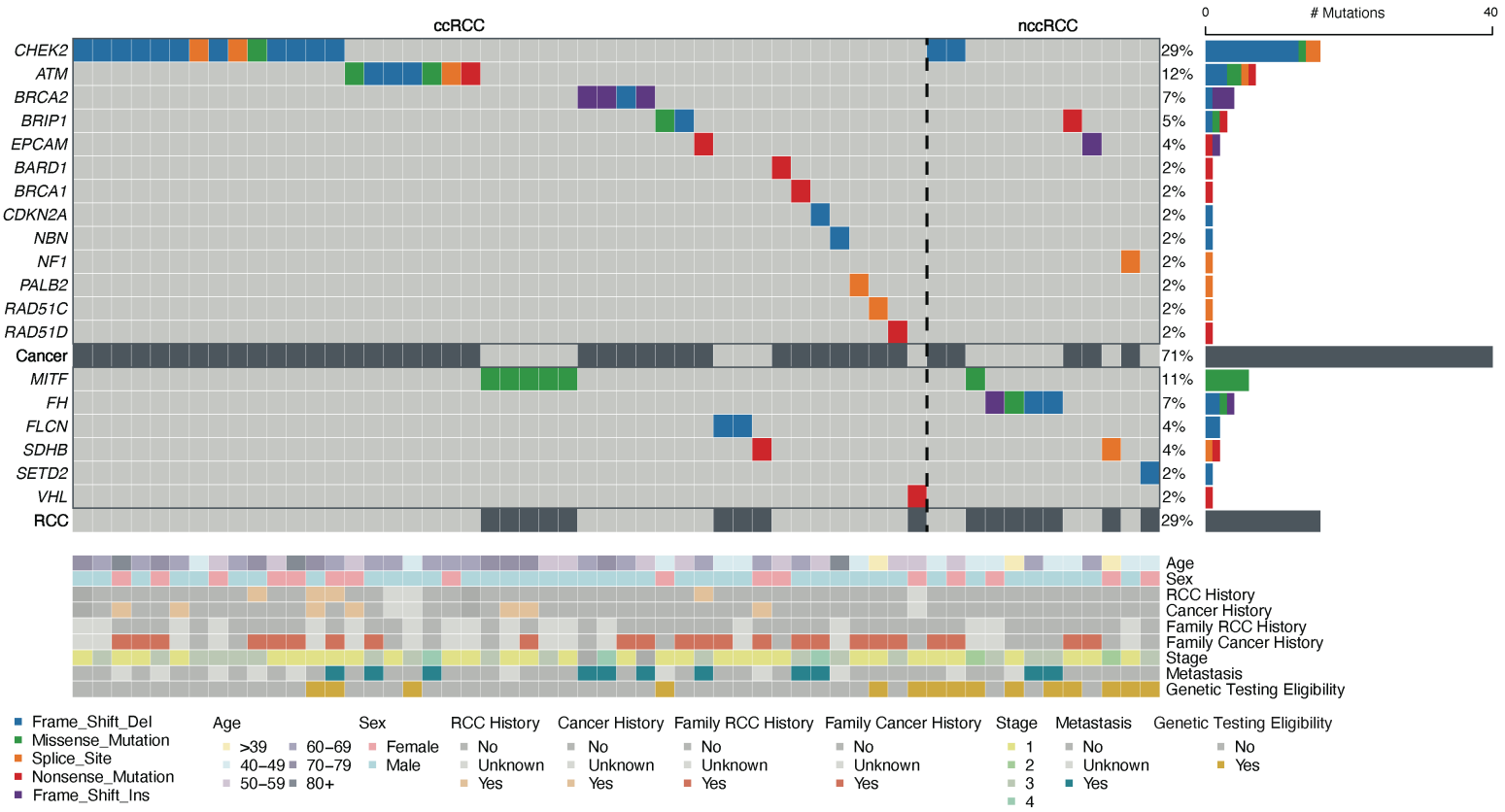


Figure 1. Oncoplot for RCC patients with germline pathogenic variants. Mutation rates are shown for general cancer susceptibility genes (Cancer) and RCC-related genes (RCC).

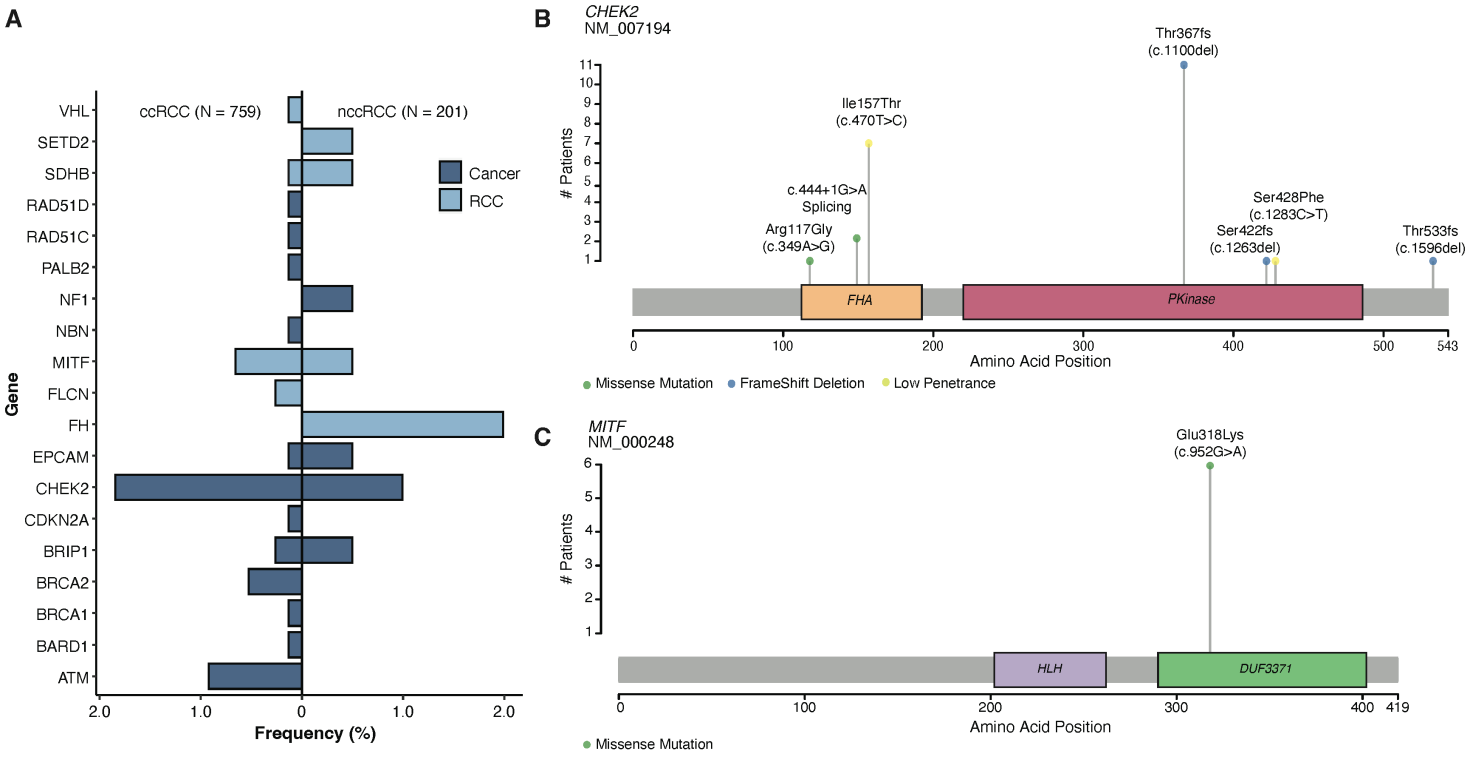


Figure 2. (A) Mutation frequency of pathogenic variants in ccRCC and nccRCC for general cancer genes (Cancer) and RCC-related genes (RCC). Germline variants observed in (B) *CHEK2* and (C) *MITF* genes.

4.4.3 Gene Burden within the Canadian Cohort

We conducted gene burden analyses to identify genes significantly mutated within ccRCC and nccRCC subtypes in our cohort. *CHEK2*, and *ATM* genes were significantly enriched for PVs in ccRCC (**Table 1**). Furthermore, *FH* gene was significantly enriched for PVs in the nccRCC subtypes, with 2% of nccRCC patients carrying germline PVs in *FH*, while the similar figure in cancer-free controls was only 0.01% (**Table 1**). Interestingly, germline *MITF* mutations were identified in 0.66% of ccRCC patients (6/759) and 0.5% of nccRCC patients (1/201), all of which had the known p.E318K variant, which has previously been linked to melanoma and RCC [**21**,**22**], and is present in 0.24% of cancer-free controls (**Figure 2C**). Although the gene-burden analyses did not show significant enrichment for *MITF* PVs in ccRCC or nccRCC, the frequency of the p.E318K variant is more than 2-fold higher among RCC patients than in the cancer-free control cohort (**Supplementary Table 5**).

Table 1. Gene-burden association tests between patients with RCC and non-cancer controls. Number of subjects with pathogenic variants are presented among all ccRCC (n=759), nccRCC (n=201) and Control cases (n=118,148).

ccRCC				
Gene	**No. of Patients (%)**	**No. of Controls (%)**	**P-value**	**OR (95% CI)**
CHEK2	14 (1.84)	474 (0.40)	3.94 × 10–5	4.8 (2.7 – 7.9)
ATM	7 (0.92)	259 (0.22)	0.016	4.5 (2.0 – 8.7)
nccRCC				
Gene	**No. of Patients (%)**	**No. of Controls (%)**	**P-value**	**OR (95% CI)**
FH	4 (2.0)	12 (0.01)	6.14 × 10–9	215.1 (64.4 – 597.8)
SETD2	1 (0.50)	3 (0.00)	0.003	219.6 (21.2 – 1341.8)
SDHB	1 (0.50)	20 (0.02)	0.033	43.1 (4.8 – 169.1)

4.4.4 Clinical characteristics of patients with germline PVs

We investigated the association between germline PVs and clinical characteristics by comparing germline PV carriers to RCC patients without any PVs (**Supplementary Table 6**). We did not observe significant differences between germline carriers and non-carriers for personal and familial cancer history, cancer stage, patient sex, or smoking status. Next, we examined PVs in any genes associated with the presence of metastasis. In patients with ccRCC, we observed a significant association between metastasis and the presence of PVs in *BRCA1/2* and *ATM* ($p=0.004$). This association was even stronger when we considered all patients, including those with non-ccRCCs ($p=0.003$).

4.4.5 Global Differences in RCC Susceptibility Genes

To develop an understanding of global differences in germline susceptibility to RCC, we compared gene burden for significantly mutated genes in patient cohorts from Canada (our study), the UK [**7**], Japan [**8**], and the USA [**9**] individually to all others combined. RCC patients from Japan were enriched for PVs in *TP53* compared to other cohorts (OR: 7.7, 95% CI: 2.3-39.7, $p=0.004$), while they were depleted for PVs in *CHEK2* (OR: 0.2, 95% CI: 0.1-0.5, $p<0.0001$) compared to the combination of other cohorts (Canada, USA, and UK) (**Table 2**, **Supplementary Table 7**). RCC patients from USA were enriched for PVs in *BAP1* (OR: 8.2, 95% CI: 1.9-29.0, $p=0.04$) and *FH* (OR: 11.0, 95% CI: 4.1-28.2, $p<0.0001$, **Table 2**, **Supplementary Table 7**) compared to the combination of other cohorts. RCC patients from Canada and UK were not enriched for any specific genes when compared to other cohorts (**Figure 3A,3B**).

Table 2. Comparing gene burdens in RCC cohorts from Japan and USA to the pool of other cohorts. Numbers of PV carriers for each gene are shown for the comparison groups. Fractions of carriers are shown in parentheses.

Japan

Gene	Japan Patients *n=1532*	Others *n=2550*	P-value	OR (95% CI)
TP53	11 (0.72)	2 (0.08)	0.004	7.7 (2.3 – 39.7)
MITF	0 (0)	16 (0.63)	0.003	0.1 (0.0 – 0.4)
CHEK2	7 (0.46)	52 (2.04)	7.3×10^{-5}	0.2 (0.1 – 0.5)

USA

Gene	USA Patients *n=254*	Others *n=3828*	P-value	OR (95% CI)
FH	7 (2.76)	10 (0.26)	9.3×10^{-5}	11.0 (4.1 – 28.2)
BAP1	3 (1.18)	6 (0.16)	0.040	8.2 (1.9 – 29.0)

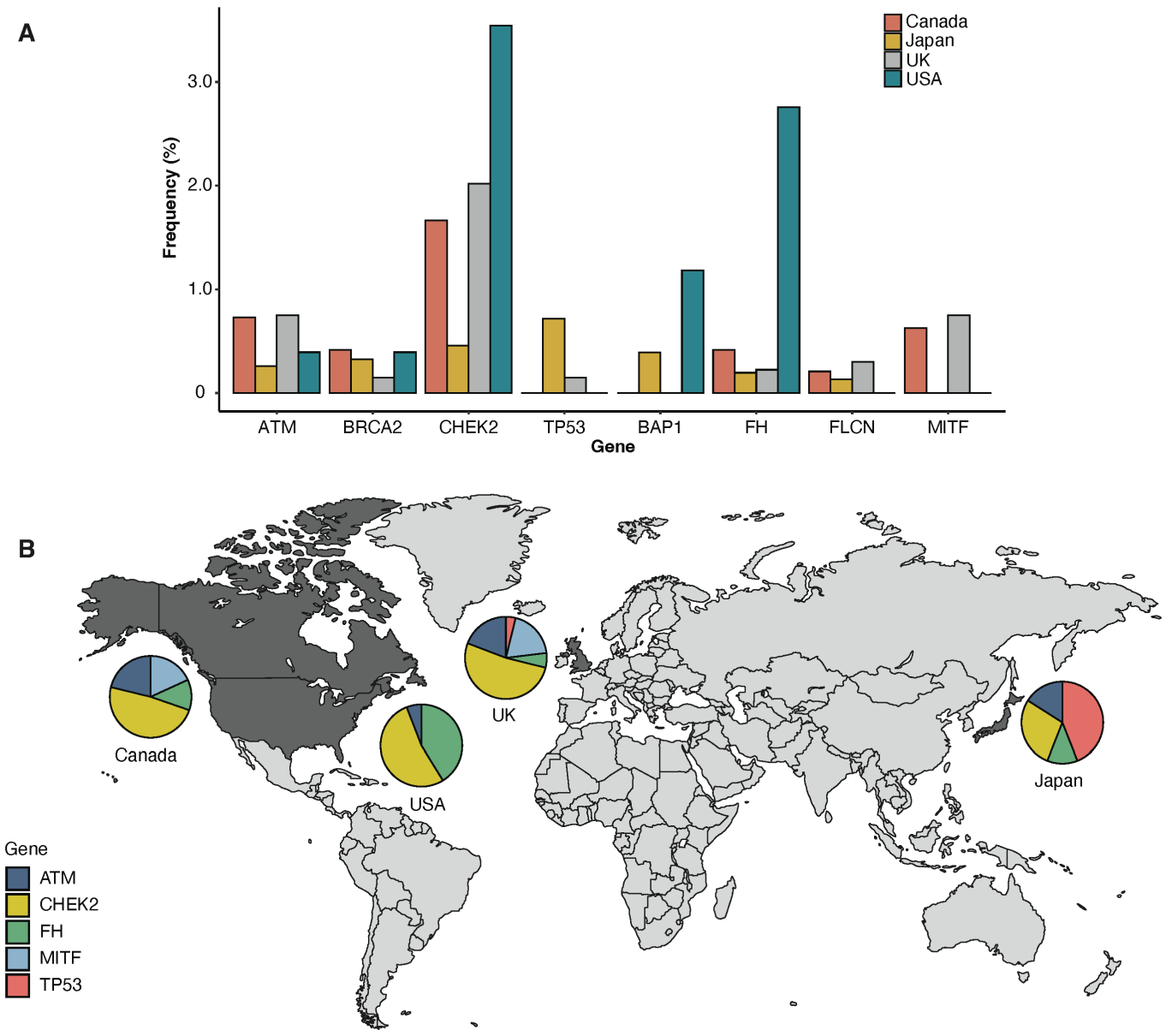


Figure 3. Differences in pathogenic variant frequency in RCC patients from Canada, Japan, USA, and the UK. (A) Mutation rate by study cohort, and (B) Most frequently mutated genes by geographical region.

4.5 DISCUSSION

We performed targeted sequencing of 19 RCC-associated genes and 27 cancer-predisposition genes in a cohort of 960 unrelated Canadian patients with RCC, including 759 patients with ccRCC, and 201 patients with nccRCC. Patients with ccRCC showed enrichment for pathogenic variants in *CHEK2*, and *ATM*, while patients with nccRCC were enriched for *FH* mutations, when compared to cancer-free controls. These results support observations reported in a recent study that different genes predispose to clear cell and non-clear cell RCC subtypes [**8**]. Furthermore, we observed a novel association between metastasis and germline mutations in *BRCA1/2*, and *ATM* genes in our cohort. Importantly, tumors with germline mutations in *BRCA1* and *BRCA2* have been shown vulnerable to PARP inhibitors, leading to the prioritized use of PARP inhibitors for BRCA-positive patients in multiple cancer types [**23**]. Recent clinical trials have already begun to investigate the utility of PARP inhibitors for the treatment of metastatic RCC for patients harboring mutations in DNA repair genes (including BRCA genes) (Trial ID: NCT03786796). In this context, our results may provide utility for identifying patients at an increased risk of developing metastases and may open new avenues for personalized prevention of metastasis in carriers of *BRCA1/2* mutations through treatment with PARP inhibitors.

Our study reveals similar patterns for rare germline variants that predispose to RCC between Canadian and the European patients. We detected three common *CHEK2* European founder variants at high frequency within Canadian RCC patients. The most well-known *CHEK2* variant, c.1100del, has frequently been studied within the European population, showing association to breast cancer risk and enrichment within other cancer types, including kidney cancer [**15**,**17**,**18**]. We observed a significant association between *CHEK2* and ccRCC, with the c.1100del variant being the most common PV within the cohort. Other *CHEK2* founder variants, p.S428F

and p.I157T, have conflicting interpretations regarding their pathogenicity, partly due to their higher frequencies among the cancer-free individuals. Recent studies indicate that these variants are enriched in RCC [**20**]; however, functional studies would be required to determine their pathogenicity potentials.

Germline translocations in *MITF* have been linked to a rare and aggressive subtype of RCC; MiTF/TFE translocation renal cell carcinoma (tRCC) [**24**]. Although MITF variants did not show significant enrichment in our patient cohort as compared to cancer-free individuals, all *MITF*-mutated patients within the Canadian cohort harbored the p.E318K variant, which is known as a predisposition locus for familial melanoma, but has also been associated with RCC [**21**]. This variant was also captured at a relatively high frequency (0.7%) among European RCC patients [**7**], and functional studies have demonstrated that the variant results in *MITF* upregulation, which in turn can induce overexpression of hypoxia-inducible factor 1α (HIF1α) [**22**]. Although it has been proposed that the p.E318K may predispose for co-occurring melanoma and RCC [**22**], only 1 of the 6 patients with the p.E318K variant in our study had a personal history of melanoma, while none had a family history of melanoma. This suggests that germline *MITF* p.E318K variant may predispose to RCC without the co-occurrence of melanoma; however, the relative high frequency of this variant among cancer-free individuals may indicate that it is a low-penetrance mutation. In line with the aforementioned similarities that we observed between Canadian and European patients, our gene-burden analyses did not show any significant differences between these cohorts. This is likely due to the largely European ancestry of Canadians. However, recent studies highlight the necessity for finer-scale population stratification [**20**], as variants are likely to have varying population frequencies even throughout Europe. On the other hand, our global gene-burden analysis highlighted key differences in genes which are enriched in patients from Japan and USA

as compared to other cohorts; whereas *TP53* germline PVs are enriched in RCC patients from Japan, *FH* and *BAP1* variants have significantly higher frequencies in the USA cohort. Although the lack of information regarding ethnicity of patients did not allow us to report on population-based genetic differences, our findings bring new insights into global differences in germline susceptibility to RCC, which may be used for refining gene lists for clinical genetic screening for RCC based on country and population-specific data.

The identification of at-risk patients is important for surgical decisions, as well as decisions regarding follow-up for both the patient and at-risk family members. In this context, characterizing the germline susceptibility to RCC, as reported here, can help refine guidelines for genetic screening for renal cancer. The current Canadian guidelines rely primarily on an early age-of-onset (<45 years), the presence of bilateral or multifocal tumors, or a first- or second-degree relative with a history of renal cancer [**25**]. A previous study assessing criteria for genetic screening in Canada identified PVs in 35% of RCC patients who were referred for genetic testing; however, the available data did not allow for an investigation on the fraction of patients with PVs who did not meet referral criteria [**26**]. We realized that within the 56 Canadian patients with identified germline PVs in our study, only 15 (27%) meet the Canadian criteria for referral for genetic screening. Notably, most of those that are eligible are patients with nccRCC (**Figure 4**). As such, 73% of patients with PVs within the Canadian cohort do not meet any of the Canadian criteria. Similarly, when we applied UK [**27**] and the US American College of Medical Genetics (ACMG)[**28**] guidelines to our cohort, they identified only 20% (11/56) and 36% (20/56) of Canadian cases with germline PVs eligible for genetic testing, respectively. We noted that the higher inclusion rate of ACMG guideline is due to the higher minimum age-of-onset applied in these criteria; <50 years in the AMCG vs. <45 years in the Canadian vs. <40 years in the UK

guidelines. Strikingly, Carlo et al. reported similar results for USA and Yngvadottir et al. for UK, with only 54% (14/26) of USA cases, and 17% (13/76) of UK cases with rare germline PVs meeting criteria for referral for genetic screening based on their country's guidelines [**7**,**9**]. Our findings together with those reported by these studies demonstrate that the current guidelines for RCC genetic testing fail to include majority of RCC cases that are affected by rare germline PVs. Specifically, combining data from our study with those from that of the UK and USA (2550 study participants) reveals that 73.4% (116 out of 158) of patients with rare germline PVs are currently excluded from genetic screening, highlighting the need for revising the inclusion criteria to capture more patients affected by actionable mutations.

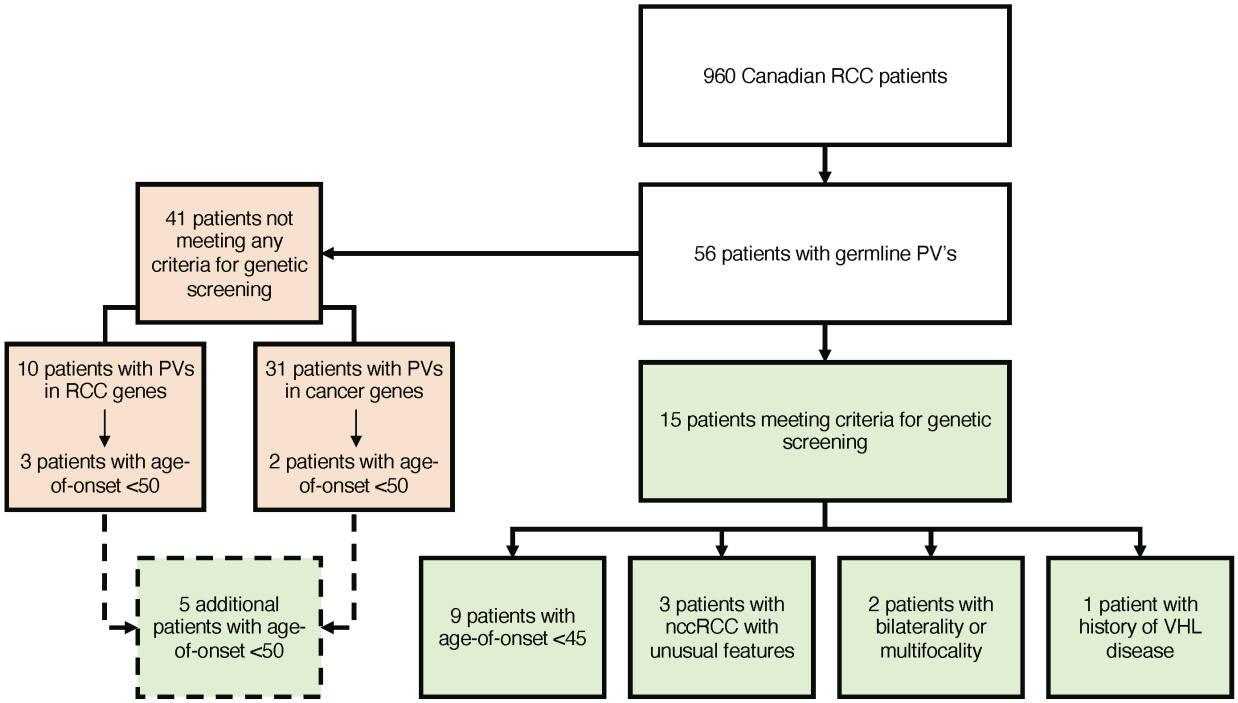


Figure 4. Patients from the Canadian RCC cohort meeting criteria for referral for genetic screening. Increasing the age-of-onset from 45 to 50 years will identify 5 of the 41 patients who do not meet current referral criteria.

In addition to more inclusive guidelines for genetic screening, our results indicate that expanding the current list of genes to include *BRCA1/2*, *CHEK2* and *ATM* for germline screening in Canada will identify more RCC patients that are carrier of rare germline PVs. Similar observations have been noted in other studies, such as within the UK, where *CHEK2* and *MITF* are not routinely included, despite their high mutation rates. Additionally, *BRCA1* and *BRCA2* are not routinely included in germline screening for RCC. Given the observed association between *BRCA1/2* and metastasis, and the vulnerability to PARP inhibitors, identifying RCC patients with germline variants in these genes would provide valuable information to guide treatment decisions, and may warrant their inclusion in genetic testing strategies. Overall, revisions to both the criteria for germline genetic susceptibility to RCC, as well as the inclusion of additional relevant genes have high potentials to identify higher number of at-risk patients and improve precision preventive strategies in RCC. Besides the lack of information about ethnicity of patients, limitations of our study include the consolidation of all nccRCC subtypes into one group, as larger cohorts of nccRCC tumors would be needed to define risk genes specific to each subtype such as papillary and chromophobe RCC. Application of whole-exome or whole-genome sequencing approaches may identify additional risk-genes that were not included in our targeted gene panel.

4.6 AUTHOR CONTRIBUTIONS

Kate I. Glennon (Investigation, Formal Analysis, Funding acquisition, Visualization, Writing – original draft); Mikiko Endo (Investigation, Writing – review & editing, Project administration); Yoshiaki Usui (Investigation, Formal Analysis, Writing – review & editing); Yusuke Iwasaki (Writing – review & editing, Software); Rodney H. Breau (Writing – review & editing); Anil Kapoor (Resources); Mark Lathrop (Writing – review & editing); Simon Tanguay (Conceptualization, Writing – review & editing, Funding acquisition); Yukihide Momozawa

(Conceptualization, Formal Analysis, Writing – review & editing, Funding acquisition, Supervision); Yasser Riazalhosseini (Conceptualization, Writing – original draft, Funding acquisition, Supervision).

4.7 CONFLICTS OF INTERESTS

The authors declare no competing interests.

4.8 FUNDING INFORMATION

KIG was a JSPS International Research Fellow, supported by a Mitacs-JSPS Summer Fellowship for this work. This work was supported by grants from Kidney Foundation of Canada (2020KHRG-673291) to YR and from Japan Agency for Medical Research and Development (JP19kk0305010) to YM.

4.9 DATA AVAILABILITY

Raw sequencing data will be deposited to EGA prior to publication, and accession codes will be provided.

4.10 ACKNOWLEDGEMENTS

The funder did not play a role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication. We are grateful to Madeleine Arseneault for technical assistance with sample handling and shipment. We thank the Laboratory for Genotyping Development in RIKEN and the RIKEN-IMS Genome Platform.

4.11 SUPPLEMENTARY MATERIALS

Supplementary Table S1. RCC-related and Cancer predisposition genes analyzed in this study.

Gene	Inclusion criteria
BAP1	RCC-related
CDC73	RCC-related
FH	RCC-related
FLCN	RCC-related
KDM5C	RCC-related
MET	RCC-related
MITF	RCC-related
MTOR	RCC-related
PBRM1	RCC-related
SDHA	RCC-related
SDHB	RCC-related
SDHC	RCC-related
SDHD	RCC-related
SETD2	RCC-related
SMARCB1	RCC-related
TSC1	RCC-related
TSC2	RCC-related
VHL	RCC-related
WT1	RCC-related
APC	Cancer susceptibility
ATM	Cancer susceptibility
BARD1	Cancer susceptibility
BMPR1A	Cancer susceptibility
BRCA1	Cancer susceptibility
BRCA2	Cancer susceptibility
BRIP1	Cancer susceptibility
CDH1	Cancer susceptibility
CDK4	Cancer susceptibility
CDKN2A	Cancer susceptibility
CHEK2	Cancer susceptibility
EPCAM	Cancer susceptibility
HOXB13	Cancer susceptibility
MLH1	Cancer susceptibility
MSH2	Cancer susceptibility
MSH6	Cancer susceptibility
MUTYH	Cancer susceptibility
NBN	Cancer susceptibility
NF1	Cancer susceptibility
PALB2	Cancer susceptibility
PMS2	Cancer susceptibility
PTEN	Cancer susceptibility
RAD51C	Cancer susceptibility
RAD51D	Cancer susceptibility
SMAD4	Cancer susceptibility
STK11	Cancer susceptibility
TP53	Cancer susceptibility

Supplementary Table S2. Clinical information for ccRCC and nccRCC groups.

Clinical Features		ccRCC	nccRCC	p-value	lower CI	upper CI	OR
Sex	Female	277	63	0.185	0.9	1.8	1.3
	Male	482	138				
Smoking	No	237	78	0.044	0.5	1.0	0.7
	Yes	410	95				
Stage of Cancer	I	409	114	0.009			
	II	41	32				
	III	269	41				
	IV	11	2				
Personal History of RCC	No	664	168	0.127	0.8	4.1	1.9
	Yes	23	11				
Personal history of other cancers	No	542	127	0.028	1.0	2.2	1.5
	Yes	145	52				
Family history of RCC	No	418	124	0.040	0.0	1.0	0.2
	Yes	28	2				
	Not Available						
Family history of other cancers	No	201	53	0.612	0.7	1.7	1.1
	Yes	245	73				
	Not Available						
nccRCC Subtype	Papillary	-	103				
	Chromophobe	-	50				
	Mixed histology	-	15				
	Unclassified	-	17				
	Other	-	16				

Supplementary Table S3. Germline pathogenic variants.

Chr	Pos	Ref	Alt	snp151	Annotation	Annotation Impact	Gene	nHET	nHomAlt
22	29091856	AG	A	rs555607708	frameshift	HIGH	CHEK2	11	0
1	45797228	C	T	rs36053993	missense &splice_region	MODERATE	MUTYH	6	0
1	45798475	T	C	rs34612342	missense	MODERATE	MUTYH	5	0
3	70014091	G	A	rs149617956	missense	MODERATE	MITF	5	1
1	241663833	CT	C	rs398123163	frameshift	HIGH	FH	2	0
22	29121230	C	T	rs121908698	splice_donor &intron	HIGH	CHEK2	2	0
1	17359573	G	A	rs74315366	stop_gained	HIGH	SDHB	1	0
1	241661223	A	ATCCATTTT	rs863223994	frameshift	HIGH	FH	1	0
1	241671944	G	A	rs587781682	missense	MODERATE	FH	1	0
2	47601174	C	T	rs397514661	stop_gained	HIGH	EPCAM	1	0
2	215595231	C	T	rs773223671	stop_gained &splice_region	HIGH	BARD1	1	0
3	10183748	C	T	rs869025619	stop_gained	HIGH	VHL	1	0
9	21971114	CGGGTCGGG TGAGAGTGGCG	C	rs730881674	frameshift	HIGH	CDKN2A	1	0
11	108119829	G	A	rs587779813	stop_gained &splice_region	HIGH	ATM	1	0
11	108121625	CAG	C	.	frameshift	HIGH	ATM	1	0
11	108121752	CAG	C	rs1374409941; rs587779817	frameshift	HIGH	ATM	1	0
11	108142134	G	T	rs192810283	splice_donor &intron	HIGH	ATM	1	0
11	108199929	T	G	rs28904921	missense	MODERATE	ATM	1	0
11	108205832	T	C	rs587782652	missense	MODERATE	ATM	1	0
11	108216623	ACTT	A	rs1340074139; rs786203976	conservative _inframe _deletion	MODERATE	ATM	1	0
13	32907420	G	GA	rs1253401667; rs80359306	frameshift	HIGH	BRCA2	1	0
13	32913648	A	AT	rs80359489	frameshift	HIGH	BRCA2	1	0
13	32944580	AC	A	rs398122605	frameshift	HIGH	BRCA2	1	0
13	32972321	T	TA	rs1423835974; rs80359773	frameshift	HIGH	BRCA2	1	0
16	23649275	T	C	rs730881897	splice_acceptor &intron	HIGH	PALB2	1	0
17	17122407	AG	A	rs878855221	frameshift	HIGH	FLCN	1	0
17	29667521	A	G	rs864622509	splice_acceptor &intron	HIGH	NF1	1	0
17	33433425	G	A	rs387906843	stop_gained	HIGH	RAD51D	1	0
17	41245159	C	A	rs62625306	stop_gained	HIGH	BRCA1	1	0
17	56787218	A	G	rs587780259	splice_acceptor &intron	HIGH	RAD51C	1	0
17	59763414	TATGG	T	rs760551339	frameshift	HIGH	BRIP1	1	0
17	59878709	C	G	rs149364097	missense	MODERATE	BRIP1	1	0

Continued Supplementary Table S3. Germline pathogenic variants.

22	29091226	TA	T	rs587780174	frameshift	HIGH	CHEK2	1	0
22	29121326	T	C	rs28909982	missense	MODERATE	CHEK2	1	0
1	17380442	C	G	.	splice_donor &intron	HIGH	SDHB	1	0
2	47601092	C	CA	.	frameshift	HIGH	EPCAM	1	0
3	47058584	CA	C	.	frameshift &stop_lost	HIGH	SETD2	1	0
17	17118345	CCA	C	.	frameshift	HIGH	FLCN	1	0
17	59793412	G	A	rs137852986	stop_gained	HIGH	BRIP1	1	0
22	29083920	TG	T	rs587781519	frameshift	HIGH	CHEK2	1	0
8	90983441	ATTTGT	A	rs587776650	frameshift	HIGH	NBN	0	1

Supplementary Table S4. Association analysis - pathogenic variants compared to gnomad control cohort.

Variant Info					Allele carriers				Firth Logistic Regression				
Chr	Pos	Ref	Alt	Gene	Carriers RCC	Non Carriers RCC	PV Carriers gnomad	Non Carriers gnomad	p-value	lower CI	upper CI	OR	p.adjust
22	29091856	AG	A	CHEK2	11	949	302	116902	3.62E-05	0.9	2.1	4.7	9.41E-04
3	70014091	G	A	MITF	6	954	278	118076	0.023	0.2	1.8	2.9	0.586
22	29121230	C	T	CHEK2	2	958	16	118000	0.002	1.3	4.1	18.7	0.051
2	47601174	C	T	EPCAM	1	959	1	102474	0.002	2.2	7.2	106.8	0.043
2	215595231	C	T	BARD1	1	959	1	102576	0.002	2.2	7.2	106.9	0.043
11	108119829	G	A	ATM	1	959	1	15414	0.033	0.3	5.3	16.1	0.865
11	108121752	CAG	C	ATM	1	959	8	102512	0.017	0.7	4.4	18.9	0.430
11	108142134	G	T	ATM	1	959	0	102228	0.001	2.8	10.8	319.6	0.015
11	108199929	T	G	ATM	1	959	10	117860	0.018	0.6	4.3	17.5	0.472
11	108205832	T	C	ATM	1	959	4	118110	0.005	1.4	5.4	41.0	0.137
11	108216623	ACTT	A	ATM	1	959	0	102458	0.001	2.8	10.8	320.3	0.015
13	32907420	G	GA	BRCA2	1	959	0	95590	0.001	2.8	10.7	298.9	0.017
13	32972321	T	TA	BRCA2	1	959	1	102220	0.002	2.2	7.2	106.5	0.043
17	33433425	G	A	RAD51D	1	959	5	117418	0.007	1.2	5.1	33.4	0.183
17	41245159	C	A	BRCA1	1	959	1	102196	0.002	2.2	7.2	106.5	0.043
17	56787218	A	G	RAD51C	1	959	6	118076	0.009	1.1	4.9	28.4	0.231
17	59763414	TATGG	T	BRIP1	1	959	3	118056	0.004	1.6	5.8	52.7	0.097
17	59878709	C	G	BRIP1	1	959	1	118044	0.001	2.3	7.3	123.0	0.034
22	29091226	TA	T	CHEK2	1	959	10	117514	0.018	0.6	4.3	17.5	0.474
22	29121326	T	C	CHEK2	1	959	22	117950	0.058	-0.1	3.5	8.2	1.000
17	59793412	G	A	BRIP1	1	959	29	116500	0.089	-0.4	3.2	6.2	1.000
22	29083920	TG	T	CHEK2	1	959	3	115554	0.004	1.6	5.7	51.6	0.101
8	90983441	ATTTGT	A	NBN	1	959	44	117832	0.162	-0.8	2.7	4.1	1.000

Supplementary Table S5. Gene burden by RCC subtype compared to gnomad control cohort.

ccRCC									
Carrier Info					Logistic Regression				
Gene	Carriers ccRCC	Non Carriers ccRCC	PV Carriers gnomad	Non Carriers gnomad	p-value	lower CI	upper CI	OR	p.adjust
MITF	5	754	279	117720	0.027	1.2	6.5	3.1	0.439
CHEK2	14	745	474	117672	2.46E-06	2.7	7.9	4.8	3.94E-05
ATM	7	752	259	117903	0.001	2.0	8.7	4.5	0.016
VHL	1	758	3	102739	0.003	5.6	353.3	58.1	0.051
CDKN2A	1	758	14	111396	0.022	1.7	61.6	15.2	0.356
SDHB	1	758	20	118064	0.034	1.3	44.4	11.4	0.552
BRCA2	4	753	230	117852	0.047	1.0	6.9	3.1	0.753
BRIP1	2	757	82	118048	0.052	1.0	13.6	4.7	0.835
RAD51D	1	758	28	117390	0.058	0.9	31.0	8.1	0.925
BARD1	1	758	49	118051	0.133	0.5	17.4	4.7	1.000
BRCA1	1	758	138	118046	0.558	0.2	6.1	1.7	1.000
EPCAM	1	758	86	118044	0.299	0.3	9.8	2.7	1.000
FLCN	2	757	124	118026	0.132	0.7	8.9	3.1	1.000
NBN	1	758	44	118046	0.113	0.6	19.5	5.2	1.000
PALB2	1	758	85	118095	0.294	0.3	9.9	2.7	1.000
RAD51C	1	758	63	118113	0.192	0.4	13.5	3.7	1.000
nccRCC									
Carrier Info					Logistic Regression				
Gene	Carriers nccRCC	Non Carriers nccRCC	PV Carriers gnomad	Non Carriers gnomad	p-value	lower CI	upper CI	OR	p.adjust
FH	4	197	12	118016	7.67E-10	64.4	597.8	215.1	6.14E-09
SETD2	1	200	3	102749	3.82E-04	21.2	1341.8	219.6	0.003
MITF	1	200	279	117720	0.239	0.4	11.4	3.2	1.000
SDHB	1	200	20	118064	0.004	4.8	169.1	43.1	0.033
NF1	1	200	34	118082	0.009	2.9	96.8	25.6	0.074
BRIP1	1	200	82	118048	0.037	1.2	39.2	10.7	0.294
EPCAM	1	200	86	118044	0.040	1.2	37.4	10.2	0.317
CHEK2	2	199	474	117672	0.133	0.6	8.8	3.1	1.000

Supplementary Table S6. Clinical information for germline carriers vs non-germline carriers.

Clinical Features		No Germline	Germline	P-value	lower CI	upper CI	OR
Sex	Female	324	16	0.315	0.8	2.7	1.4
	Male	580	40				
Smoking	No	291	24	0.242	0.4	1.3	0.7
	Yes	477	28				
Stage of Cancer	I	492	31	0.560			
	II	71	2				
	III	292	18				
	IV	10	3				
Personal History of RCC	No	785	47	0.134	0.5	6.7	2.2
	Yes	30	4				
Personal history of other cancers	No	628	41	0.731	0.4	1.7	0.8
	Yes	187	10				
Family history of RCC	No	504	38	0.250	0.0	1.8	0.0
	Yes	30	0				
	Not Available						
Family history of other cancers	No	240	14	0.399	0.7	3.0	1.4
	Yes	294	24				
	Not Available						

Supplementary Table S7. Gene burden comparison among significant risk-genes within different population cohorts.

Canada vs Other					LRF				
Gene	Allele Carriers Canadian	Non-Allele Carriers Canadian	Allele Carriers Other	Non-Allele Carriers Other	p-value	CI.min	CI.max	OR	p.adjust
CHEK2	16	944	43	3079	0.473	0.7	2.1	1.2	1.000
MITF	6	954	10	3112	0.176	0.7	5.3	2.0	1.000
ATM	7	953	15	3107	0.322	0.6	3.7	1.6	1.000
FH	4	956	13	3109	0.884	0.3	2.9	1.1	1.000
BAP1	0	960	9	3113	0.108	0.0	1.3	0.2	0.650
TP53	0	960	13	3109	0.036	0.0	0.9	0.1	0.217
UK vs Other					**LRF**				
Gene	Allele Carriers UK	Non-Allele Carriers UK	Allele Carriers Other	Non-Allele Carriers Other	p-value	CI.min	CI.max	OR	p.adjust
CHEK2	27	1309	32	2714	0.034	1.0	2.9	1.8	0.203
MITF	10	1326	6	2740	0.014	1.3	9.4	3.3	0.085
ATM	10	1326	12	2734	0.197	0.7	4.0	1.7	1.000
FH	3	1333	14	2732	0.207	0.1	1.4	0.5	1.000
BAP1	0	1336	9	2737	0.030	0.0	0.8	0.1	0.182
TP53	2	1334	11	2735	0.210	0.1	1.5	0.4	1.000
US vs Other					**LRF**				
Gene	Allele Carriers US	Non-Allele Carriers US	Allele Carriers Other	Non-Allele Carriers Other	p-value	CI.min	CI.max	OR	p.adjust
CHEK2	9	245	50	3778	0.009	1.3	5.6	2.9	0.051
MITF	0	254	16	3812	0.532	0.0	3.4	0.5	1.000
ATM	1	253	21	3807	0.956	0.1	4.1	1.0	1.000
FH	7	247	10	3818	1.55E-05	4.1	28.2	11.0	9.28E-05
BAP1	3	251	6	3822	0.007	1.9	29.0	8.2	0.040
TP53	0	254	13	3815	0.654	0.0	4.2	0.6	1.000
Japan vs Other					**LRF**				
Gene	Allele Carriers Japanese	Non-Allele Carriers Japanese	Allele Carriers Other	Non-Allele Carriers Other	p-value	CI.min	CI.max	OR	p.adjust
CHEK2	7	1525	52	2498	1.22E-05	0.1	0.5	0.2	7.30E-05
MITF	0	1532	16	2534	0.001	0.0	0.4	0.1	0.003
ATM	4	1528	18	2532	0.061	0.1	1.0	0.4	0.366
FH	3	1529	14	2536	0.095	0.1	1.2	0.4	0.571
BAP1	6	1526	3	2547	0.079	0.9	13.0	3.1	0.472
TP53	11	1521	2	2548	0.001	2.3	39.7	7.7	0.004

4.12 REFERENCES

1. Bukavina L, Bensalah K, Bray F, et al. Epidemiology of Renal Cell Carcinoma: 2022 Update. *Eur Urol* 2022; **82**(5): 529-42.
2. Motzer RJ, Jonasch E, Agarwal N, et al. Kidney Cancer, Version 3.2022, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* 2022; **20**(1): 71-90.
3. Ahrens M, Scheich S, Hartmann A, Bergmann L. Non-Clear Cell Renal Cell Carcinoma - Pathology and Treatment Options. *Oncol Res Treat* 2019; **42**(3): 128-35.
4. Karami S, Schwartz K, Purdue MP, et al. Family history of cancer and renal cell cancer risk in Caucasians and African Americans. *Br J Cancer* 2010; **102**(11): 1676-80.
5. Nguyen KA, Syed JS, Espenschied CR, et al. Advances in the diagnosis of hereditary kidney cancer: Initial results of a multigene panel test. *Cancer* 2017; **123**(22): 4363-71.
6. Carlo MI, Hakimi AA, Stewart GD, et al. Familial Kidney Cancer: Implications of New Syndromes and Molecular Insights. *Eur Urol* 2019; **76**(6): 754-64.
7. Yngvadottir B, Andreou A, Bassaganyas L, et al. Frequency of pathogenic germline variants in cancer susceptibility genes in 1336 renal cell carcinoma cases. *Hum Mol Genet* 2022; **31**(17): 3001-11.
8. Sekine Y, Iwasaki Y, Aoi T, et al. Different risk genes contribute to clear cell and non-clear cell renal cell carcinoma in 1532 Japanese patients and 5996 controls. *Hum Mol Genet* 2022; **31**(12): 1962-9.
9. Carlo MI, Mukherjee S, Mandelker D, et al. Prevalence of Germline Mutations in Cancer Susceptibility Genes in Patients With Advanced Renal Cell Carcinoma. *JAMA Oncol* 2018; **4**(9): 1228-35.

10. Momozawa Y, Akiyama M, Kamatani Y, et al. Low-frequency coding variants in CETP and CFB are associated with susceptibility of exudative age-related macular degeneration in the Japanese population. *Hum Mol Genet* 2016; **25**(22): 5027-34.

11. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 2012; **6**(2): 80-92.

12. Abou Tayoun AN, Pesaran T, DiStefano MT, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat* 2018; **39**(11): 1517-24.

13. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020; **581**(7809): 434-43.

14. Xiang HP, Geng XP, Ge WW, Li H. Meta-analysis of CHEK2 1100delC variant and colorectal cancer susceptibility. *Eur J Cancer* 2011; **47**(17): 2546-51.

15. Weischer M, Heerfordt IM, Bojesen SE, et al. CHEK2*1100delC and risk of malignant melanoma: Danish and German studies and meta-analysis. *J Invest Dermatol* 2012; **132**(2): 299-303.

16. Schmidt MK, Hogervorst F, van Hien R, et al. Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. *J Clin Oncol* 2016; **34**(23): 2750-60.

17. Hale V, Weischer M, Park JY. CHEK2 (∗) 1100delC Mutation and Risk of Prostate Cancer. *Prostate Cancer* 2014; **2014**: 294575.

18. CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. *Am J Hum Genet* 2004; **74**(6): 1175-82.

19. Bychkovsky BL, Agaoglu NB, Horton C, et al. Differences in Cancer Phenotypes Among Frequent CHEK2 Variants and Implications for Clinical Care-Checking CHEK2. *JAMA Oncol* 2022; **8**(11): 1598-606.

20. Han S, Camp SY, Chu H, et al. Integrative Analysis of Germline Rare Variants in Clear and Non-Clear Cell Renal Cell Carcinoma. *medRxiv* 2023.

21. Guhan SM, Artomov M, McCormick S, et al. Cancer risks associated with the germline MITF(E318K) variant. *Sci Rep* 2020; **10**(1): 17051.

22. Bertolotto C, Lesueur F, Giuliano S, et al. A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma. *Nature* 2011; **480**(7375): 94-8.

23. Farmer H, McCabe N, Lord CJ, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 2005; **434**(7035): 917-21.

24. Simonaggio A, Ambrosetti D, Verkarre V, Auvray M, Oudard S, Vano YA. MiTF/TFE Translocation Renal Cell Carcinomas: From Clinical Entities to Molecular Insights. *Int J Mol Sci* 2022; **23**(14).

25. Reaume MN, Graham GE, Tomiak E, et al. Canadian guideline on genetic screening for hereditary renal cell cancers. *Can Urol Assoc J* 2013; **7**(9-10): 319-23.

26. Kushnir I, Kirk L, Mallick R, et al. Application of Hereditary Renal Cell Carcinoma Risk Criteria to a Large Prospective Database. *Clin Oncol (R Coll Radiol)* 2020; **32**(1): e10-e5.

27. NHS. National Genomic Test Directory: Testing Criteria for Rare and Inherited Disease (V4 October 2022). 2022. https://www.england.nhs.uk/publication/national-genomic-test-directories/ (accessed March 23 2023).

28. Hampel H, Bennett RL, Buchanan A, Pearlman R, Wiesner GL. A practice guideline from the American College of Medical Genetics and Genomics and the National Society of

Genetic Counselors: referral indications for cancer predisposition assessment. *Genet Med* 2015; **17**(1): 70-87.

CHAPTER 5. GENERAL DISCUSSION

This thesis focused on identifying and characterizing genomic patterns underlying RCC tumors, and their clinical implications. In Chapter 2 we developed and optimized an NGS assay targeting RCC-relevant genes, applicable to various sample types including tissue and liquid biopsies. This provided a much-needed tool to facilitate future investigations into the genomics of RCC, and its clinical utility. We also demonstrate the potential of liquid biopsies for capturing relevant somatic mutations from primary RCC tumors. In Chapter 3 we generated and analyzed profiles of somatic mutation of RCC-relevant genes within the largest-to-date cohort of ccRCCs with detailed clinical annotations. Based on patterns of genes harbouring mutations, we defined a genomic classifier capable of identifying groups of ccRCC patients with divergent relapse rates. We proposed how this classifier could be used for personalized approaches to adjuvant therapy based on the risk of disease recurrence. In Chapter 4, we expanded our analysis to germline pathogenic variants that increase the risk of developing RCC. We identified risk genes for RCC within Canada and characterized differences in risk genes among cohorts from various countries. Lastly, we identified the global need to revise clinical genetic screening guidelines for RCC, as they fail to identify over 70% of patients with rare germline pathogenic variants.

Chapters 2.6, 3.6, and 4.5 discuss the biological relevance and highlight the translational findings of each chapter independently. Here we discuss how these findings relate to each other, limitations and analyses addressing these limitations that were not included in the published manuscripts, and finally how the findings are relevant for integrating genomics into public health approaches for RCC.

5.1 LEVERAGING TUMOR EVOLUTION FOR PROGNOSTIC MARKERS

Current clinical nomograms and prognostic algorithms for RCC are based solely on clinicopathological aspects[56-62], and are not very successful in assessing risk of relapse. The genomic classifier defined in Chapter 3 demonstrates how considering the underlying biology of a tumor may provide more accurate risk prediction, and how genomics can be feasibly integrated into the clinic for risk-assessment by informing decisions for adjuvant therapy. Additionally, the CAGEKID cohort presented in this chapter represents the largest-to-date cohort of ccRCCs with high confidence in the detection of somatic mutations, and in-depth clinical information.

Aside from associations to disease outcomes, the genomic patterns identified in Chapter 3 provide insight into the evolution of RCC tumors at a large scale, supporting the proposed evolutionary subtypes of ccRCC defined within the TRACeRx Renal cohort[105]. In particular, the clinical phenotypes correlated with multiple clonal driver and VHL monodriver subtypes are characterized within the genomic classifier. Of note, Turajlic et al. suggested that four to eight site biopsies would be required in order to capture the majority of subclones within a tumor[105], which is not feasible in a clinical setting. However, we demonstrate that within a single tumor biopsy, we could capture enough information about the evolutionary patterns within the tumor to be able to classify them into risk groups.

The genomic classifier does not encompass all evolutionary subtypes of ccRCC, notably due to the exclusion of genes within the PI3K/AKT/mTOR pathway, and CNAs, which are defining features of *PBRM1*-driven subtypes. The PI3K/AKT/mTOR pathway, which is implicated in ~28% of ccRCCs[68,76] is a defining feature of the *PBRM1* $\rightarrow$ *PI3K* evolutionary subtype of ccRCC. However, mutation rates of individual genes within this pathway are low, and would call for an even larger patient cohort to reach sufficient statistical power for association

analyses. Additionally, as detection of somatic CNAs (sCNAs) from a small, targeted sequencing panel is not conclusively sensitive, CNAs were not integrated into the classifier. Notably, the incorporation of CNAs with known associations to prognosis would likely improve the genomic classifier performance, in particular for evaluating the metastatic potential of ccRCCs. Analyses of metastatic ccRCCs have indicated that elevated levels of somatic CNAs, rather than increased driver SNVs/indels, are associated with metastatic competence. Specifically, the loss of chromosome 9p has been identified as a driver of metastasis and is associated with increased mortality risk[1], with loss of 14q showing a similar trend. Overall, incorporating PI3K/AKT/mTOR pathway genes and sCNAs into the classifier offers the opportunity for refinement, particularly for *PBRM1*-driven tumors, which may have different risks of disease recurrence. The *VHL*+≥3 risk group also has the potential for further refinement into subgroups, as combinations of clonal drivers may drive outcome in diverse manners.

Given the considerable sex-bias that exists in RCC, evaluating if genomic evolutionary trajectories act in a sex-dependent manner could provide insight into the molecular mechanisms driving this bias. Several sex-dependent trends have been observed across cancer types in both the context of patient survival and tumor evolution. It is likely that male and female tumors acquire mutations of different types, or at different rates[137], and pan-cancer analysis of TCGA data identified sex differences in coding and non-coding drivers, mutation prevalence, mutational signatures, and in the timing of truncal structural variants and indels[137,138]. As sex-dependent incidence rates and outcome are hallmark of RCC[139-141], it is also important to investigate evolutionary subtypes in the context of patient sex. Therefore, we expanded on the analyses described in Chapter 3 to evaluate sex-differences that may be captured by the genomic classifier. Analyzed independently, both male and female patients follow the previously described genomic

classification based on the number of driver mutations, with increased risk of disease-recurrence and RCC-related death increasing with the number of mutated driver genes (**Figure 1A-B**). Interestingly, within the *VHL*+≥3 groupings, females may have a greater risk of RCC-related death than male patients (Female HR: 10.24 [3.00;34.95], p=0.0002; Male HR: 3.46 [1.39 ;8.61]; p=0.007) (**Table 1**, **Figure 1C-D**). As in Chapter 3, multivariable models were stratified for TNM stage, and incorporated patient age, to account for differences in the makeup of male and female cohorts. Additionally, the mutation rates of genes in *VHL*+≥3 patients differ between male and female patients (**Figure 1E**). Mutations in *ATM* are more frequent in female patients within the *VHL*+≥3 group, whereas mutations in *KDM5C* are more frequent in male patients (**Table 2**). This provides evidence that the genes driving poor outcome in ccRCC may differ between male and female patients, and evolutionary trajectories of RCC should be evaluated in the context of patient sex.

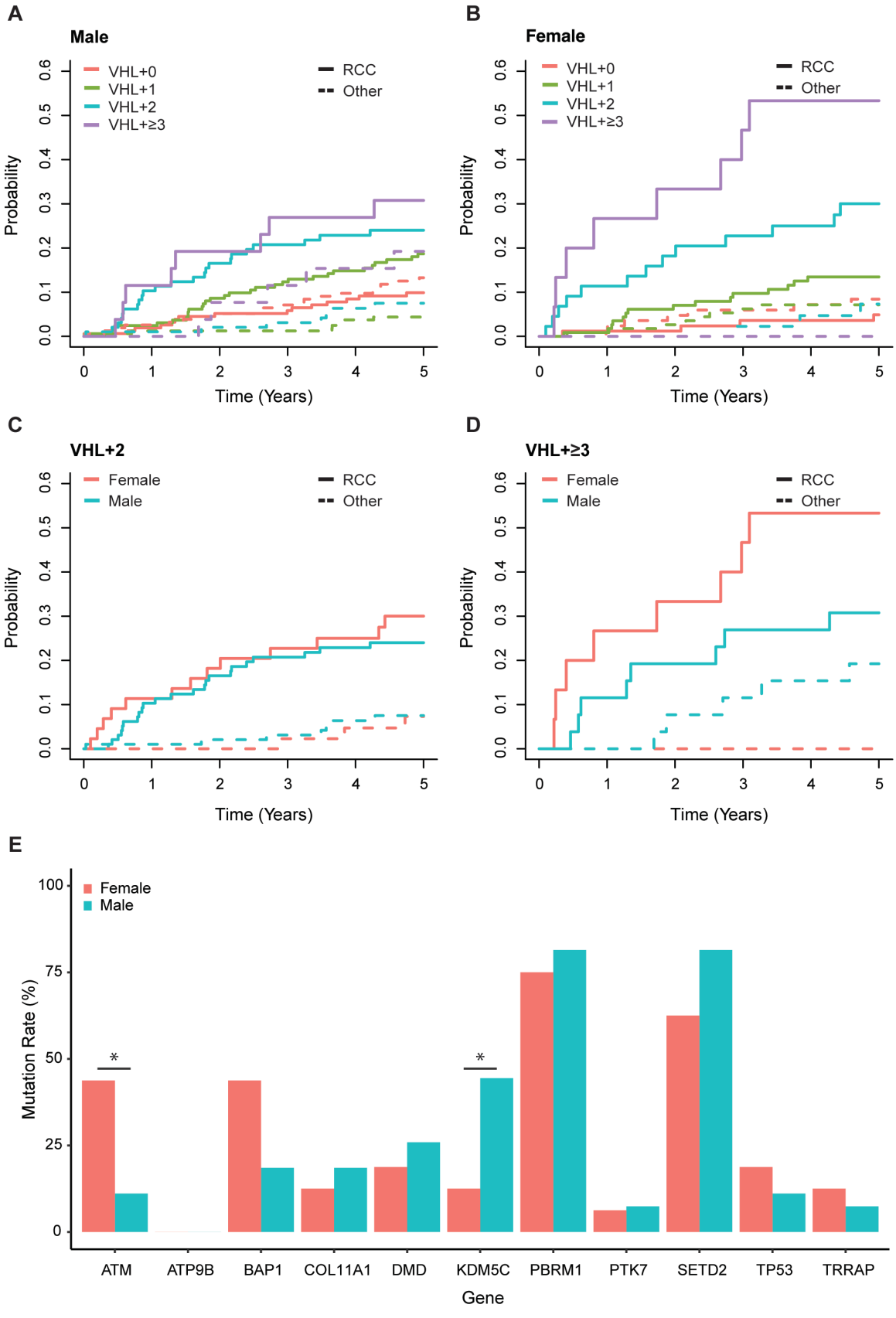

A
Male
VHL+0
VHL+1
VHL+2
VHL+≥3
RCC
Other
Probability
Time (Years)
B
Female
C
VHL+2
Female
Male
D
VHL+≥3
E
Mutation Rate (%)
ATM
ATP9B
BAP1
COL11A1
DMD
KDM5C
PBRM1
PTK7
SETD2
TP53
TRRAP
Gene

Figure 1. Genomic classification in the context of patient sex. 5-year Competing risks analysis of RCC-related death compared to non-RCC-related death ("other") for our genomic classifier in the context of patient sex. We see a similar trend as seen in the overall cohort within both (**A**) male patients (N=445) and (**B**) female patients (N=260) with. Male and female patients show similar disease-free survival within the *VHL*+2 group (**C**). Female patients show a greater probability of RCC-related death within the *VHL*+≥3 group (**D**). (**E**) Mutation rates of genes within *VHL*+≥3 category for male (N=27) and female (N=16) patients within the CAGEKID cohort. Fisher's exact tests compare mutation frequency between sexes. (* $p < 0.05$)

Table 1. Cox-Proportional-Hazards model for RCC-related and non-RCC related death within female and male groupings.

Female							
		RCC-related death			Non-RCC-related death		
Predictor	**Category**	**HR**	**95% CI**	**p-value**	**HR**	**95% CI**	**p-value**
VHL Category	VHL+0	1.00	-	-	1.00	-	-
	VHL+1	2.89	[0.94;8.83]	0.063	0.82	[0.29;2.30]	0.708
	VHL+2	3.71	[1.18;11.69]	0.025	7.6	[0.19;3.08]	0.704
	VHL+≥3	10.24	[3.00;34.95]	0.0002	-	[0.0000; Inf]	0.997
Age		1.00	[0.97;1.04]	0.938	1.06	[1.01;1.12]	0.0232
Male							
		RCC-related death			Non-RCC-related death		
Predictor	**Category**	**HR**	**95% CI**	**p-value**	**HR**	**95% CI**	**p-value**
VHL Category	VHL+0	1.00			1.00		
	VHL+1	1.65	[0.8795;3.089]	0.119	0.26	[0.1109;0.6287]	0.003
	VHL+2	2.54	[1.3042;4.950]	0.006	0.46	[0.1938;1.1054]	0.083
	VHL+≥3	3.46	[1.3944 ;8.605]	0.007	1.21	[0.4428;3.2997]	0.711
Age		1.00	[0.9776;1.026]	0.893	1.07	[1.0307;1.1021]	0.0002

*Evaluated over a 5-year time-period

Table 2. Fisher's exact test for genes mutated within male and female *VHL*+≥3 patients.

Gene	Male (N=27)	Female (N=16)	p-value
ATM	3	7	0.02437
ATP9B	0	0	-
BAP1	5	7	0.09204
COL11A1	5	2	0.695
DMD	7	3	0.719
KDM5C	12	2	0.04471
PBRM1	22	12	0.706
PTK7	2	1	1
SETD2	22	10	0.2781
TP53	3	3	0.6546
TRRAP	2	2	0.6208

5.2 LIQUID BIOPSY SOLUTIONS FOR MANAGEMENT OF RCC

Facilitating side-by-side comparisons of tissue and liquid biopsies is crucial for evaluating the utility of liquid biopsies, in particular ctDNA analysis. Chapter 2 focused to investigate the application of the RCC-NGS assay to blood-based liquid biopsies, while also demonstrating how the assay is robust when applied to genomic DNA from various sources including buffy coat, fresh frozen and FFPE tissues. We further demonstrate this in Chapter 3, through application to tumor-normal pairs from 474 patients (Cohort 3). By developing an RCC-appropriate targeted NGS assay, we have provided a robust workflow for future large-scale studies investigating the utility of ctDNA analysis in the management of kidney cancer. These studies should include longitudinal approaches and surveillance strategies for early diagnosis of relapse in at-risk individuals, and large-scale investigations of the concordance between tumors and ctDNA, including the ability to capture genomic classifiers within ctDNA.

Analysis of ctDNA has the opportunity for a role in screening and diagnosis, prognostication, and guiding treatment decisions for RCC. Screening of healthy individuals to detect the presence of ctDNA in blood-based liquid biopsies offers a solution for early detection and diagnosis of RCC, especially for those with genetic susceptibility to RCC. For individuals carrying rare germline pathogenic variants, ctDNA in the blood may be detected earlier than by imaging methods, and prior to the onset of any symptoms[131]. Liquid biopsies may also play a prognostic role in identifying patients at high-risk of disease recurrence. Even in the absence of a tumor biopsy, the somatic mutational analysis of ctDNA may be a minimally invasive method for capturing genomic classifiers for risk-stratification. ctDNA analyses may even provide more sensitive classification due to the ability of ctDNA to represent diverse subclonal mutations without the bias of spatial heterogeneity in tumor biopsy[107]. Lastly, ctDNA analysis is valuable for

post-operative screening for the detection of minimal residual disease (MRD), to monitor for disease recurrence[142], particularly for individuals who are at high risk of disease recurrence, and for real-time monitoring of response to treatment and the development of drug resistance. However, work to increase the sensitivity for ctDNA detection in low stage RCCs, along with establishing a background false-positive rate through a panel of healthy controls would be necessary for the implementation of ctDNA-based surveillance strategies.

A major limitation of the NGS assay presented in Chapter 2 was the inability to detect mutations in cfDNA from low-stage RCCs. We highlighted the need to increase assay sensitivity in this group, however the low fraction of ctDNA in the plasma of RCC patients renders this extremely difficult. Since this manuscript was published in 2021, research in liquid biopsy has explored new areas including investigating differences in fragment sizes between healthy cfDNA and tumor-derived ctDNA. The fragment size of ctDNA is slightly smaller than cfDNA from healthy cell, and studies have demonstrated the ability to enrich for smaller DNA fragments using size-selection methods, which results in increasing the fraction of tumor-derived fragments within a cfDNA sample[131]. Other studies have focused on characterizing the properties of ctDNA through 'fragmentomics', examining genomic regions with recurrently altered fragment profiles in ctDNA, and motifs in conserved regions that also could be leveraged to enrich for ctDNA[132,142,143]

We demonstrated the proof-of-principle for size-selection enrichment for RCC, increasing the tumor-fraction within a cfDNA NGS library by implementing a size selection step prior to library generation. To select for smaller fragment sizes, we used the BluePippin DNA Size Selection System, with a 3% agarose gel cassette. We applied size selection to samples from 3 RCC patients to investigate the differences the tumor fraction in cfDNA samples with and without the implementation of size-selection. We prepared libraries from size-selected and non-size-

selected cfDNA from the same patient, and conducted shallow WGS of all libraries to evaluate the tumor fraction detected within each sample. We observed that the insert-sizes for each library were representative of the size-selection (**Figure 2A**). Lastly, we estimated tumor fraction of each library using ichorCNA, and though the tumor fraction is low in all samples, we see an increase in tumor fraction in the size-selected samples (**Figure 2B**). Though preliminary, this analysis demonstrates how integrating information about the biological features of ctDNA may offer a solution for increasing the sensitivity of the assay for low stage RCCs.

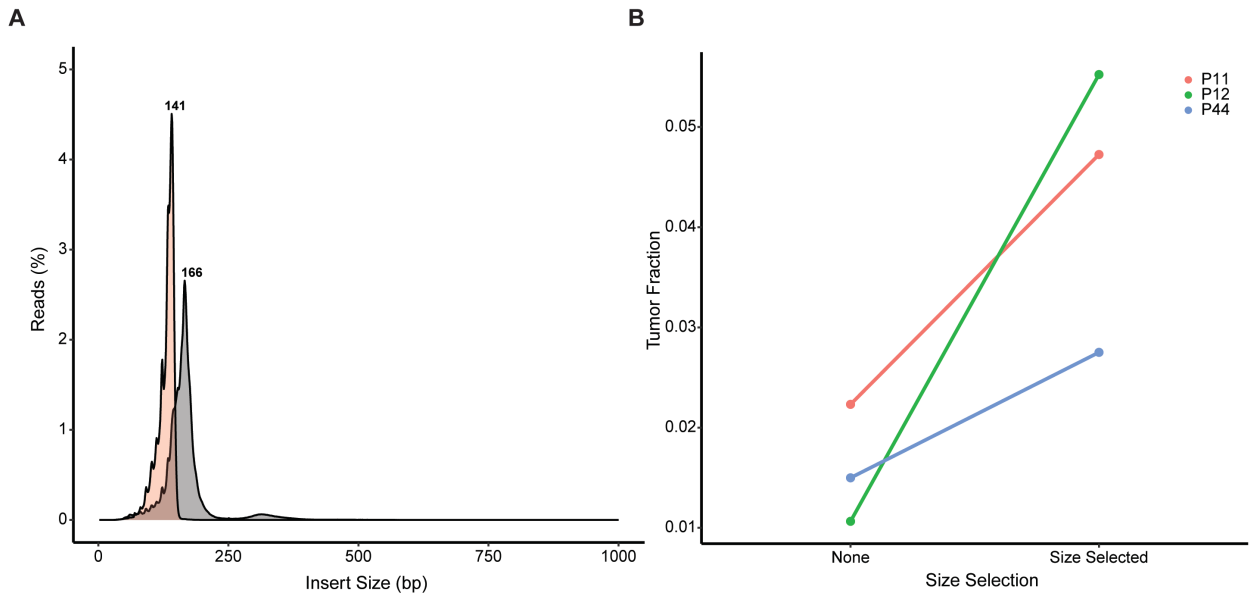


Figure 2. Enrichment of ctDNA with size-selection. (**A**) Insert sizes from cfDNA libraries prepared with (red) and without (grey) size-selection. (**B**) Estimated tumor fraction of cfDNA libraries with no size-selection ('None') and with size-selection ('Size Selected').

5.3 IMPROVED GENETIC SCREENING FOR RCC

In Chapter 3, we highlighted the need to refine clinical guidelines for referral to genetic screening in RCC. Our results show that current criteria fail to identify >70% of individuals with rare germline pathogenic variants associated to RCC, indicating that we are underestimating the

susceptibility to RCC within Canada. Indeed, the clinical impact of the Kidney Cancer Research Network of Canada (KCRNC) criteria[32] has never been assessed, and the criteria have not been validated for their ability to capture patients with germline susceptibility to RCC[144]. Other studies have also highlighted the need for more inclusive screening guidelines in the US[145], though in smaller study cohorts. Efforts to improve screening for RCC should focus on expanding criteria to capture more individuals carrying rare PVs, and in turn, their family members who may be at risk of developing RCC. Additionally, re-evaluation of gene panels used for genetic screening of RCC to include additional risk-genes and those that may guide treatment decisions will expand the impact that genetic screening offers. However, adaptations to screening guidelines must also consider the feasibility of screening large numbers of individuals, along with the available resources. Evaluation of the selection criteria for genetic screening for RCC should integrate information regarding PV frequencies within the population, how altering specific selection criteria impacts detection rates, the added benefit that a surveillance plan may have for affected carriers, and available public funding and infrastructure.

To improve the impact that genetic screening for RCC can have on public health, a focus must also be made to improve healthcare provider knowledge about genetic testing. A survey of healthcare providers across Canada revealed limited knowledge surrounding genetic testing for renal cancer, with 47.7% of healthcare providers not knowing which screening tests are available, identifying lack of provider knowledge as the main barrier to genetic screening for Canadian RCC patients[146]. This lack of knowledge is also evident through evaluation of referral rates within Canada. When the risk criteria were applied to a large prospective database of RCC patients, only 2% of patients meeting criteria actually had documentation of genetic testing[144]. The low proportion of patients referred for screening, in combination with the KCRNC criteria's failure to

identify majority of at-risk patients, indicates that susceptibility to RCC is greatly underestimated in Canada.

Improved screening for germline genetic susceptibility among RCC patients, and their at-risk family members can also "open the door" for improved preventative strategies. Screening programs for PV carriers may increase early detection of RCC, and improve survival outcomes, however there are currently no screening programs in Canada. Though improvements to the sensitivity of ctDNA detection methods are still necessary, searching for ctDNA in the blood of those at risk of developing RCC could be a potential minimally invasive screening strategy.

CHAPTER 6. CONCLUSIONS AND FUTURE DIRECTIONS

The broad aim of this thesis was to characterize genomic patterns underlying RCC and their translational relevance. Chapter 2 presents an NGS assay tailored to the analysis of tissue and liquid biopsies for RCC, providing a sensitive tool for investigating genomic patterns. Chapter 3 applies this assay to the largest-to-date cohort of ccRCCs to interrogate their genomic status, and defines a genomic classifier capable of predicting risk of disease recurrence. Notably, this classifier aligns with evolutionary subtypes of ccRCCs, and offers potential for further refinement to include additional genomic features. Lastly, Chapter 4 identifies risk genes for RCC within Canada, summarizes differences in global risk genes, and highlights the inadequate identification of at-risk individuals through current clinical guidelines for genetic screening.

Future studies should prioritize characterization of ctDNA properties in RCC, including fragmentation features that may offer solutions for increasing the sensitivity of liquid biopsy analysis for low stage tumors. Physical properties of ctDNA may also offer translational impact for quantifying MRD[142]. It is likely that an approach combining information from various liquid biopsy analytes (CTCs, ctDNA, exosomes) will offer an improved translational impact. Longitudinal studies to evaluate tumor dynamics post-surgery through ctDNA analysis will also be valuable for assessing the potential of ctDNA as a prognostic marker for RCC.

The genomic classifier is representative of some of the evolutionary trajectories of ccRCC tumors, however the integration of additional features can help to refine the classifier even further. Inclusion of additional information including CNAs, and additional genes from the PI3K/ATM/MTOR pathway can help to isolate *BAP1*- and *PBRM1*-driven evolutionary subtypes from the multiple clonal drivers. Investigation within even larger cohorts may also allow for 'weighting' of specific genes, such as *BAP1*/*PBRM1* co-occurrence within the classifier. Lastly, investigation of genomic classification in the context of response to treatment would be valuable

to assess if we can predict response to adjuvant therapy based on tumor evolutionary features. Next steps would also include assessing the feasibility for identifying genomic classifiers within liquid biopsies.

Ultimately, this study addresses key caveats that had slowed down progress in genome-based personalized medicine in RCC, including the lack of appropriate targeted NGS panels and large matched genomic/phenomic datasets, providing power for association analysis.

CHAPTER 7. BIBLIOGRAPHY

1. Turajlic S, Xu H, Litchfield K, et al. Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell* 2018; **173**(3): 581-94.e12.
2. Riazalhosseini Y, Lathrop M. Precision medicine from the renal cancer genome. *Nat Rev Nephrol* 2016; **12**(11): 655-66.
3. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021; **71**(3): 209-49.
4. Znaor A, Lortet-Tieulent J, Laversanne M, Jemal A, Bray F. International variations and trends in renal cell carcinoma incidence and mortality. *Eur Urol* 2015; **67**(3): 519-30.
5. Chow WH, Dong LM, Devesa SS. Epidemiology and risk factors for kidney cancer. *Nat Rev Urol* 2010; **7**(5): 245-57.
6. Scelo G, Larose TL. Epidemiology and Risk Factors for Kidney Cancer. *J Clin Oncol* 2018; **36**(36): Jco2018791905.
7. Bukavina L, Bensalah K, Bray F, et al. Epidemiology of Renal Cell Carcinoma: 2022 Update. *Eur Urol* 2022; **82**(5): 529-42.
8. Theis RP, Dolwick Grieb SM, Burr D, Siddiqui T, Asal NR. Smoking, environmental tobacco smoke, and risk of renal cell cancer: a population-based case-control study. *BMC Cancer* 2008; **8**: 387.
9. Gati A, Kouidhi S, Marrakchi R, et al. Obesity and renal cancer: Role of adipokines in the tumor-immune system conflict. *Oncoimmunology* 2014; **3**(1): e27810.
10. Weikert S, Boeing H, Pischon T, et al. Blood pressure and risk of renal cell carcinoma in the European prospective investigation into cancer and nutrition. *Am J Epidemiol* 2008; **167**(4): 438-46.
11. Deckers IA, van den Brandt PA, van Engeland M, et al. Polymorphisms in genes of the renin-angiotensin-aldosterone system and renal cell cancer risk: interplay with hypertension and intakes of sodium, potassium and fluid. *Int J Cancer* 2015; **136**(5): 1104-16.
12. Cumberbatch MG, Rota M, Catto JW, La Vecchia C. The Role of Tobacco Smoke in Bladder and Kidney Carcinogenesis: A Comparison of Exposures and Meta-analysis of Incidence and Mortality Risks. *Eur Urol* 2016; **70**(3): 458-66.

13. Wang F, Xu Y. Body mass index and risk of renal cell cancer: a dose-response meta-analysis of published cohort studies. *Int J Cancer* 2014; **135**(7): 1673-86.
14. Colt JS, Schwartz K, Graubard BI, et al. Hypertension and risk of renal cell carcinoma among white and black Americans. *Epidemiology* 2011; **22**(6): 797-804.
15. Hofmann JN, Corley DA, Zhao WK, et al. Chronic kidney disease and risk of renal cell carcinoma: differences by race. *Epidemiology* 2015; **26**(1): 59-67.
16. Woldrich JM, Mallin K, Ritchey J, Carroll PR, Kane CJ. Sex differences in renal cell cancer presentation and survival: an analysis of the National Cancer Database, 1993-2004. *J Urol* 2008; **179**(5): 1709-13; discussion 13.
17. Levi F, Ferlay J, Galeone C, et al. The changing pattern of kidney cancer incidence and mortality in Europe. *BJU Int* 2008; **101**(8): 949-58.
18. Karami S, Lan Q, Rothman N, et al. Occupational trichloroethylene exposure and kidney cancer risk: a meta-analysis. *Occup Environ Med* 2012; **69**(12): 858-67.
19. Debelle FD, Vanherweghem JL, Nortier JL. Aristolochic acid nephropathy: a worldwide problem. *Kidney Int* 2008; **74**(2): 158-69.
20. Ricketts CJ, De Cubas AA, Fan H, et al. The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep* 2018; **23**(1): 313-26.e5.
21. Muglia VF, Prando A. Renal cell carcinoma: histological classification and correlation with imaging findings. *Radiol Bras* 2015; **48**(3): 166-74.
22. Leibovich BC, Lohse CM, Crispen PL, et al. Histological subtype is an independent predictor of outcome for patients with renal cell carcinoma. *J Urol* 2010; **183**(4): 1309-15.
23. Gossage L, Eisen T, Maher ER. VHL, the story of a tumour suppressor gene. *Nat Rev Cancer* 2015; **15**(1): 55-64.
24. Lobo J, Ohashi R, Amin MB, et al. WHO 2022 landscape of papillary and chromophobe renal cell carcinoma. *Histopathology* 2022; **81**(4): 426-38.
25. Linehan WM, Spellman PT, Ricketts CJ, et al. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N Engl J Med* 2016; **374**(2): 135-45.
26. Shuch B, Amin A, Armstrong AJ, et al. Understanding pathologic variants of renal cell carcinoma: distilling therapeutic opportunities from biologic complexity. *Eur Urol* 2015; **67**(1): 85-97.

27. Cheng L, Zhang S, MacLennan GT, Lopez-Beltran A, Montironi R. Molecular and cytogenetic insights into the pathogenesis, classification, differential diagnosis, and prognosis of renal epithelial neoplasms. *Hum Pathol* 2009; **40**(1): 10-29.
28. Davis CF, Ricketts CJ, Wang M, et al. The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* 2014; **26**(3): 319-30.
29. Moch H, Amin MB, Berney DM, et al. The 2022 World Health Organization Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours. *Eur Urol* 2022; **82**(5): 458-68.
30. Sirohi D, Smith SC, Agarwal N, Maughan BL. Unclassified renal cell carcinoma: diagnostic difficulties and treatment modalities. *Res Rep Urol* 2018; **10**: 205-17.
31. Haas NB, Nathanson KL. Hereditary kidney cancer syndromes. *Adv Chronic Kidney Dis* 2014; **21**(1): 81-90.
32. Reaume MN, Graham GE, Tomiak E, et al. Canadian guideline on genetic screening for hereditary renal cell cancers. *Can Urol Assoc J* 2013; **7**(9-10): 319-23.
33. Hampel H, Bennett RL, Buchanan A, Pearlman R, Wiesner GL. A practice guideline from the American College of Medical Genetics and Genomics and the National Society of Genetic Counselors: referral indications for cancer predisposition assessment. *Genet Med* 2015; **17**(1): 70-87.
34. Richards FM, Crossey PA, Phipps ME, et al. Detailed mapping of germline deletions of the von Hippel-Lindau disease tumour suppressor gene. *Hum Mol Genet* 1994; **3**(4): 595-8.
35. Ong KR, Woodward ER, Killick P, Lim C, Macdonald F, Maher ER. Genotype-phenotype correlations in von Hippel-Lindau disease. *Hum Mutat* 2007; **28**(2): 143-9.
36. Zbar B, Tory K, Merino M, et al. Hereditary papillary renal cell carcinoma. *J Urol* 1994; **151**(3): 561-6.
37. Dharmawardana PG, Giubellino A, Bottaro DP. Hereditary papillary renal carcinoma type I. *Curr Mol Med* 2004; **4**(8): 855-68.
38. Toro JR, Nickerson ML, Wei MH, et al. Mutations in the fumarate hydratase gene cause hereditary leiomyomatosis and renal cell cancer in families in North America. *Am J Hum Genet* 2003; **73**(1): 95-106.

39. Merino MJ, Torres-Cabala C, Pinto P, Linehan WM. The morphologic spectrum of kidney tumors in hereditary leiomyomatosis and renal cell carcinoma (HLRCC) syndrome. *Am J Surg Pathol* 2007; **31**(10): 1578-85.
40. Khoo SK, Bradley M, Wong FK, Hedblad MA, Nordenskjöld M, Teh BT. Birt-Hogg-Dubé syndrome: mapping of a novel hereditary neoplasia gene to chromosome 17p12-q11.2. *Oncogene* 2001; **20**(37): 5239-42.
41. Benhammou JN, Vocke CD, Santani A, et al. Identification of intragenic deletions and duplication in the FLCN gene in Birt-Hogg-Dubé syndrome. *Genes Chromosomes Cancer* 2011; **50**(6): 466-77.
42. Patard JJ. Incidental renal tumours. *Curr Opin Urol* 2009; **19**(5): 454-8.
43. Gray RE, Harris GT. Renal Cell Carcinoma: Diagnosis and Management. *Am Fam Physician* 2019; **99**(3): 179-84.
44. Naito S, Kato T, Tsuchiya N. Surgical and focal treatment for metastatic renal cell carcinoma: A literature review. *Int J Urol* 2022; **29**(6): 494-501.
45. Barata PC, Rini BI. Treatment of renal cell carcinoma: Current status and future directions. *CA Cancer J Clin* 2017; **67**(6): 507-24.
46. Oliver RT, Nethersell AB, Bottomley JM. Unexplained spontaneous regression and alpha-interferon as treatment for metastatic renal carcinoma. *Br J Urol* 1989; **63**(2): 128-31.
47. Klapper JA, Downey SG, Smith FO, et al. High-dose interleukin-2 for the treatment of metastatic renal cell carcinoma : a retrospective analysis of response and survival in patients treated in the surgery branch at the National Cancer Institute between 1986 and 2006. *Cancer* 2008; **113**(2): 293-301.
48. Dizman N, Philip EJ, Pal SK. Genomic profiling in renal cell carcinoma. *Nat Rev Nephrol* 2020; **16**(8): 435-51.
49. Escudier B, Eisen T, Stadler WM, et al. Sorafenib in advanced clear-cell renal-cell carcinoma. *N Engl J Med* 2007; **356**(2): 125-34.
50. Zhou L, Liu XD, Sun M, et al. Targeting MET and AXL overcomes resistance to sunitinib therapy in renal cell carcinoma. *Oncogene* 2016; **35**(21): 2687-97.
51. Choueiri TK, Kaelin WG, Jr. Targeting the HIF2-VEGF axis in renal cell carcinoma. *Nat Med* 2020; **26**(10): 1519-30.

52. Lavacchi D, Pellegrini E, Palmieri VE, et al. Immune Checkpoint Inhibitors in the Treatment of Renal Cancer: Current State and Future Perspective. *Int J Mol Sci* 2020; **21**(13).
53. Wei SC, Duffy CR, Allison JP. Fundamental Mechanisms of Immune Checkpoint Blockade Therapy. *Cancer Discov* 2018; **8**(9): 1069-86.
54. Motzer RJ, Escudier B, McDermott DF, et al. Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med* 2015; **373**(19): 1803-13.
55. Choueiri TK, Motzer RJ. Systemic Therapy for Metastatic Renal-Cell Carcinoma. *N Engl J Med* 2017; **376**(4): 354-66.
56. Zisman A, Pantuck AJ, Wieder J, et al. Risk group assessment and clinical outcome algorithm to predict the natural history of patients with surgically resected renal cell carcinoma. *J Clin Oncol* 2002; **20**(23): 4559-66.
57. Sorbellini M, Kattan MW, Snyder ME, et al. A postoperative prognostic nomogram predicting recurrence for patients with conventional clear cell renal cell carcinoma. *J Urol* 2005; **173**(1): 48-51.
58. Kattan MW, Reuter V, Motzer RJ, Katz J, Russo P. A postoperative prognostic nomogram for renal cell carcinoma. *J Urol* 2001; **166**(1): 63-7.
59. Yaycioglu O, Roberts WW, Chan T, Epstein JI, Marshall FF, Kavoussi LR. Prognostic assessment of nonmetastatic renal cell carcinoma: a clinically based model. *Urology* 2001; **58**(2): 141-5.
60. Karakiewicz PI, Briganti A, Chun FK, et al. Multi-institutional validation of a new renal cancer-specific survival nomogram. *J Clin Oncol* 2007; **25**(11): 1316-22.
61. Frank I, Blute ML, Cheville JC, Lohse CM, Weaver AL, Zincke H. 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* 2002; **168**(6): 2395-400.
62. Leibovich BC, Blute ML, Cheville JC, et al. Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. *Cancer* 2003; **97**(7): 1663-71.

63. Correa AF, Jegede O, Haas NB, et al. Predicting Renal Cancer Recurrence: Defining Limitations of Existing Prognostic Models With Prospective Trial-Based Validation. *J Clin Oncol* 2019; **37**(23): 2062-71.
64. Lam JS, Shvarts O, Leppert JT, Pantuck AJ, Figlin RA, Belldegrun AS. 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* 2005; **174**(2): 466-72; discussion 72; quiz 801.
65. Zbar B, Brauch H, Talmadge C, Linehan M. Loss of alleles of loci on the short arm of chromosome 3 in renal cell carcinoma. *Nature* 1987; **327**(6124): 721-4.
66. Shen C, Beroukhim R, Schumacher SE, et al. Genetic and functional studies implicate HIF1α as a 14q kidney cancer suppressor gene. *Cancer Discov* 2011; **1**(3): 222-35.
67. Beroukhim R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. *Nature* 2010; **463**(7283): 899-905.
68. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 2013; **499**(7456): 43-9.
69. Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* 2011; **469**(7331): 539-42.
70. Nargund AM, Pham CG, Dong Y, et al. The SWI/SNF Protein PBRM1 Restrains VHL-Loss-Driven Clear Cell Renal Cell Carcinoma. *Cell Rep* 2017; **18**(12): 2893-906.
71. Gossage L, Murtaza M, Slatter AF, et al. Clinical and pathological impact of VHL, PBRM1, BAP1, SETD2, KDM6A, and JARID1c in clear cell renal cell carcinoma. *Genes Chromosomes Cancer* 2014; **53**(1): 38-51.
72. Voss MH, Reising A, Cheng Y, et al. Genomically annotated risk model for advanced renal-cell carcinoma: a retrospective cohort study. *The Lancet Oncology* 2018; **19**(12): 1688-98.
73. Guo G, Gui Y, Gao S, et al. Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. *Nat Genet* 2011; **44**(1): 17-9.
74. Hsieh JJ, Chen D, Wang PI, et al. Genomic Biomarkers of a Randomized Trial Comparing First-line Everolimus and Sunitinib in Patients with Metastatic Renal Cell Carcinoma. *Eur Urol* 2017; **71**(3): 405-14.

75. Ho TH, Choueiri TK, Wang K, et al. Correlation Between Molecular Subclassifications of Clear Cell Renal Cell Carcinoma and Targeted Therapy Response. *Eur Urol Focus* 2016; **2**(2): 204-9.
76. Sato Y, Yoshizato T, Shiraishi Y, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet* 2013; **45**(8): 860-7.
77. Yang P, Cornejo KM, Sadow PM, et al. Renal cell carcinoma in tuberous sclerosis complex. *Am J Surg Pathol* 2014; **38**(7): 895-909.
78. Shuch B, Ricketts CJ, Vocke CD, et al. Germline PTEN mutation Cowden syndrome: an underappreciated form of hereditary kidney cancer. *J Urol* 2013; **190**(6): 1990-8.
79. Menko FH, Maher ER, Schmidt LS, et al. Hereditary leiomyomatosis and renal cell cancer (HLRCC): renal cancer risk, surveillance and treatment. *Fam Cancer* 2014; **13**(4): 637-44.
80. Linehan WM. Genetic basis of kidney cancer: role of genomics for the development of disease-based therapeutics. *Genome Res* 2012; **22**(11): 2089-100.
81. Simonaggio A, Ambrosetti D, Verkarre V, Auvray M, Oudard S, Vano YA. MiTF/TFE Translocation Renal Cell Carcinomas: From Clinical Entities to Molecular Insights. *Int J Mol Sci* 2022; **23**(14).
82. Caliò A, Segala D, Munari E, Brunelli M, Martignoni G. MiT Family Translocation Renal Cell Carcinoma: from the Early Descriptions to the Current Knowledge. *Cancers (Basel)* 2019; **11**(8).
83. Elias R, Zhang Q, Brugarolas J. The von Hippel-Lindau Tumor Suppressor Gene: Implications and Therapeutic Opportunities. *Cancer J* 2020; **26**(5): 390-8.
84. Latif F, Tory K, Gnarra J, et al. Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science* 1993; **260**(5112): 1317-20.
85. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 1971; **68**(4): 820-3.
86. Maxwell PH, Wiesener MS, Chang GW, et al. The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* 1999; **399**(6733): 271-5.
87. Xia X, Lemieux ME, Li W, et al. Integrative analysis of HIF binding and transactivation reveals its role in maintaining histone methylation homeostasis. *Proc Natl Acad Sci U S A* 2009; **106**(11): 4260-5.

88. Hakimi AA, Tickoo SK, Jacobsen A, et al. TCEB1-mutated renal cell carcinoma: a distinct genomic and morphological subtype. *Mod Pathol* 2015; **28**(6): 845-53.
89. Hodges C, Kirkland JG, Crabtree GR. The Many Roles of BAF (mSWI/SNF) and PBAF Complexes in Cancer. *Cold Spring Harb Perspect Med* 2016; **6**(8).
90. Xie Y, Sahin M, Sinha S, et al. SETD2 loss perturbs the kidney cancer epigenetic landscape to promote metastasis and engenders actionable dependencies on histone chaperone complexes. *Nat Cancer* 2022; **3**(2): 188-202.
91. Farley MN, Schmidt LS, Mester JL, et al. A novel germline mutation in BAP1 predisposes to familial clear-cell renal cell carcinoma. *Mol Cancer Res* 2013; **11**(9): 1061-71.
92. Joseph RW, Kapur P, Serie DJ, et al. Loss of BAP1 protein expression is an independent marker of poor prognosis in patients with low-risk clear cell renal cell carcinoma. *Cancer* 2014; **120**(7): 1059-67.
93. Hakimi AA, Ostrovnaya I, Reva B, et al. Adverse outcomes in clear cell renal cell carcinoma with mutations of 3p21 epigenetic regulators BAP1 and SETD2: a report by MSKCC and the KIRC TCGA research network. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2013; **19**(12): 3259-67.
94. Kwon J, Lee D, Lee SA. BAP1 as a guardian of genome stability: implications in human cancer. *Exp Mol Med* 2023; **55**(4): 745-54.
95. Peña-Llopis S, Christie A, Xie XJ, Brugarolas J. Cooperation and antagonism among cancer genes: the renal cancer paradigm. *Cancer Res* 2013; **73**(14): 4173-9.
96. Peña-Llopis S, Vega-Rubín-de-Celis S, Liao A, et al. BAP1 loss defines a new class of renal cell carcinoma. *Nat Genet* 2012; **44**(7): 751-9.
97. Kapur P, Peña-Llopis S, Christie A, et al. Effects on survival of BAP1 and PBRM1 mutations in sporadic clear-cell renal-cell carcinoma: a retrospective analysis with independent validation. *The Lancet Oncology* 2013; **14**(2): 159-67.
98. Motzer RJ, Escudier B, Oudard S, et al. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet* 2008; **372**(9637): 449-56.
99. Kucejova B, Peña-Llopis S, Yamasaki T, et al. Interplay between pVHL and mTORC1 pathways in clear-cell renal cell carcinoma. *Mol Cancer Res* 2011; **9**(9): 1255-65.

100. Brugarolas J. Molecular genetics of clear-cell renal cell carcinoma. *J Clin Oncol* 2014; **32**(18): 1968-76.
101. Stephens PJ, Greenman CD, Fu B, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 2011; **144**(1): 27-40.
102. Mitchell TJ, Turajlic S, Rowan A, et al. Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. *Cell* 2018; **173**(3): 611-23.e17.
103. Gerlinger M, Horswell S, Larkin J, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet* 2014; **46**(3): 225-33.
104. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012; **366**(10): 883-92.
105. Turajlic S, Xu H, Litchfield K, et al. Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell* 2018; **173**(3): 595-610.e11.
106. Bettegowda C, Sausen M, Leary RJ, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* 2014; **6**(224): 224ra24.
107. Parikh AR, Leshchiner I, Elagina L, et al. Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat Med* 2019; **25**(9): 1415-21.
108. Zieren RC, Zondervan PJ, Pienta KJ, Bex A, de Reijke TM, Bins AD. Diagnostic liquid biopsy biomarkers in renal cell cancer. *Nat Rev Urol* 2023.
109. Yamamoto Y, Uemura M, Fujita M, et al. Clinical significance of the mutational landscape and fragmentation of circulating tumor DNA in renal cell carcinoma. *Cancer Sci* 2019; **110**(2): 617-28.
110. Hauser S, Zahalka T, Ellinger J, et al. Cell-free circulating DNA: Diagnostic value in patients with renal cell cancer. *Anticancer Res* 2010; **30**(7): 2785-9.
111. van der Toom EE, Verdone JE, Gorin MA, Pienta KJ. Technical challenges in the isolation and analysis of circulating tumor cells. *Oncotarget* 2016; **7**(38): 62754-66.
112. Nel I, Gauler TC, Bublitz K, et al. Circulating Tumor Cell Composition in Renal Cell Carcinoma. *PLoS One* 2016; **11**(4): e0153018.
113. Liu S, Tian Z, Zhang L, et al. Combined cell surface carbonic anhydrase 9 and CD147 antigens enable high-efficiency capture of circulating tumor cells in clear cell renal cell carcinoma patients. *Oncotarget* 2016; **7**(37): 59877-91.

114. Grange C, Tapparo M, Collino F, et al. Microvesicles released from human renal cancer stem cells stimulate angiogenesis and formation of lung premetastatic niche. *Cancer Res* 2011; **71**(15): 5346-56.

115. Dong L, Zieren RC, Horie K, et al. Comprehensive evaluation of methods for small extracellular vesicles separation from human plasma, urine and cell culture medium. *J Extracell Vesicles* 2020; **10**(2): e12044.

116. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer* 2011; **11**(6): 426-37.

117. Stroun M, Lyautey J, Lederrey C, Olson-Sand A, Anker P. About the possible origin and mechanism of circulating DNA apoptosis and active DNA release. *Clin Chim Acta* 2001; **313**(1-2): 139-42.

118. Trejo-Becerril C, Pérez-Cárdenas E, Taja-Chayeb L, et al. Cancer progression mediated by horizontal gene transfer in an in vivo model. *PLoS One* 2012; **7**(12): e52754.

119. García-Olmo DC, Domínguez C, García-Arranz M, et al. Cell-free nucleic acids circulating in the plasma of colorectal cancer patients induce the oncogenic transformation of susceptible cultured cells. *Cancer Res* 2010; **70**(2): 560-7.

120. Cimadamore A, Massari F, Santoni M, et al. Molecular characterization and diagnostic criteria of renal cell carcinoma with emphasis on liquid biopsies. *Expert Rev Mol Diagn* 2020; **20**(2): 141-50.

121. Siena S, Sartore-Bianchi A, Garcia-Carbonero R, et al. Dynamic molecular analysis and clinical correlates of tumor evolution within a phase II trial of panitumumab-based therapy in metastatic colorectal cancer. *Ann Oncol* 2018; **29**(1): 119-26.

122. Cabanero M, Tsao MS. Circulating tumour DNA in EGFR-mutant non-small-cell lung cancer. *Curr Oncol* 2018; **25**(Suppl 1): S38-s44.

123. Reinert T, Schøler LV, Thomsen R, et al. Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. *Gut* 2016; **65**(4): 625-34.

124. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol* 2013; **10**(8): 472-84.

125. Wan JCM, Massie C, Garcia-Corbacho J, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* 2017; **17**(4): 223-38.

126. Goldberg SB, Narayan A, Kole AJ, et al. Early Assessment of Lung Cancer Immunotherapy Response via Circulating Tumor DNA. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2018; **24**(8): 1872-80.
127. Phallen J, Sausen M, Adleff V, et al. Direct detection of early-stage cancers using circulating tumor DNA. *Sci Transl Med* 2017; **9**(403).
128. Gormally E, Vineis P, Matullo G, et al. TP53 and KRAS2 mutations in plasma DNA of healthy subjects and subsequent cancer occurrence: a prospective study. *Cancer Res* 2006; **66**(13): 6871-6.
129. Abbosh C, Birkbak NJ, Wilson GA, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* 2017; **545**(7655): 446-51.
130. Beaver JA, Jelovac D, Balukrishna S, et al. Detection of cancer DNA in plasma of patients with early-stage breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2014; **20**(10): 2643-50.
131. Mouliere F, Chandrananda D, Piskorz AM, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med* 2018; **10**(466).
132. Cristiano S, Leal A, Phallen J, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* 2019; **570**(7761): 385-9.
133. Smith CG, Moser T, Mouliere F, et al. Comprehensive characterization of cell-free tumor DNA in plasma and urine of patients with renal tumors. *Genome Med* 2020; **12**(1): 23.
134. Robinson DR, Wu YM, Lonigro RJ, et al. Integrative clinical genomics of metastatic cancer. *Nature* 2017; **548**(7667): 297-303.
135. Choueiri TK, Plimack E, Arkenau HT, et al. Biomarker-Based Phase II Trial of Savolitinib in Patients With Advanced Papillary Renal Cell Cancer. *J Clin Oncol* 2017; **35**(26): 2993-3001.
136. Farmer H, McCabe N, Lord CJ, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 2005; **434**(7035): 917-21.
137. Li CH, Haider S, Shiah YJ, Thai K, Boutros PC. Sex Differences in Cancer Driver Genes and Biomarkers. *Cancer Res* 2018; **78**(19): 5527-37.
138. Li CH, Prokopec SD, Sun RX, Yousif F, Schmitz N, Boutros PC. Sex differences in oncogenic mutational processes. *Nature communications* 2020; **11**(1): 4330.

139. Yuan Y, Liu L, Chen H, et al. Comprehensive Characterization of Molecular Differences in Cancer between Male and Female Patients. *Cancer Cell* 2016; **29**(5): 711-22.
140. Laskar RS, Muller DC, Li P, et al. Sex specific associations in genome wide association analysis of renal cell carcinoma. *Eur J Hum Genet* 2019; **27**(10): 1589-98.
141. Brannon AR, Haake SM, Hacker KE, et al. Meta-analysis of clear cell renal cell carcinoma gene expression defines a variant subgroup and identifies gender influences on tumor biology. *Eur Urol* 2012; **61**(2): 258-68.
142. Vessies DCL, Schuurbiers MMF, van der Noort V, et al. Combining variant detection and fragment length analysis improves detection of minimal residual disease in postsurgery circulating tumour DNA of stage II-IIIA NSCLC patients. *Mol Oncol* 2022; **16**(14): 2719-32.
143. Markus H, Chandrananda D, Moore E, et al. Refined characterization of circulating tumor DNA through biological feature integration. *Sci Rep* 2022; **12**(1): 1928.
144. Kushnir I, Kirk L, Mallick R, et al. Application of Hereditary Renal Cell Carcinoma Risk Criteria to a Large Prospective Database. *Clin Oncol (R Coll Radiol)* 2020; **32**(1): e10-e5.
145. Carlo MI, Mukherjee S, Mandelker D, et al. Prevalence of Germline Mutations in Cancer Susceptibility Genes in Patients With Advanced Renal Cell Carcinoma. *JAMA Oncol* 2018; **4**(9): 1228-35.
146. Violette PD, Kamel-Reid S, Graham GE, et al. Knowledge of genetic testing for hereditary kidney cancer in Canada is lacking: The results of the Canadian national hereditary kidney cancer needs assessment survey. *Can Urol Assoc J* 2014; **8**(11-12): E832-40.

APPENDIX

Permission from co-first authors Naveen Vasudev and Ghislaine Scelo to include the manuscript published in *Clinical Cancer Research* (Chapter 3) in the thesis.

10/3/23, 2:51 PM Mail - Kate Glennon - Outlook

Re: Permission to include CCR manuscript in thesis

Naveen Vasudev <N.Vasudev@leeds.ac.uk>
Wed 9/13/2023 9:08 AM
To:Kate Glennon <kate.glennon@mail.mcgill.ca>

Dear Kate

Re: Vasudev, Scelo, Glennon et al. Application of genomic sequencing to refine patient stratification for adjuvant therapy in renal cell carcinoma. CCR 2023

I am writing to confirm the following:

1. I agree to the use of the above publication as part of your manuscript based thesis
2. I will not use the above publication in a manuscript based thesis

Sincerely

Naveen

Dr Naveen Vasudev
Associate Professor / Honorary Consultant in Medical Oncology
Leeds Institute of Medical Research at St James's & St James's Institute of Oncology
Tel: 0113 – 2067680
Clinical Secretary: Lorraine Yeowart 0113 20-68218
Academic PA: 0113 20-67592
https://medicinehealth.leeds.ac.uk/medicine/staff/840/dr-naveen-vasudev

10/3/23, 2:49 PM Mail - Kate Glennon - Outlook

Re: Permission to include CCR manuscript in thesis

Ghislaine Scelo <ghislaine.scelo@gmail.com>
Fri 9/22/2023 8:00 AM
To:Kate Glennon <kate.glennon@mail.mcgill.ca>

Hi Kate,

Congratulations on finishing you PhD thesis and appologies for the delay. Please see below.

Re: Vasudev, Scelo, Glennon et al. Application of genomic sequencing to refine patient stratification for adjuvant therapy in renal cell carcinoma. CCR 2023
I am writing to confirm the following:
1. I agree to the use of the above publication as part of your manuscript based thesis
2. I will not use the above publication in a manuscript based thesis

Kind regards,

Ghislaine