THE MOLECULAR UNDERPINNINGS OF WILD-TYPE VON HIPPEL-LINDAU CLEAR CELL RENAL CELL CARCINOMA

Aashil A Batavia

Diss. ETH N^{o} 28525

Diss. ETH Nº 28525

THE MOLECULAR UNDERPINNINGS OF WILD-TYPE VON HIPPEL-LINDAU CLEAR CELL RENAL CELL CARCINOMA

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES
(Dr. sc. ETH Zurich)

Presented by

AASHIL A BATAVIA

M.Sc Bioinformatics and Systems Biology

The University of Manchester

Born on 30.01.1993

Accepted on the recommendation of

Prof. Dr. Niko Beerenwinkel
Prof. Dr. Mitch Levesque
Prof. Dr. Holger Moch

2022

Abstract

The following work was conducted within the Department of Pathology and Molecular Pathology at the University Hospital Zurich (USZ) and The Department of Biosystems Science and Engineering at Eidgenössische Technische Hochschule Zürich. In this dissertation, I assess the clinical and molecular phenotypes of wild-type *von Hippel-Lindau* renal cell carcinoma with clear cell histology (^{wt}*VHL* ccRCC). *VHL* aberrations are central for tumourigenesis in ccRCC. The identification but frequent dismissal of ^{wt}*VHL* ccRCC in scientific literature has led to a profound lack of understanding as to the tumorigenesis and molecular phenotypes of these carcinoma. Unveiling the molecular and clinical characteristics of these rare malignancies is the goal of my work.

In Chapter **I** the context from which my works originates is set. I introduce the varied landscape of renal cell carcinoma with clear cell histology and the importance of *VHL* aberrations as evolutionarily truncal events within these tumours, initiating tumourigenesis. I describe the role of other factors e.g. the SWI/SNF gene family and pathways e.g PI3k–AKT–mTOR which, in combination with *VHL* aberrations, promote tumour progression in ccRCC. I also further introduce a new tumour entity now termed ELOC RCC (previously TCEB1 RCC) characterised by the biallelic inactivation of *ELOC* which lies within the same E3 ubiquitin ligase complex as protein VHL (pVHL). In Chapter **II** I present my work identifying rare ELOC RCC with clear cell histology from a patient cohort derived from the USZ Tissue Biobank. Genomic assessment of FFPE derived tumours allowed the identification of those with Chromosome 8q deletions and *ELOC* mutations resulting in its biallelic inactivation. I identify a novel *ELOC* variant which I describe using *in silico* structure modelling and structural docking analysis. Using mass spectrometry I show biallelic inactivations of *ELOC* result in the reduced expression of Elongin-C and may cause tumourigenesis via similar pathways as the biallelic inactivation of *VHL*. Chapter **III** contains the clinical and molecular characterisation of the remaining ^{wt}*VHL* ccRCC which do not contain biallelic *ELOC* inactivations. I obtain, assess and integrate genomic, epigenomic, transcriptomic, proteomic and clinical data from both my own cohort and publicly available sources to assess the molecular landscape of ^{wt}*VHL* ccRCC in comparison to classical ccRCC. I find ^{wt}*VHL* ccRCC tumours confer a poorer prognosis than those with *VHL* aberrations caused by an increase in invasive properties. Specifically I find, across multiple omic layers, factors promoting cell division, EMT and migration are elevated in ^{wt}*VHL* ccRCC in comparison to classical ccRCC. Chapter **IV** forms the conclusion of this work, summarising the findings and limitations as well as providing an outlook for future research.

Kurzfassung

Die folgende Arbeit wurde am Institut für Pathologie und Molekularpathologie des UniversitätsSpitals Zürich und des Departments für Biosystems Science and Engineering an der Eidgenössischen Technischen Hochschule ausgeführt. In dieser Doktorarbeit gehe ich auf die klinischen und molekularen Phänotypen von Nierenzellkarzinomen (NZK) mit intaktem (Wildtyp) von Hippel-Lindau Gen und klarzelliger Histologie ein (^{wt}*VHL* klarzellige NZK bzw. ^{wt}*VHL* kNZK). VHL Genveränderungen sind der Schlüssel für die Entstehung von kNZK. Die Identifizierung, aber häufige Vernachlässigung, von ^{wt}*VHL* kNZK in der wissenschaftlichen Literatur führte zu einer starken Wissenslücke hinsichtlich Entstehung und molekularen Phänotyps dieser Karzinome. Das Ziel meiner Arbeit ist die Entschlüsselung der molekularen und klinischen Charakteristika dieser seltenen Tumoren.

Kapitel **I** steckt den Rahmen meiner Arbeit ab. Ich beginne mit den unterschiedlichen Formen des Nierenzellkarzinoms mit klarzelliger Histologie und der Bedeutung von *VHL* Änderungen als evolutionärem Hauptereignis für die Entstehung dieser Tumore. Ich beschreibe die Rolle anderer Faktoren, wie zum Beispiel die der SWI/SNF Genfamilie und von Signalpfaden, wie dem von PI3K-AKT-mTOR, die zusammen mit VHL Mutationen die Tumorprogression in kNZK fördern. Ich stelle auch eine neue Tumorentität vor, ELOC NZK (vormals TCEB1 NZK), die durch biallelische Inaktivierung von *ELOC* charakterisiert ist und die sich im selben E3 Ubiquitin Ligase Komplex befindet wie das VHL Protein. In Kapitel **II** führe ich aus wie ich mit Hilfe einer aus der USZ Nierentumor-Biobank stammenden Patientenkohorte seltene ELOC NZK mit klarzelliger Histologie identifiziere. Die genomische Analyse Formalin fixierter, in Paraffin eingebetteter Tumoren erlaubten die Identifizierung solcher mit Chromosom 8q Deletionen und *ELOC* Mutationen, was zu einer biallelischen Inaktivierung führt. Ich identifiziere eine neue *ELOC* Mutationsvariante, welche ich mit Hilfe von in silico Strukturmodellierung und struktureller «Docking» Analyse beschreibe. Durch Massenspektrometrie zeigen wir, dass eine biallelische Inaktivierung von *ELOC* zu reduzierter Expression von Elongin-C führt. Dies könnte über ähnliche Signalwege wie bei einer biallelischen *VHL* Inaktivierung zu einer Tumorentstehung führen. Kapitel **III** beinhaltet die klinische und molekulare Charakterisierung der restlichen ^{wt}*VHL* kNZK welche keine biallelische *ELOC* Inaktivierung aufweisen. Ich erhalte, bearbeite und integriere genomische, epigenomische, transkriptomische, proteomische und klinische Daten sowohl von meiner eigenen Kohorte als auch von öffentlich verfügbaren Quellen, um die molekularen Eigenheiten von ^{wt}*VHL* kNZK mit denen der klassischen kNZK zu vergleichen. Ich stellte fest, dass ^{wt}*VHL* kNZK eine schlechtere Prognose aufweisen als Tumoren mit VHL Änderungen, was durch eine Zunahme invasiver Merkmale verursacht wird. Insbesondere zeigte sich, dass über mehrere omic layer hinweg, Faktoren die die Zellteilung, EMT und Migration fördern, bei ^{wt}*VHL* ccRCC im Vergleich zu klassischem ccRCC erhöht sind. Kapitel **IV** ist der Schlussfolgerung aus dieser Arbeit gewidmet, in welchem Resultate und Limitationen zusammengefasst, so wie ein Ausblick auf die zukünftige Forschung gegeben werden.

Contents

I

Introduction

The following review article published in *Histopathology* laid the foundation on which my work lies. Tumour classifications, as endorsed by the World Health Organisation, are currently vastly driven by histological features, however, in the modern era there is a transition toward a classification system incorporating morphological, immunohistochemical and molecular features[1]. This is reflected in the 2022 WHO classification of urogenital tumours[2] summarised toward the latter end of this chapter. In the subsequent article, I outline the complex and heterogeneous nature of clear cell renal cell carcinoma with a focus on tumours with no aberrations in the *von Hippel-Lindau* gene. My contributions involve obtaining the literature, assessing its validity and writing the article.

[1] Pfister, S. M. et al. A Summary of the Inaugural WHO Classification of Pediatric Tumors: Transitioning from the Optical into the Molecular Era. Cancer Discov. 12, 331–355 (2022).
[2] International Agency for Research on Cancer. WHO Classification of Tumours ; Urinary and male genital tumours. 5th ed. (2022).

Clear cell renal cell carcinoma with wild-type von Hippel-Lindau gene: a non-existent or new tumour entity?

Aashil A Batavia, Peter Schraml & Holger Moch

Department of Pathology and Molecular Pathology, University and University Hospital Zurich, Zurich, Switzerland

Keywords: clear cell renal cell carcinoma, TCEB1-deficient ccRCC, VHL wild-type

Abstract

The current World Health Organisation (WHO) classification of renal tumours is based on characteristic histological features or specific molecular alterations. *von Hippel-Lindau* (*VHL*) alteration is the hallmark of clear cell renal cell carcinoma (ccRCC). After identification of the *MiT* translocation family of tumours, clear cell papillary renal cancer and others, the group of ccRCC with wild-type *VHL* is small. *TCEB1* mutation combined with chromosome 8q loss is an emerging tumour entity with wild-type *VHL*. Inactivation of *TCEB1* increases HIF stabilisation via the same mechanism as *VHL* inactivation. Importantly, recent molecular analyses suggest the existence of another '*VHL* wild-type' evolutionary subtype of clear cell RCC in addition to *TCEB1* mutated RCC and clear cell papillary renal cancer. These tumours are characterised by aggressive behaviour, high tumour cell proliferation rate, elevated chromosomal instability and frequent presence of sarcomatoid differentiation. Future clinicopathological studies will have to provide data to determine whether *TCEB1* tumours and clear cell RCC with wild-type *VHL* are separate tumour entities or represent variants of a clear cell RCC tumour family.

Introduction

Tumour classification is important for outcome prediction and treatment decisions. In the past, tumour classification was based mainly on morphological parameters assessed by light microscopy. The histological classification of renal tumours has developed over time and refers to various characteristic features. Different renal tumour entities are named according to predominant cytoplasmic features and staining characteristics (clear cell and chromophobe RCC). The clear cytoplasm of clear cell renal cell carcinoma (ccRCC) is due to the accumulation of glycogen and lipids that are dissolved by deparaffinising formalin-fixed paraffin-embedded (FFPE) sections using organic solvents.[1] Architectural features (papillary RCC), a combination of such features (clear cell papillary RCC), the anatomical location of the tumour (collecting duct RCC) or a correlation with background renal disease (acquired cystic disease associated RCC) are also used for tumour classification. In recent years there has been a trend to refer to molecular alterations that are pathognomonic for RCC subtypes (e.g. MiT family translocation renal carcinomas, SDH-deficient renal cancer or HLRCC-associated RCC with specific *FH* mutations). Emerging renal cancer entities include

ALK-rearrangement-associated RCC. Molecular findings coming from large sequencing efforts and novel genotype/phenotype correlations will have a dramatic impact on future classification systems.

The relevance of *von Hippel-Lindau* gene (*VHL*) alterations for sporadic ccRCC was identified following the study of families with VHL syndrome. The prevalence of *VHL* alterations in sporadic ccRCC has been reported to range from 56% to 91% (Table 1). Some novel renal tumour entities of the 2016 WHO classification are also characterised by clear cytoplasm and show significant morphological overlap with ccRCC.[2] Therefore, translocation carcinomas, acquired cystic disease (ACD)-associated RCC and clear cell papillary RCC have certainly influenced the prevalence of *VHL* alterations in older series, as they would have been diagnosed as ccRCC. *TCEB1* renal cancer is an emerging tumour entity. These tumours also have a characteristic clear cytoplasm. The aim of this review is to discuss the consequence of these new molecular findings for future renal tumour classifications.

Table 1: ccRCC studies with analyses addressing chromosome 3p loss, VHL mutation and VHL promoter methylation

Patients (N)	3p loss (%)	VHL mutated (%)	VHL methylated (%)	Bi-allelic inactivation (%)	wtVHL (1)* (%)	wtVHL (2)* (%)	Reference
96	78.4	69.9	20.4	74.1	14.3	11.5	Banks et al.[52]
385	91	52.3	7	56.1	38.7	5.1	The Cancer Genome Atlas Research[8]
240	94.2	85.4	10	91.7	2.9	5.4	Sato et al.[2]
50	80	70	12	80	2	18	Beroukhim et al.[53]
124†	93†	38.4	6.6	41.1	51.6	7.3	Brauch et al.[54]
155†	90†	56.8	5.1	54.8	35.5	9.7	Kondo et al.[55]

* Mono-allelic (1) or bi-allelic (2) wild-type VHL
† Informative cases only

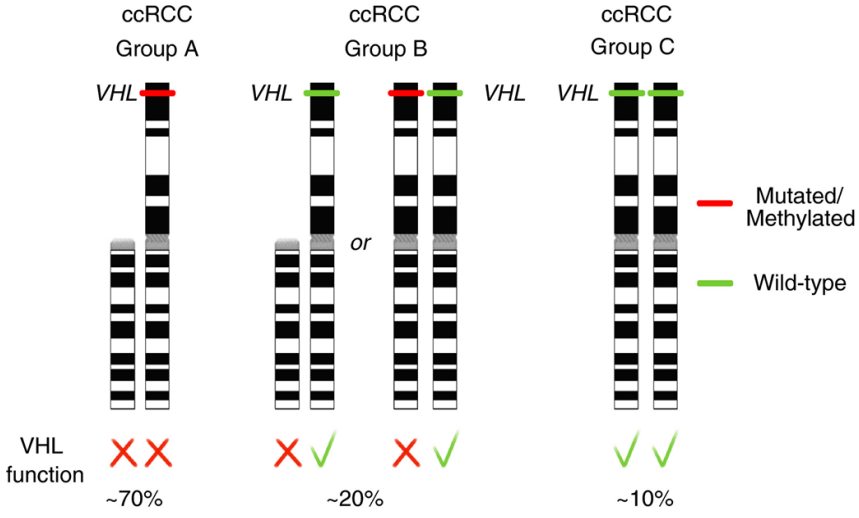

Figure 1: Three groups of clear cell renal carcinoma (ccRCC) and their von Hippel-Lindau (VHL) status: VHL bi-allelic inactivated, mono-allelic and bi-allelic VHL wild-type tumours.

The VHL gene in ccRCC

Among the several histological subtypes of renal cell carcinomas, clear cell renal carcinoma remain the most prevalent, making up 70% of all renal cancers.[2,3] For decades, it has been well accepted that the majority of ccRCC are characterised by bi-allelic loss of the *VHL* tumour suppressor gene located on chromosome 3p25-26.[4,5] However, *VHL* alterations are absent in other renal tumour subtypes, e.g. papillary RCC, clear cell papillary RCC, translocation carcinomas and chromophobe carcinomas. Patients suffering from the autosomal-dominant von Hippel– Lindau syndrome, an inherited familial disorder, receive a single inactivated copy of the *VHL* gene. Due to the haplosufficient nature of *VHL*, those suffering from VHL syndrome demonstrate a one-hit model for tumorigenesis in which a single inactivation step is required for a tumour to form.[6,7] In sporadic ccRCC, *VHL* inactivation follows the Knudson two-hit model for tumorigenesis where two active copies of *VHL* are sequentially altered.[7] Bi-allelic inactivation of *VHL* occurs via mutations, copy deletion and promoter hypermethylation (see Figure 1).[5,8–10]

The *VHL* gene product, pVHL, is a multi-adapter protein that binds more than 30 different binding partners.[11] pVHL is a key component of the E3 ubiquitin ligase complex and is essential for the oxygen-dependent ubiquitin-mediated proteolytic degradation of HIF-a.[12] The loss of function of pVHL in ccRCC leads to stabilisation of the HIF-a subunits which, in turn, enhances the transcription of HIF gene targets such as *VEGF*, *PDGF*, *EPO*, *CA9* and *CXCR4*. These proteins are known to be important in metastatic processes.[13] pVHL also interacts with several effector proteins that regulate cellular processes, including microtubule stability, activation of p53, neuronal apoptosis, cellular senescence and aneuploidy, ubiquitination of RNA polymerase II and regulation of nuclear factor kappa B (NF-kB) activity.[11,13]

Clear cell RCC beyond VHL

It is becoming increasingly evident that other tumour suppressor genes alongside *VHL* on the 3p arm are also key for tumorigenesis and tumour evolution in ccRCC. The genes coding Polybromo-1 (PBRM1), BRCA1-associated protein-1 (BAP1) and Set domain containing 2 (SETD2) are all located adjacent to *VHL* on chromosome 3 and commonly lost along with VHL upon 3p arm deletion.[8] *PBRM1* encodes for BAF180, a key component for the PBAF SWI/SNF chromatin remodelling complex associated with DNA damage repair, replication and cell proliferation.[14,15] *BAP1* codes for a nuclear deubiquitinase and plays a key role in histone modification,[16] and *SETD2* encodes a histone H3K36 methyltransferase and is thought to maintain genome stability and coordinate DNA repair[17]. Several large-scale sequencing studies focusing on ccRCC have shown that, after *VHL*, these three genes are the most frequently mutated in ccRCC, making evident their importance in ccRCC development and progression.[8,14,18–20] *PBRM1* mutations are present in approximately 40% of ccRCC, whereas *BAP1* and *SETD2* are found to be mutated in approximately 10% of ccRCC.

Reduced expression of *PBRM1* has been associated previously with poorer prognosis[21]; however, recently Miao *et al.*[22] have demonstrated an increased clinical benefit of immune checkpoint therapy in patients with inactivated *PBRM1*, as defined by the Response Evaluation Criteria In Solid Tumours (RECIST) guidelines. This indicates that *PBRM1* loss induces changes resulting in an increased susceptibility to immunotherapy.[22] Univariate and multivariate analysis has shown that *BAP1*

mutations have been associated with reduced overall survival, which is not true for both *PBRM1* and *SETD2* mutations.[8,20] Moreover, *BAP1* mutations were found to be mutually exclusive with *PBRM1* mutations while *SETD2* mutations are not seen in the presence of *BAP1* mutations, but are found either alone or in combination with *PBRM1*. The fact that the most frequently mutated genes after *VHL* play roles in chromatin remodelling and histone modifications highlight the importance of epigenetic regulation in ccRCC. Notably, the SWI/SNF family is largely implicated in ccRCC not only with the frequent *PBRM1* mutations but also with less frequent *ARID1A*, *SMARCA2* and *SMARCA4* mutations, which all form components of the SWI/SNF family.[20]

In addition to these epigenetic regulators, many other gene mutations and somatic copy number aberrations (SCNA) have been associated with ccRCC. Alterations in *PTEN*, *PI3Kca* and *mTOR* are also prevalent in ccRCC, which are all key players in the PI3k–AKT–mTOR pathway regulating cell proliferation and HIF stabilisation.[8,20,23] With regard to SCNA in ccRCC, 5q gains are the second most frequent alterations following 3p deletions with the *SQSTM1* gene, thought to be the pathogenic target for this gain.[24,25] SCNA are common in ccRCC, with alterations in nearly all chromosomes being observed; however, loss of 14q and chromosome 9 have been specifically associated with a poorer prognosis in ccRCC.[26–28]

Studies have shown that VHL inactivation alone is insufficient for tumorigenesis,[29] highlighting the importance of the SCNA and secondary driver genes for formation and progression of ccRCC. Turajlic et al.[18] further demonstrate the importance of these elements for the evolution of ccRCC. Following the identification of driver event ordering and mutual exclusivity from 1206 tumour regions from 101 patients, seven evolutionary subtypes of ccRCC were described. Five of the seven evolutionary subtypes included at least one of the secondary driver mutations along with VHL. The 'PBRM1→SETD2' subtype was defined by mutations in PBRM1 followed by SETD2, the 'PBRM1→PI3k' subgroup was defined by PBRM1 mutations followed by alterations in the PI3K/AKT pathway, e.g. TSC1 and mTOR, the 'PBRM1→SCNA' subgroup was defined by mutations in PBRM1 followed by a driver SCNA event, the 'BAP1 driven' subgroup was defined by BAP1 as the single driver event with VHL, and finally, the 'multiple clonal drivers' evolutionary subtype, which was defined by tumours in which two or more secondary driver genes where clonally mutated.[18]

Clear cell RCC with wild-type VHL

If MiT-family translocation RCC and clear cell papillary RCC are classified correctly, the vast majority of ccRCC is characterised by the inactivation of *VHL*. However, there remains a subset of ccRCC cases with active *VHL* considered wild-type *VHL* (^{wt}*VHL*) ccRCC. Most ccRCC studies have focused on the frequencies of *VHL* mutation and chromosome 3p loss to study ccRCC with inactivated *VHL*. Therefore, details with regard to the molecular background of ^{wt}*VHL* ccRCC cases are limited. Several groups have stratified their data by *VHL* mutation types and found that missense mutations may exert different effects on both pVHL stability and binding capability.[11,30,31]

All three methods of inactivation – mutation, 3p loss and promoter hypermethylation – have been the focus of just a few studies. However, data concerning the number of ccRCC retaining one or two ^{wt}*VHL* alleles were indirectly described in several of these studies, which are listed in detail in Table 1

and summarised in Figure 1. That being said, Dagher *et al.* made the first attempt to isolate and compare active *VHL* ccRCC to inactivated *VHL* ccRCC. From their retrospective analysis of 98 patients, Dagher *et al.* identified 11 patients in whom both *VHL* copies were active and a further 22 samples in whom only a single copy of *VHL* was active, i.e. a single hit of *VHL* inactivation. Active *VHL* ccRCC tumours in which at least a single active copy of *VHL* was remaining showed a more aggressive profile than those with inactivated *VHL*. The active *VHL* tumours had higher nuclear differentiation grades and were significantly associated with sarcomatoid components, dense lymphocyte infiltration, metastasis and an increased expression of intratumoral VEGF. ccRCC cases with two active copies of *VHL* ('truly' ^{wt}*VHL* ccRCC) were also shown to also have significantly lower survival compared to those with a single or bi-allelic inactivation of *VHL*.[32] The aggressive nature of ^{wt}*VHL* ccRCC was further confirmed in another study by Turajlic *et al.*,[33] in which they examined the routes and timing of ccRCC tumour metastasis. The result was the division of their cohort, made up of 575 primary and 335 metastatic biopsies across 100 patients, into two groups: cases with rapid progression and those with attenuated progression. They found that the rapid progression group was enriched with ^{wt}*VHL* tumours, associated with lower intratumor heterogeneity and increased genomic instability.[33]

With the use of immunohistochemistry, active *VHL* ccRCC tumours have also been associated with programmed death-ligand 1 (PD-L1) expression, particularly those with two active copies of *VHL*.[34] PD-L1 is a transmembrane protein which has been shown to be expressed in a large number of cancers, including melanomas, breast and lung cancers. PD-L1 binds to its receptor programmed cell death protein 1 (PD-1) present on the cell surface of T cells, attenuating their anti-tumour response and thereby promoting cancer survival and progression.[35] PD-L1 is a key target for cancer immunotherapy; preventing the interaction of PD-L1 with PD-1 would increase tumour recognition and increase the immune response against the cancerous cells. As traditional broad cancer treatments are ineffective against renal carcinomas, targeted therapies are being developed for the treatment of ccRCC.[36] Due to the strong associations of PD-L1 expression with ^{wt}*VHL* ccRCC, Kammerer-Jacquet *et al.*[34] suggest that these cases would benefit from PD-1/PD-L1 targeted therapies.

Distinguishing 'true' from 'non-true' ^{wt}VHL ccRCC by histology

Although no molecular characterisation of 'true' ccRCC with ^{wt}*VHL* has been carried out to date, there have been attempts to distinguish histologically between ccRCC and ^{wt}*VHL* renal carcinomas. The histology of *VHL* mutated ccRCC can be seen in Figure 2A. Clear cell papillary renal cell carcinomas (CCPRCC) and renal angiomyoadenomatous tumours (RAT) are described as two relatively novel renal carcinoma subgroups. These tumours have markedly similar histologies, with clear cytoplasm within their epithelial cells with low grade-nuclei (Figure 2B). CCPRCC and RAT have also been shown to have similar morphological and immunohistochemical properties to ccRCC, but are not associated with 3p loss or *VHL* mutation, i.e. are ^{wt}*VHL*.[37–39] In spite of this, a few cases thought to be CCPRCC have shown *VHL* mutations and it is not clear if these are truly CCPRCC or ccRCC. It is thought that these subtypes of RCC are not rare, and cases are present around the world which are simply not recognised. CCPRCC is thought to have a prevalence rate of approximately 1–4% and that approximately 6% of low-grade ccRCC are, in fact, CCPRCC.[40,41] Furthermore, following immunohistochemical and fluorescence in-situ hybridisation (FISH) analysis, few RAT cases have been reclassified to translocation RCC (tRCC). tRCC are characterised by chromosomal translocations

of Xp11.2, the location of TFE3 and, less frequently, 6p21, the location of TFEB,[39] but are not associated with VHL inactivation. It may be that a proportion of ccRCC with ^{wt}VHL are CCPRCCs, RAT or tRCC. Molecular characterisation of the Cancer Genome Atlas ccRCC cases showed that five cases, which were all ^{wt}VHL tumours, had SFPQ–TFE3 fusions specific to tRCCs with clear cell histology.[8] As there is a clear genetic marker for tRCCs, identification and reclassification of ^{wt}VHL ccRCC to tRCC would be possible. However, due to the similarities between RAT, CCPRCC and ^{wt}VHL ccRCC morphologies, and considering that RAT and CCPRCC have yet to be identified as molecularly distinguishable subgroups of RCC, classification of ^{wt}VHL ccRCC to CCPRCC and RAT would be difficult. (See section 'TCEB1 RCC and CCPRCC/RAT are separate entities' for more information concerning RAT and CCPRCC).

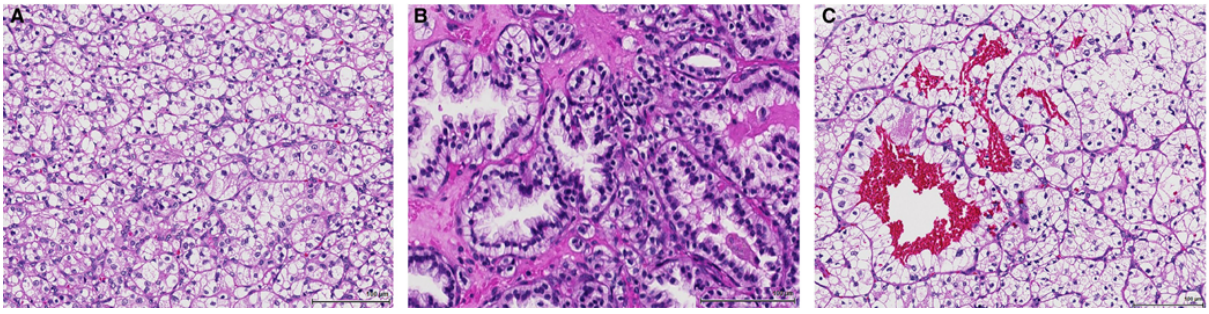

Figure 2: Haematoxylin and eosin staining showing: **A**, clear cell renal cell carcinoma with VHL mutation; **B**, clear cell papillary renal cell carcinoma with clear cytoplasm, varying papillary architecture and low-grade nuclei; **C**, tumour with chromosome 8q deletion suggestive for TCEB1 alteration with the typical morphology of a clear cell renal cell carcinoma.

TCEB1-deficient ^{wt}VHL ccRCC

Sato *et al.* were the first to identify inactivation of the *TCEB1* gene in a fraction of ^{wt}*VHL* ccRCC cases.[20] *TCEB1* codes for Elongin C that, like the VHL protein, is part of the E3 ubiquitin ligase complex. Similar to *VHL* inactivation by mutation and chromosome 3p loss in ccRCC, *TCEB1* mutations are accompanied with a chromosome 8q loss resulting in complete inactivation of *TCEB1*. No differences were found between the clinicopathological properties of cases in which *TCEB1* was inactivated and those with *VHL* inactivation. Examination of HIF-a expression via immunohistochemistry showed elevated levels of HIFa within *TCEB1*-inactivated ccRCC compared to those without *VHL* or *TCEB1* mutations and normal kidney tissue. These results support the idea that the inactivation of *TCEB1* is thought to increase HIF stabilisation via the same mechanism as *VHL* inactivation. Using the Sato *et al.*[20] cohort, along with ccRCC cases from the TCGA, Hakimi *et al.*[42] carried a more detailed analysis of the *TCEB1* inactivated ccRCC. The morphology of these tumours was distinct, but there were no 'specific' histological features. Following extensive molecular analysis of 11 *TCEB1* inactivated tumours, it was found that mutations exclusively affected two residues: Y79C/S/F/N or A100P, both essential for the interaction between Elongin C and pVHL.[42] All 11 tumours were wild-type for *VHL* and had no 3p loss with no alterations affecting the function of *PBRM1*, *SETD2* or *BAP1*. Given a similar morphology of ccRCC and *TCEB1* tumours, together with the fact that inactivation of *TCEB1* increases HIF stabilisation via the same mechanism as *VHL* inactivation, one could argue that *TCEB1* tumours can be regarded as part of a spectrum of ccRCC. A tumour with chromosome 8q deletion suggestive of *TCEB1* alteration is shown in Figure 2C.

TCEB1 RCC and CCPRCC/RAT are separate entities

The 2016 WHO classification has not designated TCEB1 RCC as its own tumour entity. For the time being they are regarded as an emerging or provisional entity, following the controversy at the 2015 WHO consensus conference in Zurich regarding the overlap of these tumours with the group of RCC associated with prominent (angio)leiomyomatous stroma. Two groups have provided seminal work characterising RCC associated with prominent (angio)leimyomatous stroma or with a clear cell papillary RCC-like morphology. The Ondrej Hes *et al.* group has pointed out that a prominent (angio)leimyomatous stroma is not a feature of a specific entity, but can be seen in otherwise typical ccRCC and even in papillary RCC. [43–45] RAT is an indolent low-grade neoplasm with overlapping histological features to clear cell papillary RCC (CCPRCC). RAT and CCPRCC are discussed as two separate entities or as two ends of a spectrum within one tumour family,[44,46] similar to the mixed epithelial and stroma tumour family encompassing tumours ranging from predominantly cystic tumours (adult cystic nephromas) to more solid variants (MEST). There are no exact criteria for distinguishing RAT and CCPRCC. Whereas RAT exhibits a voluminous stromal component, CCPRCC have a less prominent smooth muscle stroma. RAT and CCPRCC have an almost equal immunophenotype and molecular background. Both are characterised by strong and diffuse CK7 positivity and a lack of *VHL* gene abnormalities.[44]

Before the acceptance of CCPRCC as a unique entity, CK7 immunoreactivity in ccRCC has been regarded as a favourable prognostic feature or as a specific subset of ccRCC.[47,48] Therefore, it is reasonable to argue that RAT and CCPRCC are two related tumours characterised by the absence of *VHL* alterations. This is consistent with molecular results from Williamson *et al.*[49,50] Williamson's group identified *VHL* alterations in ccRCC with borderline features of CCPRCC as well as in VHL-related ccRCC with features of CCPRCC. Williamson *et al.* argue that there remain tumours with CCPRCC-like morphology, more diffuse CK7 staining and *VHL* alterations. Such tumours should be regarded as a manifestation of ccRCC and potentially represent a more aggressive neoplasm than RAT/CCPRCC. For diagnostic purposes, a diagnostic algorithm using CK7, TFE3, MET, Parafibromin and hKIM-1 has been proposed.[39,43] In selected cases, it is necessary to perform molecular analysis to identify *VHL* alterations to separate ccRCC from RAT/CCPRCC. This is clinically relevant because RAT/CCPRCC have indolent behaviour in almost all cases.

Very recently, Petersson *et al.*[45] studied a cohort with prominent smooth-muscle stroma by next-generation sequencing (NGS). They identified a group of tumours with an epithelial component indistinguishable from conventional ccRCC and distinct from CCPRCC. These tumours lack both aberrations related to the function of the *VHL* gene and hot-spot mutations in *TCEB1*. This is consistent with findings by Favazza *et al.*, who reviewed whole-slide images of 418 tumours in the published TCGA ccRCC database [51]; 93% had *VHL* alterations. The remaining wild-type *VHL* group contained translocation renal cell carcinomas, *TCEB1* mutant RCC, papillary RCC and CCPRCC (n = 27). Six cases had ccRCC-associated gene alterations (*PBRM1*, *SMARCA4*, *BAP1*, *SETD2*), leaving 11 specimens, including two high-grade or sarcomatoid RCC and two cases with prominent fibromuscular stroma (not *TCEB1* mutant). These two tumours may fall into the category of RCCs with leiomyomatous stroma and ^{wt}*VHL*, described by Petersson *et al.*[45]

Non-TCEB1-deficient ^{wt}VHL ccRCC

Although *TCEB1* inactivation is able to account for a fraction of ^{wt}*VHL* ccRCC they are extremely rare, with approximately 12 cases reported worldwide.[18,20,42] Vitally, these do not account for all ^{wt}*VHL* ccRCC cases, still leaving the root cause for many ^{wt}*VHL* ccRCC cases unknown. Referring back to the seven evolutionary subtypes of ccRCC identified by Turajlic *et al.*,[18] one of the identified evolutionary trajectories was '*VHL* wild-type'. Cases within this group were defined by active *VHL* within their tumours, i.e. no bi-allelic inactivation of *VHL*. The one case with *TCEB1*-inactivated ccRCC was not included within this group. The *VHL* wild-type group was associated with elevated cell proliferation markers and higher genomic instability. The driver phylogenetic trees for this group of patients showed various alterations as initiators for tumorigenesis, such as SCNA, *SETD2* mutations and *PTEN* mutations. Markedly, three wild-type *VHL* cases had evolutionary pathways initiated by *PBRM1* followed by either SCNA (two cases) or *SETD2* alterations (one case). Additionally, another ^{wt}*VHL* case had a driver phylogenetic tree similar to that of the 'multiple clonal drivers' subgroup in that alterations within several key driver genes, e.g. *PBRM1*, *BAP1* and *SETD2*, were the route cause for tumorigenesis. This suggests that in the absence of *VHL* alteration it may be possible for alterations in *PBRM1* alone or in combination with specific alterations within other known key driver genes to initiate ccRCC tumorigenesis.[18]

Conclusion

In recent years, significant advances have been made in the molecular characterisation of renal tumours with clear cytoplasm. A subset of tumours with clear cell histology and wild-type *VHL* can now be reclassified as other entities, including translocation RCC and the CCPRCC/RAT tumour family. Clear cell papillary RCC is defined as its own tumour entity with characteristic morphology. *TCEB1* inactivation is able to account for another fraction of ^{wt}*VHL* ccRCC. RCC with mutations in the *TCEB1* gene represents a distinct genomic subtype of RCC. Given its 'non-specific' clear cell morphology, the diagnosis requires genomic testing. Importantly, the inactivation of *TCEB1* increases HIF stabilisation via the same mechanism as *VHL* inactivation. Therefore, renal cancers with inactive *VHL* and *TCEB1* mutation-associated RCC could represent variants of ccRCC with different initiating molecular events. According to this literature review, it is evident that a small group of RCC with clear cell morphology and wild-type *VHL* (true ^{wt}*VHL* ccRCC) exists in addition to *TCEB1* mutated RCC and CCPRCC/RAT. ccRCC with active *VHL* (^{wt}*VHL*) develop via evolutionary pathways initiated by *PBRM1* and *SETD2* alterations or they fall into the 'multiple clonal drivers' subgroup. There is evidence in the literature that ^{wt}*VHL* ccRCC have aggressive behaviour. Future classification systems will consequently use the molecular background to define tumour entities. It has to be decided, in a future classification of RCC, if a diagnosis of ccRCC with ^{wt}*VHL* can be made. Currently, such tumours should be formally diagnosed as 'unclassified RCC'.

Acknowledgements

This study was supported by the Swiss National Science Foundation (SNSF grant number S-8770103-01).

Conflicts of interest

The authors declare no conflicts of interest.

Bibliography

1. Ericsson JL, Seljelid R, Orrenius S. Comparative light and electron microscopic observations of cytoplasmic matrix in renal carcinomas. Virchows Arch. Pathol. Anat. Physiol. Klin. Med. 1966; 341; 204–223.
2. Moch H, Humphrey PA, Ulbright TM, Reuter VE. WHO classification of tumours of the urinary system and male genital organs. Lyon: International Agency for Research on Cancer, 2016.
3. Moch H. World cancer report 2014, 2016 ed. Geneva: WHO Press, 2014; 436–443.
4. Latif F, Tory K, Gnarra J et al. Identification of the vonHippelLindau disease tumor-suppressor gene. Science 1993; 260; 1317–1320.
5. Gnarra JR, Tory K, Weng Y et al. Mutations of the VHL tumor suppressor gene in renal-carcinoma. Nat. Genet. 1994; 7; 85–90.
6. Knudson AG. Mutation and cancer – statistical study of retinoblastoma. Proc. Natl Acad. Sci. USA 1971; 68; 820–823.
7. Maher ER, Yates JRW, Ferguson-Smith MA. Statistical-analysis of the 2 stage mutation model in vonHippel-Lindau disease, and in sporadic cerebellar hemangioblastoma and renal-cell carcinoma. J. Med. Genet. 1990; 27; 311–314.
8. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature 2013; 499; 43–49.
9. Chen M, Ye YQ, Yang HS et al. Genome-wide profiling of chromosomal alterations in renal cell carcinoma using high-density single nucleotide polymorphism arrays. Int. J. Cancer 2009; 125; 2342–2348.
10. Herman JG, Latif F, Weng YK et al. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal-carcinoma. Proc. Natl Acad. Sci. USA 1994; 91; 9700–9704.
11. Razafinjatovo C, Bihr S, Mischo A et al. Characterization of VHL missense mutations in sporadic clear cell renal cell carcinoma: hotspots, affected binding domains, functional impact on PVHL and therapeutic relevance. BMC Cancer 2016; 16; 638.
12. Maxwell PH, Wiesener MS, Chang GW et al. The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. Nature 1999; 399; 271–275.
13. Frew IJ, Moch H. A clearer view of the molecular complexity of clear cell renal cell carcinoma. Annu. Rev. Pathol. 2015; 10; 263–289.
14. 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; 539–542.
15. Reisman D, Glaros S, Thompson EA. The SWI/SNF complex and cancer. Oncogene 2009; 28; 1653–1668.
16. Wang SS, Gu YF, Wolff N et al. BAP1 is essential for kidney function and cooperates with VHL in renal tumorigenesis. Proc. Natl Acad. Sci. USA 2014; 111; 16538–16543.
17. Kanu N, Gronroos E, Martinez P et al. SETD2 loss-of-function promotes renal cancer branched evolution through replication stress and impaired DNA repair. Oncogene 2015; 34; 5699– 5708.
18. Turajlic S, Xu H, Litchfield K et al. Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. Cell 2018; 173; 595–610.

19. Dalgliesh GL, Furge K, Greenman C et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature 2010; 463; 360–363.

20. Sato Y, Yoshizato T, Shiraishi Y et al. Integrated molecular analysis of clear-cell renal cell carcinoma. Nat. Genet. 2013; 45; 860–U191.

21. Pawlowski R, Muhl SM, Sulser T, Krek W, Moch H, Schraml P. Loss of PBRM1 expression is associated with renal cell carcinoma progression. Int. J. Cancer 2013; 132; E11–E17.

22. Miao D, Margolis CA, Gao WH et al. Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. Science 2018; 359; 801–805.

23. Guo HF, German P, Bai SS et al. The PI3K/AKT pathway and renal cell carcinoma. J. Genet. Genomics 2015; 42; 343–353.

24. 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; 611–623.

25. Li LJ, Shen C, Nakamura E et al. SQSTM1 is a pathogenic target of 5q copy number gains in kidney cancer. Cancer Cell 2013; 24; 738–750.

26. Moore LE, Jaeger E, Nickerson ML et al. Genomic copy number alterations in clear cell renal carcinoma: associations with case characteristics and mechanisms of VHL gene inactivation. Oncogenesis 2012; 1; E14.

27. Herbers J, Schullerus D, Muller H et al. Significance of chromosome arm 14q loss in nonpapillary renal cell carcinomas. Genes Chromosomes Cancer 1997; 19; 29–35.

28. Schraml P, Struckmann K, Bednar R et al. CDKNA2A muta tion analysis, protein expression, and deletion mapping of chromosome 9p in conventional clear-cell renal carcinomas: evidence for a second tumor suppressor gene proximal to CDKN2A. Am. J. Pathol. 2001; 158; 593–601.

29. Mandriota SJ, Turner KJ, Davies DR et al. HIF activation identifies early lesions in VHL kidneys: evidence for site-specific tumor suppressor function in the nephron. Cancer Cell 2002; 1; 459–468.

30. Razafinjatovo CF, Stiehl D, Deininger E, Rechsteiner M, Moch H, Schraml P. VHL missense mutations in the p53 binding domain show different effects on p53 signaling and HIF alpha degradation in clear cell renal cell carcinoma. Oncotarget 2017; 8; 10199–10212.

31. Rechsteiner MP, von Teichman A, Nowicka A, Sulser T, Schraml P, Moch H. VHL gene mutations and their effects on hypoxia inducible factor HIF alpha: identification of potential driver and passenger mutations. Cancer Res. 2011; 71; 5500– 5511.

32. Dagher J, Kammerer-Jacquet S-F, Brunot A et al. Wild-type VHL clear cell renal cell carcinomas are a distinct clinical and histologic entity: a 10-year follow-up. Eur. Urol. Focus 2016; 1; 284–290.

33. Turajlic S, Xu H, Litchfield K et al. Tracking cancer evolution reveals constrained routes to metastases: TRACERx renal. Cell 2018; 173; 581–594.

34. Kammerer-Jacquet SF, Crouzet L, Brunot A et al. Independent association of PD-L1 expression with noninactivated VHL clear cell renal cell carcinomaa finding with therapeutic potential. Int. J. Cancer 2017; 140; 142–148.

35. Zitvogel L, Kroemer G. Targeting PD-1/PD-L1 interactions for cancer immunotherapy. Oncoimmunology 2012; 1; 1223– 1225.

36. Barata PC, Rini BI. Treatment of renal cell carcinoma: current status and future directions. CA Cancer J. Clin. 2017; 67; 507– 524.

37. Wolfe A, Dobin SM, Grossmann P, Michal M, Donner LR. Clo nal trisomies 7,10 and 12, normal 3p and absence of vhl gene mutation in a clear cell tubulopapillary carcinoma of the kidney. Virchows Arch. 2011; 459; 457–463.

38. Michal M, Hes O, Nemcova J et al. Renal angiomyoadenomatous tumor: morphologic, immunohistochemical, and molecular genetic study of a distinct entity. Virchows Arch. 2009; 454; 89–99.

39. Deml KF, Schildhaus HU, Comperat E et al. Clear cell papillary renal cell carcinoma and renal angiomyoadenomatous tumor two variants of a morphologic, immunohistochemical, and genetic distinct entity of renal cell carcinoma. Am. J. Surg. Pathol. 2015; 39; 889–901.

40. Gill S, Kauffman EC, Kandel S, George S, Schwaab T, Xu B. Incidence of clear cell papillary renal cell carcinoma in lowgrade renal cell carcinoma cases: a 12-year retrospective clinicopathologic study from a single cancer center. Int. J. Surg. Pathol. 2016; 24; 207–212.

41. Zhou HJ, Zheng SJ, Truong LD, Ro JY, Ayala AG, Shen SS. Clear cell papillary renal cell carcinoma is the fourth most common histologic type of renal cell carcinoma in 290 consecutive nephrectomies for renal cell carcinoma. Hum. Pathol. 2014; 45; 59–64.

42. Hakimi AA, Sato Y, Morikawa T et al. TCEB1-mutated renal cell carcinoma: a distinct genomic and morphologic subtype. J. Urol. 2014; 191; E249–E249.

43. Petersson F, Grossmann P, Hora M et al. Renal cell carcinoma with areas mimicking renal angiomyoadenomatous tumor/ clear cell papillary renal cell carcinoma. Hum. Pathol. 2013; 44; 1412–1420.

44. Peckova K, Grossmann P, Bulimbasic S et al. Renal cell carcinoma with leiomyomatous stroma – further immunohistochemical and molecular genetic characteristics of unusual entity. Ann. Diagn. Pathol. 2014; 18; 291–296.

45. Petersson F, Martinek P, Vanecek T et al. Renal cell carcinoma with leiomyomatous stroma: a group of tumors with indistinguishable histopathologic features, but 2 distinct genetic profiles: next-generation sequencing analysis of 6 cases negative for aberrations related to the VHL gene. Appl. Immunohistochem. Mol. Morphol. 2018; 26; 192–197.

46. Srigley JR, Delahunt B, Eble JN et al. The International Society of Urological Pathology (ISUP) Vancouver classification of renal neoplasia. Am. J. Surg. Pathol. 2013; 37; 1469–1489.

47. Mertz KD, Demichelis F, Sboner A et al. Association of cytokeratin 7 and 19 expression with genomic stability and favorable prognosis in clear cell renal cell cancer. Int. J. Cancer 2008; 123; 569–576.

48. Mai KT, Kohler DM, Belanger EC, Robertson SJ, Wang D. Sporadic clear cell renal cell carcinoma with diffuse cytokeratin 7 immunoreactivity. Pathology 2008; 40; 481–486.

49. Williamson SR, Zhang SB, Eble JN et al. Clear cell papillary renal cell carcinoma-like tumors in patients with von Hippel-Lindau disease are unrelated to sporadic clear cell papillary renal cell carcinoma. Am. J. Surg. Pathol. 2013; 37; 1131–1139.

50. Williamson SR, Gupta NS, Eble JN et al. Clear cell renal cell carcinoma with borderline features of clear cell papillary renal cell carcinoma combined morphologic, immunohistochemical, and cytogenetic analysis. Am. J. Surg. Pathol. 2015; 39; 1502– 1510.

51. Favazza L, Chitale DA, Barod R et al. Renal cell tumors with clear cell histology and intact vhl and chromosome 3p: a histological review of tumors from the cancer genome atlas database. Mod. Pathol. 2017; 30; 1603–1612.

52. Banks RE, Tirukonda P, Taylor C et al. Genetic and epigenetic analysis of von Hippel-Lindau (VHL) gene alterations and relationship with clinical variables in sporadic renal cancer. Cancer Res. 2006; 66; 2000–2011.

53. Beroukhim R, Brunet JP, Di Napoli A et al. Patterns of gene expression and copy-number alterations in von-Hippel Lindau disease-associated and sporadic clear cell carcinoma of the kidney. Cancer Res. 2009; 69; 4674–4681.

54. Brauch H, Weirich G, Brieger J et al. VHL alterations in human clear cell renal cell carcinoma: association with advanced tumor stage and a novel hot spot mutation. Cancer Res. 2000; 60; 1942–1948.

55. Kondo K, Yao M, Yoshida M et al. Comprehensive mutational analysis of the VHL gene in sporadic renal cell carcinoma: relationship to clinicopathological parameters. Genes Chromosomes Cancer 2002; 34; 58–68.

Renal tumour classifications as designated by the 2022 WHO guidelines

The 5th edition of the World Health Organisation classification of Urogenital tumours released in 2022 has for the first time introduced molecular-driven tumour classifications in addition to the traditional morphology-based system (based on cytoplasmic features, architectural features, anatomical locations, and familial predisposition syndromes). These molecularly classified tumours may show heterogeneous morphological phenotypes and therefore cannot be diagnosed by their morphology alone (Table 2).

Clear cell renal cell carcinoma are composed of cells with clear cytoplasms and distinct cell membranes with a characteristic molecular background e.g. *VHL* inactivation mentioned above. Clear cell RCC can be architecturally diverse, however, solid alveolar and acinar patterns are commonly observed. These criteria for classification as ccRCC have not altered in the new guidelines. ELOC-mutated renal cell carcinoma, however, are now considered a separate molecularly defined RCC subtype. Although, due to their broad morphological spectrum, their differential diagnosis remains ccRCC or clear cell papillary RCC with prominent fibromuscular septation and Cytokeratin 7 (CK7) positivity, requiring molecular assessment for classification as ELOC-mutated RCC.

Table 2: Renal tumour classification according to the World Health Organisation as of 2022 [1]

Renal cell tumours	
Clear cell renal tumours	
8310/3	Clear cell renal cell carcinoma
8316/1	Multilocular cystic renal neoplasm of low malignant potential
Papillary renal tumours	
8260/0	Papillary adenoma
8260/3	Papillary renal cell carcinoma†
Oncocytic and chromophobe renal tumours	
8290/0	Oncocytoma
8317/3	Chromophobe cell renal carcinoma
N/A	Other oncocytic tumours of the kidney
Collecting duct tumours	
8319/3	Collecting duct carcinoma
Other renal tumours	
8323/1	Clear cell papillary renal cell tumour†
8480/3	Mucinous tubular and spindle cell carcinoma
8316/3	Tubulocystic renal cell carcinoma
8316/3	Acquired cystic disease–associated renal cell carcinoma
8311/3	Eosinophilic solid and cystic renal cell carcinoma
8312/3	Renal cell carcinoma, NOS
Molecularly defined renal carcinomas	
8311/3	TFE3-rearranged renal cell carcinomas
8311/3	TFEB-altered renal cell carcinomas
8311/3	ELOC (formerly TCEB1)-mutated renal cell carcinoma

8311/3 Fumarate hydratase–deficient renal cell carcinoma
8311/3 Hereditary leiomyomatosis and renal cell carcinoma (HLRCC) syndrome–associated renal cell carcinoma
8311/3 Succinate dehydrogenase–deficient renal cell carcinoma
8311/3 ALK-rearranged renal cell carcinomas
8510/3 Medullary carcinoma, NOS
8510/3 SMARCB1-deficient medullary-like renal cell carcinoma
8510/3 SMARCB1-deficient undifferentiated renal cell carcinoma, NOS
8510/3 SMARCB1-deficient dedifferentiated renal cell carcinomas of other specific subtypes

Metanephric tumours

8325/0 Metanephric adenoma
9013/0 Metanephric adenofibroma
8935/1 Metanephric stromal tumour

Mixed epithelial and stromal renal tumours

8959/0 Mixed epithelial and stromal tumour
8959/0 Adult cystic nephroma
8959/0 Paediatric cystic nephroma

Renal mesenchymal tumours

Adult renal mesenchymal tumours
8860/0 Angiomyolipoma
8860/0 Oncocytic angiomyolipoma
8860/0 Angiomyolipoma with epithelial cysts
8860/1 Angiomyolipoma, epithelioid
9161/1 Haemangioblastoma
8361/0 Juxtaglomerular tumour
8361/0 Functioning juxtaglomerular cell tumour
8361/0 Non-functioning juxtaglomerular cell tumour
8966/0 Renomedullary interstitial cell tumour

Paediatric renal mesenchymal tumours
8967/0 Ossifying renal tumour of infancy
8960/1 Mesoblastic nephroma
8960/1 Classic congenital mesoblastic nephroma
8960/1 Cellular congenital mesoblastic nephroma
8960/1 Mixed congenital mesoblastic nephroma
8963/3 Malignant rhabdoid tumour of the kidney
8964/3 Clear cell sarcoma of kidney

Embryonal neoplasms of the kidney

Nephroblastic tumours
N/A Nephrogenic rests
N/A Perilobar nephrogenic rests
N/A Intralobar nephrogenic rests
N/A Nephroblastomatosis
8959/1 Cystic partially differentiated nephroblastoma
8960/3 Nephroblastoma

Miscellaneous renal tumours	
Germ cell tumours of the kidney	
9084/0	Prepubertal-type teratoma
9084/3	Teratoma with carcinoid (neuroendocrine tumour)
9071/3	Yolk sac tumour, NOS
9085/3	Mixed teratoma–yolk sac tumour

ICD-O-3.2 code followed by ICD-O label (subtypes are indicated in grey text, with the label indented)

N/A, not available (provisional entity).

† Labels marked with a dagger have undergone a change in terminology of a previous code.

II

ELOC Renal Cell Carcinoma

This chapter is composed of my work concerning the identification and molecular assessment of ELOC Renal Cell Carcinoma within a cohort of tumours present within the University Hospital of Zurich's Tissue Biobank. As of 2022, the World Health Organisation has classified ELOC RCC as a separate tumour entity[3]. Attempts have been made to distinguish ELOC RCC as a more aggressive subtype and physiologically distinct to classical *VHL* inactivated ccRCC[4,5]. However, given the presence of both ELOC and pVHL within the E3 ubiquitin ligase complex[6] there may be considerable overlap between *ELOC* inactivated RCC with clear cell morphology and classical *VHL* inactivated ccRCC with respect to their molecular landscapes and routes to tumourigenesis[7]. With only 17 *ELOC* ccRCC having been identified worldwide, the identification of additional *ELOC* RCC and an assessment of their molecular and clinical phenotypes are of great value to the community. My contributions to the subsequent article is as follows: Identifying the cohort and experimentally determining *VHL* status, obtaining and analysing the sequencing data, obtaining and analysing the mass spectrometry data, performing the *in silico* modelling of the protein structures, assessing the impact of variants on those structures, all statistical analysis, production of the visualisations and writing the article.

[3] International Agency for Research on Cancer. *WHO Classification of Tumours ; Urinary and male genital tumours. 5th ed.* (2022).

[4] Hakimi, A. A. *et al.* TCEB1-mutated Renal Cell Carcinoma: A Distinct Genomic and Morphologic Subtype. *Journal of Urology* **191**, E249–E249 (2014).

[5] DiNatale, R. G. *et al.* Putative Drivers of Aggressiveness in TCEB1-mutant Renal Cell Carcinoma: An Emerging Entity with Variable Clinical Course. *Eur. Urol. Focus* **7**, 381–389 (2021).

[6] Stebbins, C. E., Kaelin, W. G. & Pavletich, N. P. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science* **284**, 455–461 (1999).

[7] Sato, Y. *et al.* Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.* **45**, 860–867 (2013).

Consequences of *ELOC* alterations in renal cancer

Aashil A. Batavia[1,2,3], Dorothea Rutishauser[1], Bettina Sobottka[1], Peter Schraml[1], Niko Beerenwinkel[2,3] and Holger Moch[1]*

[1]*Department of Pathology and Molecular Pathology, University and University Hospital Zurich, Schmelzbergstrasse 12, 8091 Zurich,Switzerland*

[2]*Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, 4058 Basel*

[3]*SIB Swiss Institute of Bioinformatics, Mattenstrasse 26, 4058 Basel, Switzerland*

* Corresponding author

Running title: RCC with *ELOC* and *VHL* alterations

Conflicts of interest: The authors declare no potential conflicts of interest

Abstract

Approximately 70% of clear cell renal cell carcinoma are characterised by the biallelic inactivation of *VHL* on chromosome 3p. *ELOC*-mutated renal cell carcinoma with biallelic *ELOC* inactivation on chromosome 8q are considered a novel subtype of renal cancer possessing a morphological overlap with ccRCC; however, the frequency and consequences of *ELOC* alterations in ^{wt}*VHL* and ^{mut}*VHL* ccRCC is unclear. In this study, we characterise 123 renal tumours with clear cell morphology with known *VHL* mutation status to assess morphological and molecular consequences of *ELOC* inactivation. Using Oncoscan and whole exome sequencing we identify 18 *ELOC* deleted RCC, three of which contain *ELOC* mutations resulting in the biallelic inactivation of *ELOC*. Biallelic *ELOC* and biallelic *VHL* aberrations were mutually exclusive, although two *ELOC*-mutated RCC showed monoallelic *VHL* alterations. Using High Ambiguity Driven Biomolecular Docking we report a novel *ELOC* variant containing a duplication event disrupting ELOC-pVHL interaction alongside the frequently seen Y79C alteration. Using HRM mass spectrometry we show RCC with biallelic *ELOC* alterations have significantly reduced ELOC expression but similar CAIX and VEGFA expression when compared to classical ccRCC with *VHL* inactivation. These data demonstrate that RCC with *ELOC* and *VHL* alterations have comparable downstream effects with similar pathways to ccRCC tumourigenesis indicating that both entities are closely related.

Statement of significance

ELOC RCC have very recently been endorsed as a novel RCC entity by the World Health Organisation. Biallelic *ELOC* and *VHL* aberrations are essential diagnostic criteria for ccRCC and *ELOC* RCC. We show monoallelic *VHL* and *ELOC* alterations can coincide contrary to the notion that *ELOC* and *VHL* aberrations are mutually exclusive. Only biallelic inactivation of *ELOC* results in a reduction in ELOC protein expression, driving tumorigenesis similarly to classical ccRCC with *VHL* inactivation; functioning as a classical tumour suppressor gene.

Introduction

Clear cell renal cell carcinoma (ccRCC) is the most predominant histological subtype of renal cell carcinoma which in turn is amongst the 10 most commonly diagnosed malignancies in the world [1,2]. The majority of ccRCC are characterised by the biallelic loss of the von Hippel Lindau (*VHL*) tumour suppressor gene located on chromosome 3. The biallelic inactivation of *VHL* can result from the combination of chromosomal 3p loss, *VHL* mutation and/or *VHL* promoter hypermethylation [3,4]. As a key component of the E3 ubiquitin ligase complex, the VHL protein or pVHL is essential for the ubiquitin-mediated proteolytic degradation of HIF-α [5]. In the absence of pVHL in ccRCC, HIF-α is stabilised promoting its translocation into the nucleus in complex with HIF-β culminating in the overexpression of HIF target genes. Many of these genes are involved in metastatic processes such as dedifferentiation, VEGF production, and cell migration [6].

Whereas loss of chromosome 3p is found in >90% of ccRCC, *VHL* point mutations are reported in only 60%-70% patients, and epigenetic silencing in a further 5-10%. Past studies have largely focused their analysis on ccRCC with inactivated *VHL* while largely ignoring a subset of tumours retaining two wild-type *VHL* alleles [4,7,8]. As the number of these tumours is small, molecular analysis of wild-type *VHL* (^{wt}*VHL*) ccRCC are limited [9]. Sato et al. first identified *ELOC*, also known as *TCEB1*, inactivation in a fraction of ^{wt}*VHL* ccRCC following the assessment of 240 ccRCC samples. *ELOC* codes for Elongin C, a binding partner of pVHL in the E3 ubiquitin ligase complex [10]. As with the biallelic loss of *VHL*, *ELOC* mutations are accompanied by the loss of chromosome 8q resulting in its biallelic inactivation. *ELOC*-mutated RCC have shown an elevated expression of HIF-α in comparison to tumours with active *VHL* and active *ELOC*, suggesting similar downstream consequences within the same molecular pathway promoting tumorigenesis as in *VHL* inactivated ccRCC.

In 2022, ELOC-mutated RCC was endorsed as a novel, molecularly defined tumour entity by the World Health Organisation (WHO) [11]. Thus far 17 ccRCC with *ELOC* mutations have been identified: 8 in the Sato et al cohort [10], 5 in the MSK-IMPACT cohort [12], 3 in The Cancer Genome Atlas cohort[4] and 1 in the TRACERx renal cohort [7]. All identified *ELOC* mutations have been found to affect 3 codons of Elongin C present at pVHL interaction sites: Y79, I95 and A100. The insertion of a Cysteine in place of a Tyrosine at amino acid position 79 is the most common aberration present in 10 of the 17 cases[12]. All mutations were accompanied by chromosome 8q deletions leading to the biallelic inactivation of *ELOC*. Given the collaborative function of pVHL and Elongin C, *ELOC* deletions in *ELOC*-mutated RCC and *ELOC* non-mutated ccRCC may impact total Elongin C expression contributing to HIF-stabilisation.

Here we use a retrospective cohort of 464 ccRCC samples obtained from our Renal Tumour Biobank to determine the effect of *ELOC* aberrations on RCC with clear cell morphology. We identify three *ELOC*-mutated RCC of which two disprove the notion that *ELOC* mutations and *VHL* aberrations are mutually exclusive amongst RCC with clear cytoplasm. Structural assessments of *ELOC* mutations found in our cohort including a novel variant prevent the VHL:V155-ELOC:Y79 interaction from taking place. We find ELOC protein expression is significantly reduced in *ELOC*-mutated RCC but CAIX and VEGFA expression is no different from classical *VHL* inactivated ccRCC.

Methods

Patients and Tissue Specimens

Adult renal epithelial neoplasms with available pathology material and diagnosed at the University Hospital Zurich (USZ) form the basis for this analysis. We use 464 formalin fixed and paraffin embedded (FFPE) RCC samples from the USZ's Department of Pathology and Molecular Pathology tissue biobank ranging from 1993-2019, which were primarily diagnosed as clear cell RCC according to the 2016 WHO classifications [1]. These cases were included in prior publications evaluating other pathologic features including immunohistochemistry and molecular analyses [13–15]. The study did not include consultation material. The RCC selected for further analysis in this study were retrieved retrospectively during a study of *VHL* sequencing of all ccRCC, for which matched normal renal tissue was available. Hematoxylin-eosin stained sections from each tumour were reviewed by a pathologist (HM).

Clinical characteristics (age, sex and presentation), size, pT stage, pN, and pM stage were evaluated from the pathology reports. In each case of this study, the following microscopic features were evaluated: WHO/ISUP grade, presence of fibromuscular bands, nodular configuration, architecture (solid, cystic, alveolar and papillary), presence of voluminous cells with clear cytoplasm, and clear cell papillary RCC-like nuclear arrangement. As part of this study, immunohistochemistry (CA-IX, CK7, CD10, TFE3, HMB45) was separately performed to exclude MiTF translocation RCC or other renal tumour entities. When indicated, molecular analyses (TFE3 and TFEB FISH) were done in many cases during routine work-up.

Our retrospective study fulfilled the legal conditions according to the Swiss Law "Humanforschungsgesetz (HFG)", which, in exceptional cases, allows the use of biomaterial and patient data for research purposes without informed consent, if i) it is impossible or disproportionately difficult to obtain patient consent; ii) there is no documented refusal; iii) research interests prevail the individual interest of a patient. Law abidance of this study was reviewed and approved by the ethics commission of the Canton Zurich (BASEC_2019-01959).

DNA extraction

Genomic DNA (gDNA) was extracted from normal and tumour FFPE samples with a minimum of 70% tumour cells using 0.6 mm punch needles and the Maxwell 16 FFPE Tissue DNA Purification Kit from Promega. The quality and yield of extracted DNA was determined using the NanoDrop spectrophotometer and Qubit Fluorometric quantification from Thermo Fisher Scientific.

Sanger Sequencing

To determine the *VHL* status of our ccRCC cohort (n = 464), Sanger sequencing was carried out as stated previously [14,15]. Sequencing was carried out using the BigDye Terminator v1.1 Cycle Sequencing Kit (Thermo Fisher Scientific Inc.). Mutations were identified using the AF010238 (GenBank Accession) sequence as the *VHL* reference for comparison and CLC Main Workbench 8.1. Tumours without *VHL* mutations were selected for further analysis.

Copy number determination

RCC with *ELOC* alterations (*ELOC* mutations or chromosome 8q deletions) are largely observed in RCC without chromosome 3p deletions. Therefore, we assessed genome-wide copy number aberrations (CNA) in 96 samples without *VHL* mutations applying Affymetrix OncoScan® CNV FFPE Microarrays at IMGM laboratories, Munich and the Children's Hospital Zurich[16]. 27 cases with *VHL* mutations were used as a control group and assessed using the Affymetrix CytoScan HD as part of our ongoing biobanking initiative at the University Hospital Zurich [17].

From the raw paired .CEL files, log2 ratios (L2R) and B-allele frequencies (BAF) were determined, following normalisation, and segmented using Allele-Specific Copy number Analysis of Tumors (ASCAT) [18] via EaCoN. ASCAT provides allele-specific copy number profiles, an estimation of the number of aberrant cells and global ploidy of the tumour samples. An absolute L2R cutt-off value of 0.05 is used when calling losses and gains. The genewise copy number calls of *VHL* and *ELOC* were also verified using Nexus Copy Number 10.0 (Biodiscovery, Inc., El Segundo, CA, USA).

Bisulfite sequencing

Bisulfite sequencing was used to determine the methylation status of the *VHL* promoter region. 500ng of gDNA was bisulfite converted using the EpiTect Bisulfite Kit (QIAGEN) and the protocol for FFPE derived DNA. The following primers were then used to amplify the *VHL* promoter prior to Sanger sequencing using BigDye Terminator v1.1 Cycle Sequencing Kit: forward: 5'-gagttttttaggttatttttgtaat-3', reverse: 5'-tcaccctaaatatatcctacctcaaaa-3'. The methylation status of the promoter site for each sample was assessed using CLC Main Workbench 8.1. Genomic DNA extracted from HEK 293 cells are used as positive controls for the bisulfite conversion and A-704 cells known to have methylated *VHL* promoters were used to validate the protocol [19].

Whole-exome sequencing and variant calling

200ng of FFPE derived gDNA from 18 *ELOC* deleted ccRCC with matched normals were sequenced at the Functional Genomics Centre Zurich. The quality of each sample was assessed using the Agilent 4200 TapeStation. The Agilent SureSelect^{XT2} Human All Exon V6+UTR workflow and kit was used for library prep. Samples were sequenced on the Illumina Novaseq producing 150 bp paired-end reads.

The mean coverage for tumour and normal samples was 159x and 100x respectively. A computational data analysis pipeline was built using Snakemake adapting the framework of NGS-pipe with updated methods [20]. Adapter clipping and trimming of low-quality bases was carried out on raw paired-end FASTQ files using trimmomatic v0.36 [21]. MultiQC v1.9 was used to assess the quality of the resulting FASTQ files [22]. Trimmed FASTQ files were aligned to the human genome reference build hg19 using the Burrows-Wheeler Aligner algorithm (BWA-mem) [23]. Paired BAM files were merged upon which secondary alignments and PCR duplicates were removed using Picard Tools and Samtools v1.4 [24,25]. Regions with indels were locally realigned in order to improve alignment quality, which preceded base recalibration used to adjust the base quality scores correcting for systematic technical errors resulting from the sequencer. A custom panel of normals (PON) was produced to account for any

technical biases. The PON consists of 17 normal samples (one sample was removed due to poor coverage). Single-nucleotide variants (SNVs), insertions and deletions (indels) were called for each tumour sample using Mutect2 given the matched normal, the PON, and population allele frequencies of common and rare variants from The Genome Aggregation Database (gnomAD) [26]. SNVs and indels were called using only the PON and germline resource for the one tumour sample (12T) where the matched normal was removed. The Genome analysis toolkit v4.8.1.0 was used for base recalibration, production of the PON and SNV and indel calling by Mutect2.

Structural impact of *ELOC* mutations

The consequences of identified *ELOC* mutations were assessed using the High Ambiguity Driven Biomolecular Docking (HADDOCK) webserver v2.4 [27] and UCSF Chimera v1.15 [28]. A structure of the pVHL-ELOC-ELOB complex was obtained from the Protein Data Bank (accession: 1VCB) [29] and manipulated using pdb-tools [30]. As the location of the *ELOC* mutations identified lie within the pVHL binding domain, we focused on the interaction surface between ELOC and pVHL and removed the chain corresponding to Elongin B (ELOB) from subsequent analysis. Docking between ELOC structural variants and pVHL is determined using a three step refinement process implemented via HADDOCK. In brief, the geometric parameters are frozen to first identify the optimum interaction surfaces for the given chains. In the second step flexibility is introduced to the interacting partners freeing the side chains to orient themselves. In the final step, the structure is immersed with water molecules to reflect the interior of a cell causing the side chains to be positioned in an orientation most similar to the true structure. Multiple models are produced to find the optimum binding pose of each structure. Single amino acid changes were introduced directly into the structure of ELOC prior to refinement to obtain the desired structures for input into HADDOCK. Structures containing a duplication i.e. structures with F77_T78dup are modelled using SWISS-MODEL[31] with the original reference ELOC structure (within 1VCB) serving as a template. Water molecules present in 1VCB were removed from all structures as they are not required for the refinement process. We applied HADDOCK to the wild-type ELOC structure and four mutant structures: Y79C, K80E, F77_T78dup and one containing both the F77_T78 duplication and the K80E substitution. The best model in each instance, defined by the lowest HADDOCK score:

$$HADDOCK\ Score\ water\ refinement = 1.0\ E_{vdw} + 0.2\ E_{elec} + 1.0\ E_{desol}$$

where E_{vdw} is the Van der Waal (VdW) energy, E_{elec} is the electrostatic intermolecular energy and E_{desol} is the desolvation energy, was visualised in UCSF Chimera for validation. Using UCSF Chimera we identified the changes in side chain contacts from the wild-type state in each of the four mutated structures. Significant differences between the VdW, electrostatic and desolvation energies along with the buried surface area (BSA) were identified using the Kruskal–Wallis one-way analysis of variance test and applying the Dunn test for pairwise comparisons between the different mutant structures; q-values were reported following multiple testing correction using the Benjamini-Hochberg Procedure[32].

Immunohistochemistry

Multiple 2 µm thick FFPE sections from the 18 ccRCC samples with chromosomal *ELOC* deletions were taken and stained against CAIX, CK7, CD10, TFE3, HMB-45 and with 34BET12 on a Benchmark Ultra platform (Ventana) with protocols used for routine diagnostics (See Table S1). Assessment of histopathological characteristics and manual scoring of the IHC stainings were carried out by BS and HM from the University Hospital Zurich.

Protein extraction and digestion

2-4 punches per patient samples using a 0.6 mm needle were collected and processed. Samples were deparaffinised with Xylenes and rehydrated with ethanol. Samples were subsequently lysed with Biognosys' Lysis Buffer and Bioruptor device (Diagenode, Seraing, Belgium). Extracted proteins were reduced with DDT and alkylated with iodaciodoacetamide. After precipitation with acetone, proteins were resuspended with Biognosys' Denature Buffer and digested first with LysC (Wako Chemicals, 1ug per sample) for 3 hours and then with trypsin (Promega Co., 1:20 protease to total protein ratio) overnight at 37°C. The resulting tryptic peptides were cleaned with a C18 MicroSpin plate (The Nest Group, Inc.™, MA, USA) and dried down using a SpeedVac system. Peptides were resuspended again in 1% acetonitrile (ACN) and 0.1% formic acid (FA) and spiked with iRTcalibration peptides [33]. Peptide concentrations were determined using the Micro BCA Protein Assay Kit (Thermo Fisher Scientific Inc.).

HRM mass spectrometry acquisition

For the data-independent acquisition (DIA) 1 µg of the resulting peptide mixture per samples was injected onto an in-house packed 60 cm long analytical reversed phase column consisting of a PicoFrit emitter (75 µm ID, 10µm tip; New Objective, Woburn, MA, USA) and 1.7 µm Charged Surface Hybrid C18 particles (Waters). The non-linear LC gradient was mixed with buffer A (1% ACN in water with 0.1% FA) and buffer B (80% ACN in water with 0.1% FA) according to the following scheme: 1-59% solvent B in 85 minutes followed by 59-90% B in 10 seconds, 90% B for 8 minutes, 90%-1% in 10 seconds and at the end 1% for 5 min. The column temperature was set to 60°C and a flow rate of 250 nl/min was used over the entire gradient.

The analytical system used was a Thermo Scientific EASY-nLC 1200 nano-liquid chromatography system connected to a ThermoScientific QExactiveHF-X mass spectrometer equipped with a Nanospray Flex ™ IonSource. The DIA method consisted of one full range MS1 scan and 29 DIA segments (MS2) and was adopted from Bruderer et al. [34] Generation of a sample-specific spectral library was done using shotgun LC-MS/MS on a Thermo Q-Exactive HF-X instrument.

Proteomic data processing

For the HRM data analysis the MS data was analysed using Spectronaut™ Pulsar software (Biognosys AG). The false discovery rate (FDR) on peptide and protein level was set to 1 % and the data was filtered using row-based extraction. The assay library (protein inventory) generated in this project was used for the analysis. Two samples contained a lower number of identified proteins and showed high relative abundance of blood proteins. In order to exclude a possible effect of these proteins on the quantitative data set we remove 49 blood specific proteins (derived from Hortin et al.[35]) prior to log2

transformation and global normalisation of the data. We apply Welch's one-way ANOVA for the comparison of ELOC, CAIX and VEGFA protein expression levels reporting q-values obtained by applying the Benjamini-Hochberg Procedure for FDR correction.

Data Accessibility

The Oncoscan SNP-array data generated in this study are publicly available in GEO via the accession number GSE201277. The accessions of the samples used can be found in Additional File 1. The WES data have been deposited on the Sequence Read Archive accessed via the project accession PRJNA831852. The mass spectrometry data is available on ProteomeXchange via PRIDE with the accession number PXD033291. The structural models have been placed on ModelArchive.

Results

ELOC-mutated RCC are characterised by the biallelic inactivation of the *ELOC* gene. *ELOC* mutations have been considered independently from any form of *VHL* inactivation, i.e., *VHL* mutations, chromosome 3p deletions and *VHL* promoter hypermethylation. In our cohort of 464 ccRCC, we identified 96 samples without *VHL* mutations (20.7%). Genome-wide copy number assessment and bisulfite sequencing was carried out on these 96 tumours. In addition, CNAs were also determined for 27 ccRCC with *VHL* mutations (^{mut}*VHL*; Figure 1). Genome-wide copy number assessment of these 123 tumours allowed the identification of 18 tumours with monoallelic chromosomal ELOC deletions (see Figure 1 and Table 1). Of these 18 samples with *ELOC* deletions, 3 have no *VHL* aberrations, 4 possess a monoallelic inactivation of *VHL* (3 with chromosomal *VHL* deletions and 1 with a hypermethylated promoter region), and 11 possess a biallelic inactivation of *VHL*. In total, *VHL* mutations were found in 7 of the 18 samples.

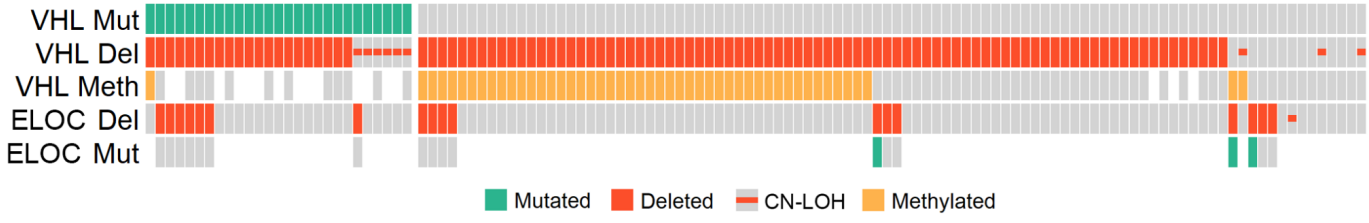

Figure 1: *VHL* and *ELOC* aberrations in 27 ^{mut}*VHL* and 96 samples with no *VHL* mutations. 18 tumours show monoallelic *ELOC* deletions. Whole Exome sequencing of these 18 samples identified 3 *ELOC*-mutated tumours possessing biallelic *ELOC* inactivation i.e. *ELOC* deletion with an *ELOC* mutation.

Table 1: Clinicopathological features of 18 *ELOC* deleted renal carcinomas

Sample	Gender	pT Stage	ISUP Grade	Age at diagnosis (years)	Last follow up since OP (Months)	Survival (1 = Dead, 0 = Alive)
1T	Male	1b	2	58	7	1
2T	Female	1b	3	62	212	1
3T	Female	3a	2	76	8	1
4T	Male	3b	4	68	55	0
5T	Male	3b	3	58	99	1
6T	Female	1a	2	74	88	1
7T	Female	1b	2	78	12	1
8T	Female	1b	2	59	83	0
9T	Male	1a	3	54	91	0
10T	Male	3	3	68	80	1
12T	Male	1a	2	52	151	0
13T	Male	3b	2	73	151	0
14T	Male	3a	3	71	49	1
15T	Male	1b	2	54	221	0
16T	Male	1a	2	53	126	0
17T	Male	1a	2	48	66	0
18T	Male	3a	3	69	36	0
19T	Male	3a	3	61	3	0

Three *ELOC*-mutated RCC with biallelic inactivation of *ELOC* are found: one without *VHL* gene alterations and two cases with monoallelic *VHL* aberrations (Figure 2). Tumour 7T possesses a 3p deletion while tumour 18T possesses a hypermethylated *VHL* promoter region (Figure 2). Along with 3p deletion, tumour 7T possesses losses in Chr1p, Chr8, Chr9, Chr14 and Chr18 with focal amplifications in Chr3q, Chr4q, Chr12p, Chr15q and Chr16q. Sample 18T has fewer aberrations with deletions in Chr6, Chr8, Chr9, and Chr14 with an amplification in Chr20. These aberrations have all previously been identified in ccRCC with chr14 being amongst the most common chromosomal aberrations following 3p deletion [4,7,8].

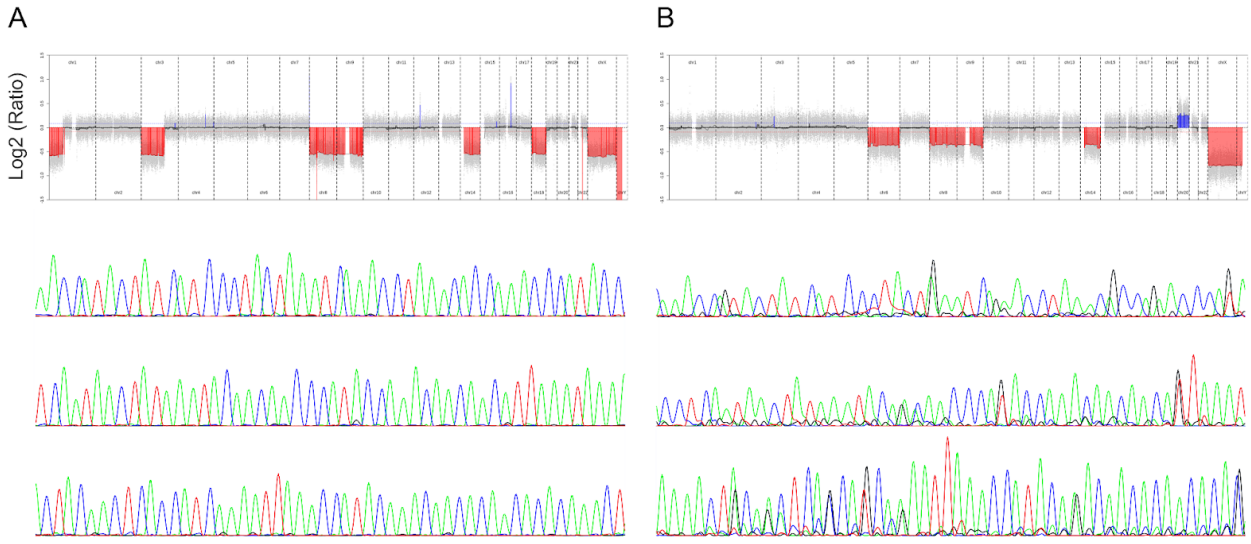

Figure 2: Copy number aberrations and the *VHL* promoter methylation trace following bisulfite sequencing for two samples with *ELOC* mutations and *VHL* aberrations. **A)** Tumour sample 7T possesses chromosome 3p loss with an absolute log2 ratio above 0.5 with no promoter hypermethylation. **B)** Tumour sample 18T possesses no chromosome 3p loss while the *VHL* promoter is hypermethylated as shown by the presence of guanines (black trace) adjacent to cytosines (blue trace).

Clinico-pathologic features of RCC with chromosomal ELOC deletion

The stage and grade distribution of 18 *ELOC*-inactivated RCC is shown in Table 1. The age of the patients ranged from 48 to 78 years (mean 61 years). The male to female ratio was 2.6:1. *ELOC*-inactivated RCC showed clear cells with voluminous cytoplasms and prominent borders similar to ccRCC. Additionally, all tumours were characterised by fibromuscular bands resulting in a nodular configuration. There was no apical nuclear arrangement found in clear cell papillary renal cell carcinoma. Tumours showed a solid alveolar architecture and cystic features in 13 cases. Five tumours contain tubular/branching tubular structures. Immunohistocemically, all 18 tumours were positive for CAIX and negative for TFE3, HMB45 and high molecular weight cytokeratins. The expression of CD10 varied across the samples with 14 containing a high number of positive cells (>90%) while three tumours showed heterogeneous CD10 expression (Table 2)

Table 2: Histopathologic and immunohistochemical features of *ELOC* deleted renal carcinoamas. The highlighted samples are those with biallelic inactivations on *ELOC*.

Sample	Fibromuscular bands	Nodular configuration	Architectural patterns	ccPRCC-like nuclear arrangement	Clear cells with voluminous cytoplasm and prominent borders	CA-IX	CK7	CD10	TFE3	HMB-45	34BET12
1T	+	+	Cys, SA	-	+	100	0	0	0	0	0
2T	+	+	SA	-	+	100	0	0	0	0	0
3T	+	+	SA	-	+	100	0	100	0	0	0
4T	+	+	SA, Cys	-	+	100	0	90	0	0	0
5T	+	+	SA, Cys, TU/BTU	-	+	10	0	100	0	0	0
6T	+	+	SA, Cys	-	+	100	20	90	0	0	0
8T	+	+	Cys, SA, TU/BTU	-	+	100	10	90	0	0	0
9T	+	+	SA, Cys	-	+	100	5	50	0	0	0
10T	+	+	SA, Cys	-	+	100	single cells	100	0	0	0
12T	+	+	SA, Cys	-	+	100	single cells	100	0	0	0
13T	+	+	SA, Cys, TU/BTU, Pap	-	+	100	0	90	0	0	0
14T	+	+	SA	-	+	100	0	10	0	0	0
16T	+	+	Cys, SA	-	+	100	10	90	0	0	0
17T	+	+	SA, Cys	-	+	100	30	90	0	0	0
19T	+	+	SA	-	+	100	0	100	0	0	0
7T	**+**	**+**	**SA, Cys**	**-**	**+**	**100**	**0**	**100**	**0**	**0**	**0**
15T	**+**	**+**	**TU/BTU, Cys, Pap**	**-**	**+**	**100**	**90**	**90**	**0**	**0**	**0**
18T	**+**	**+**	**TU/BTU, SA**	**-**	**+**	**100**	**90**	**100**	**0**	**0**	**0**

Staining scores are given as percentage of cells; - = negative ; + = positive; Cys = Cystic; SA = solid alveolar; TU/BTU = tubular/branching-tubular ; pap = papillary

Identification of ELOC mutations and their structural consequences.

All 18 samples were sent for whole-exome sequencing to allow the identification of classical biallelically inactivated ELOC ccRCC with no *VHL* inactivation whilst also granting the ability to locate ccRCC with both billalleic *ELOC* inactivations and *VHL* aberrations. 8 genes were identified to contain significant mutations ($q < 0.1$) when accounting for the background mutation rates, mutation clusters and site conservation using Mutsig2CV[36] (Figure 3A). Of the 8 genes, mutations in *VHL* and *PBRM1* are the most prevalent as is common in ccRCC with two tumours experiencing mutations in both genes. Three samples, 15T, 7T and 18T, contain *ELOC* mutations. 18T and 15T experience an A>G missense mutation at position 236 resulting in the Y79C substitution frequently observed in ELOC ccRCC. Novel *ELOC* mutations are found in sample 7T which include both an A>G missense mutation at position 238 and a conservative inframe insertion at position 230-235 resulting in a K80E amino acid substitution at position 80 and a Phe_Thr duplication at positions 77-78. Using a systematic structure-function analysis Takagi et al. identified two regions within the 112 residue of Elongin C central to its role in pVHL binding, specifically residues 71-80 and 92-111[29]. The mutations identified in our cohort impact residues 77-80 falling within these pVHL binding regions (Figure 3B).

We assess the impact of the *ELOC* mutations identified in our cohort on the pVHL-ELOC complex using a reference wild-type structure (PDB accession: 1VCB; the pVHL-ELOC complex is shown in Figure 3C) [37]. The HADDOCK refinement tool [27] is applied to each state we identify in our cohort namely the wild-type structure, a structure carrying a T79C substitution and another carrying both the F77_Y78 duplication and the K80E substitution produced using SWISS-MODEL (see methods). We believe the application of the refinement tool provides structure estimates more reflective of the true *in vivo* confirmation compared to a direct sidechain substitution seen in previous work as the change in the physical confirmation caused by the side chain alteration is accounted for. VdWs energy, electrostatic energy and the desolvation energy are calculated for each structure along with the buried surface area (BSA).

Extracting energy terms and the BSA from the models provided for each ELOC state shows F77_T78dup + K80E models possess higher VdWs energies ($\mu^{F77\_T78dup+K80E:VdW} = -64.05$ kcal mol^{-1}) and lower BSAs ($\mu^{F77_T78dup+K80E:BSA} = 2112.48$ Å^2) compared to both the wild-type ($\mu^{WT:VdW} = -73.60$ kcal mol^{-1}, $q = 1.36 \times 10^{-7}$; $\mu^{WT:BSA} = 2217.22$ Å^2, $q = 9.86 \times 10^{-7}$) and Y79C ($\mu^{Y79C:VdW} = -73.52$ kcal mol^{-1}, $q = 3.59 \times 10^{-6}$; $\mu^{Y79C:BSA} = 2192.41$ Å^2, $q = 7.78 \times 10^{-5}$) structures (Figure 3D-E). Both F77_T78dup + K80E ($\mu^{F77\_T78dup+K80E:Desol} = -15.35$ kcal mol^{-1}) and Y79C models ($\mu^{Y79C:Desol} = -14.41$ kcal mol^{-1}) have higher desolvation energies compared to the wild-type models ($\mu^{WT:Desol} = 16.79$ kcal mol^{-1}; q-value = 0.028 vs F77_T78dup + K80E; q-value = 0.023 vs Y79C; Figure S1). Here, the desolvation energy is the energy required for the dispersion of H_2O molecules necessary for the interaction between ELOC and pVHL to take place. The fall in buried surface area, measured by the difference between the surface area of the two proteins in complex and the surface area of both proteins independently, suggests a reduction in the interaction interface between pVHL and the F77_T78dup + K80E ELOC mutant structure. We measure no significant differences between the wild-type ELOC structure and Y79C mutant with respect to their BSA or VdW and electrostatic energies. With the knowledge that alterations of Y79 disrupt ELOC-pVHL interaction [29] and given the frequency with which Y79C substitution are observed in ELOC ccRCC[12], we visualise the effects of the Y79C substitution and assess the change in side chain interactions between ELOC and pVHL against the wild-type reference. Within the wild-type pVHL-ELOC complex, contacts are identified

between Y79 of ELOC and V155 of pVHL, however, the replacement of the Tyrosine, a large aromatic residue, with a smaller Cysteine leads to the absence of this interaction (Figure 3C).

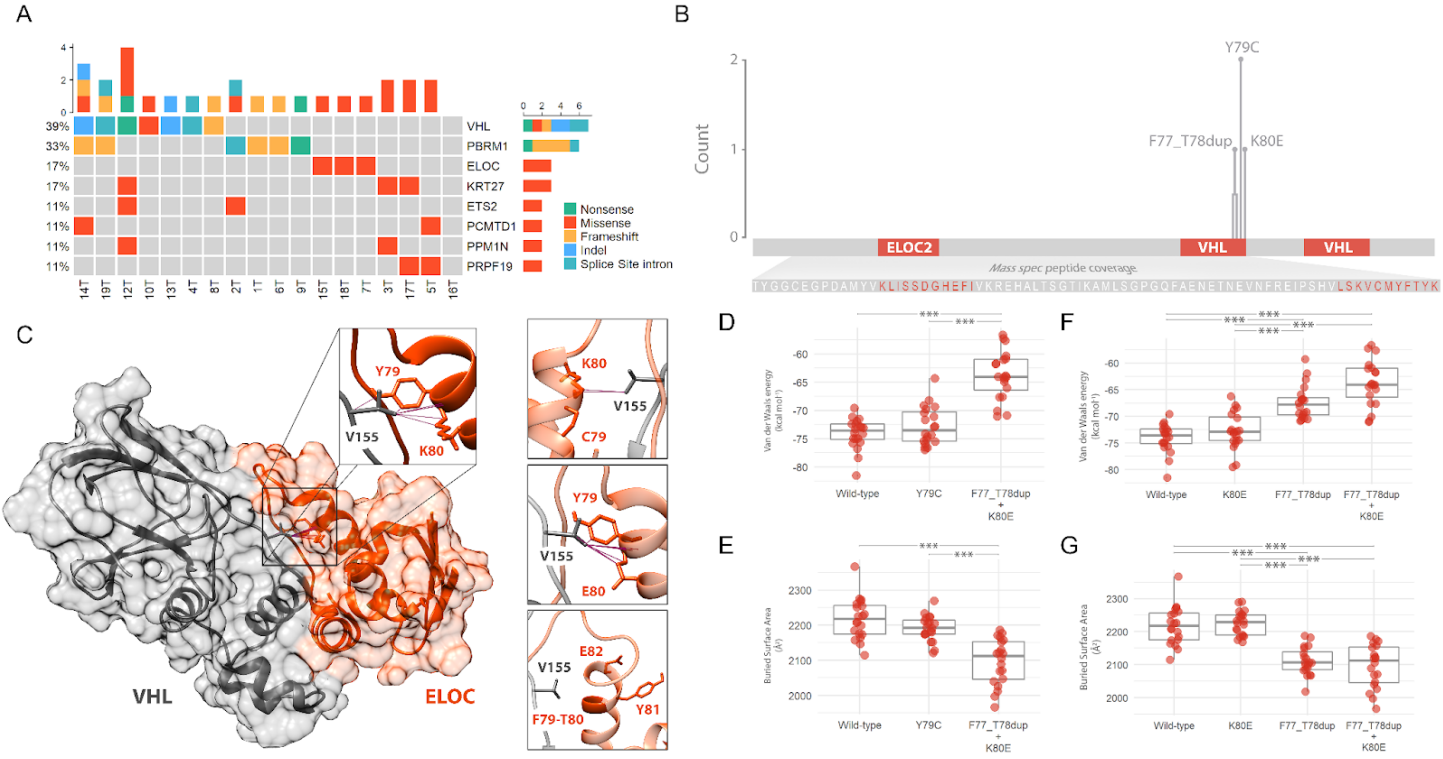

Figure 3: A) Summary of the mutations identified in our cohort of 18 ccRCC with monoallelic *ELOC* deletions. **B)** Positions and counts of *ELOC* mutations identified in the cohort, ELOC residues essential for the interaction of ELOC with pVHL and ELOB as identified by Takagi et al. (orange), and the sequence coverage of the 7 peptides identified during mass spectrometry; the identified peptides cover 74 of the 112 amino acids forming ELOC. **C)** Structure of the pVHL-ELOC complex showing the interaction of V155 on pVHL with Y79 and K80 on ELOC in the wild-type state (left). The structural impact of both the Y79C and K80E substitutions and the F77_T78dup + K80E mutation on their interaction with V155 (right top, middle, and bottom, respectively). Purple lines indicate contacts between atoms with Van der Waal (VdW) overlaps ≥ -1.3 Å. The VdW energy and buried surface areas (BSA) for the ELOC wild-type, Y79C and F77_T78dup + K80E models are shown in **D** & **E**. The VdW and BSA data following the assessment of the models containing the F77_T78dup and K80E mutations alone and in combination are shown in **F** and **G** (*** = q ≤ 0.001).

The F77_Y78dup + K80E ELOC mutant consists of two events: a duplication and a substitution. We assess the impact of each of these alone and in combination as identified in the tumour to determine if both events are necessary to alter ELOC-pVHL interaction. There is no difference between the models containing K80E and the wild-type models with respect to their VdW, electrostatic and desolvation energies or the BSA. Likewise, we also find no difference between the models containing the F77_Y78 duplication alone and those with both the F77_Y78 duplication and K80E substitution. We do, in contrast, find that models containing the F77_Y78 duplication (with and without K80E) have higher VdW and desolvation energies along with a lower BSA than both the wild-type models and those containing the K80E mutation alone (Figure 3F-G & Table S2 for means and q-values). These data suggest the F77_Y78 duplication alone may be sufficient to disrupt the pVHL-ELOC interaction. This is further supported by and reflected in the visualisation of these alterations (Figure 3C). The

interaction between V155 and residue 80 of ELOC is not disrupted whether it be a Lysine or a Glutamic acid residue. The duplication, however, inserts an additional rotation in the α-helix in the interface between pVHL and ELOC preventing the interaction of both Y79 and K80/E80 with V155 on pVHL.

Clinico-pathologic features of ELOC-mutated RCC

The male to female ratio was 2:1 of patients with *ELOC*-mutations. The age of the patients ranged from 54 to 78 years. Both male patients were still alive following their last follow up after 221 and 36 months after surgery (Patient 15T and 18T respectively; Table 1). The third patient was a female (Patient 7T). She died 13 months after surgery, but it is unclear whether the cause of death was related to the carcinoma. Two *ELOC*-mutated RCC were characterised by a high percentage of CK7 positive cells (Table 2 and Figure 4). Samples from patient 7T and 18T both presented with solid alveolar architecture with 7T also possessing cystic features. The tumour from patient 15T also possesses cystic features along with papillary and tubular/branching-tubular structures; the latter is also seen in patient 18T. The stage and grade differed between the three patients. Patient 7T possessed a stage 1b grade 3 tumour, patient 15T possessed a stage 1b and grade 2 tumour, and patient 18T possessed a stage 3a and grade 3 tumour. Both patients 7T and 15T suffered from additional malignancies. Patient 7T suffered from pancreatic cancer with liver metastasis and patient 15T went on to suffer from both prostate cancer and melanoma.

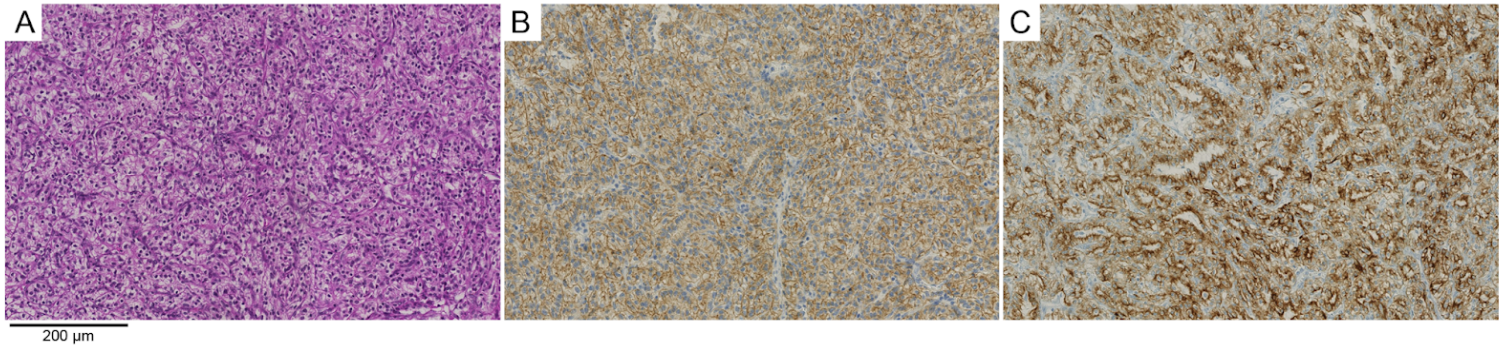

Figure 4: Histology and immunohistochemistry of an *ELOC*-mutated renal cell carcinoma. **(A)** Tumour cells with scanty clear cell cytoplasm and low nuclear grade This morphology is unusual for a clear cell renal cell carcinoma (Hematoxylin & Eosin staining). Strong **(B)** CAIX and **(C)** CK7 expression shown by immunohistochemistry.

Downstream impact of ELOC deletions

To assess both the impact of *ELOC* deletions on ELOC expression and the downstream effect of the biallelic inactivation of *ELOC* on the classical ccRCC pathway of HIF stabilisation, we investigated the protein landscape of *ELOC*-deleted ccRCC. Mass spectrometry was performed on the 18 samples with *ELOC* deletions and 31 samples in which *ELOC* remained intact. We identify seven unique peptides specific for ELOC in our analysis covering 66% of its amino acid sequence including the ELOB binding region and one of the two pVHL binding regions (Figure 4B). Of the 31 samples with no *ELOC* aberrations 20 were classical billellic *VHL* inactivated ccRCC. We see a significant reduction in ELOC protein expression in the three samples containing both an *ELOC* deletion and mutation in comparison to samples with intact *ELOC* ($q = 1.55 \times 10^{-11}$). Those samples with a biallelic inactivation of *ELOC* also have a significant reduction in ELOC protein expression compared to those with chromosomal *ELOC* deletions alone ($q = 0.008$; Figure 5A).

A characteristic feature of classical *VHL* inactivated ccRCC is the overexpression of CAIX, which is also used as a diagnostic marker for *VHL* inactivated ccRCC. Both CAIX and VEGFA overexpression is the result of HIF-α stabilisation and translocation into the nucleus upon which there is an overexpression of hypoxia-associated genes. We assess CAIX and VEGFA expression in biallelically *ELOC* inactivated ccRCC in comparison to those with chromosomal *ELOC* deletion and those with only a biallelic inactivation of *VHL*. We find no significant difference between CAIX ($p = 0.563$; Figure 5B) and VEGFA ($p = 0.845$; Figure 5C) expression levels between these three groups on the protein level, suggesting the downstream effects of *ELOC* inactivation are similar to those of *VHL* inactivation resulting in ccRCC.

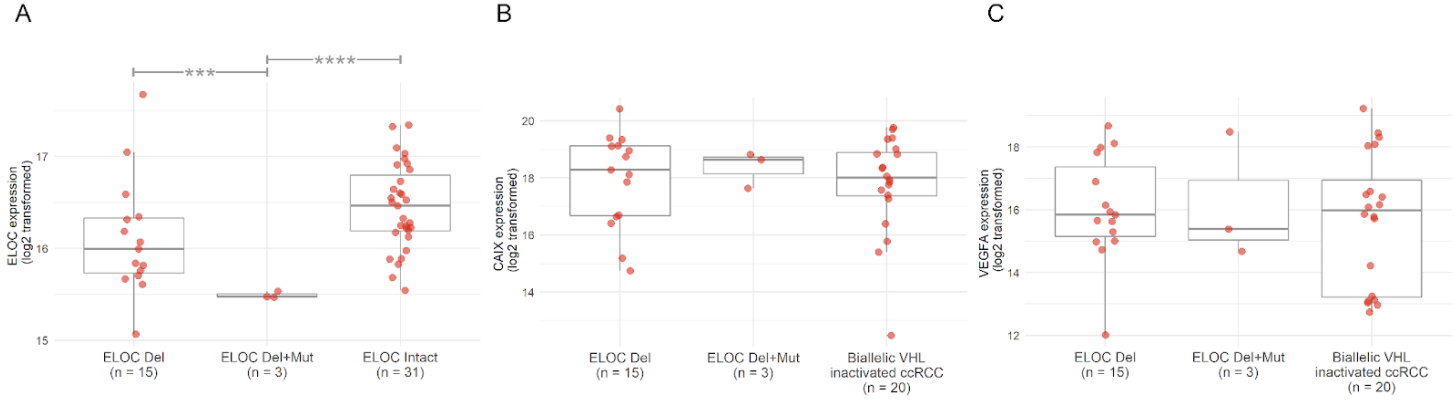

Figure 5: **A)** ELOC protein expression measured by mass spectrometry in renal cell carcinomas with monoallelic chromosomal *ELOC* deletion (n = 15), *ELOC* deletion and mutation (n=3) and those without *ELOC* aberration (n = 31). Renal cell carcinomas with biallelic inactivation of *ELOC* have a significantly lower ELOC expression compared to those with *ELOC* intact and those with only a monoallelic *ELOC* deletion (*** = $q \leq 0.001$ and **** = $q \leq 0.0001$; Welch's one-way ANOVA). No significant difference was found between classical *VHL* inactivated ccRCC (n = 20) and those with *ELOC* aberrations for either **B)** CAIX protein expression or **C)** VEGFA protein expression.

Discussion

In this study, we characterise 18 RCC with chromosomal *ELOC* deletions of which three can be regarded as *ELOC*-mutated RCC possessing biallelic inactivations of *ELOC*. We describe a novel ELOC mutation variant and combinations of *ELOC* mutations with *VHL* aberrations. We also show that the biallelic inactivation of *ELOC* results in the reduction in ELOC protein expression compared to those with intact *ELOC* and those with a monoallelic chromosomal *ELOC* deletion. Finally, we show that the impact of *ELOC* inactivation on the protein expression of CAIX and VEGFA is similar to ccRCC containing biallelic *VHL* inactivations.

We identified and characterised three new *ELOC*-mutated RCC. In 2022 *ELOC*-mutated RCC has been endorsed by the WHO as a novel renal tumour entity harbouring mutations in the *ELOC* (*TCEB1*) gene at chromosome 8q21.1; these mutations are accompanied by loss of the second allele, typically by the loss of chromosome 8 in its entirety [10,12,38]. With this new WHO classification, a transition into a molecular era of renal tumour classification has started [39]. Molecularly defined epithelial renal tumours include SMARCB1-deficient medullary RCC, TFEB-rearranged RCC, and Alk-rearranged RCC. These molecularly defined RCC subtypes are extremely rare and cannot be diagnosed by morphology alone. To date, 17 *ELOC*-mutated RCC have been identified, none of which contain co-occurring aberrations of *VHL* [10,12]. Known ELOC variants have all been identified within pVHL binding domains affecting three residues: A100, I95, and Y79, with 15 of the 17 known ELOC ccRCC samples containing Y79 variants [10,12]. Two of the ELOC ccRCC here possess the frequent Y79C alteration found in 10 out of the 17 previously identified ELOC tumours. However, we additionally describe a novel ELOC variant containing a duplication of the Phenylalanine and Tyrosine residues at positions 77 and 78 respectively. Similar to previously identified ELOC variants, this novel duplication lies within the pVHL binding domain known to be essential for the interaction of ELOC with pVHL and formation of the pVHL-ELOC-ELOB complex as identified by Tekagi et al. [29]. Structurally, like DiNatale et al., we show the frequent Y79C variant disrupts the pVHL:V155-ELOC:Y79 interaction. The F77_T78 duplication indirectly prevents the same pVHL:V155-Y79 interaction from taking place. Several alterations impacting the V155 residue in pVHL have been previously reported in those suffering from von Hippel-Lindau syndrome Type 1 and sufferers of ccRCC (data from the VHLdb[40]). A V155L substitution is the most common variant observed at position 155 [40] and was also identified in patients within the USZ Tissue Biobank. This suggests the V155 residue plays a key role in pVHL function and the V155-Y79 interaction may be essential for the recruitment of pVHL to the pVHL-ELOC-ELOB complex.

Two of the *ELOC*-mutated RCC have co-occurring monoallelic *VHL* aberrations (chromosomal *VHL* deletion or a hypermethylated *VHL* promoter region). The proportion of wild-type *VHL* tumours found here, i.e., 20.7% of ccRCCs with no *VHL* mutations, lies within the expected range also found in other large cohort studies [9]. During an assessment of the evolutionary history of ccRCC, mutations in *PBRM1*, *SETD2*, *BAP1*, and *PTEN* together with single *VHL* inactivations were shown to promote ccRCC tumourigenesis [7]. However, the two monoallelic *VHL* inactivated ccRCC with biallelic inactivation of *ELOC* do not possess mutations in these genes suggesting the biallelic inactivation of *ELOC* is driving tumorigenesis, justifying the classification as a separate tumour entity. Until now *VHL* aberrations have not been found to co-occur with *ELOC* mutations in ccRCC but we dispel the notion that these aberrations are perfectly mutually exclusive. However, whether and to which degree

ELOC and *VHL* mutations are mutually exclusive remains to be elucidated with a larger cohort of *ELOC* mutated samples. Further, samples to assess the effect of the monoallelic inactivation of *VHL* and *ELOC* together with respect to the tumorigenesis of ccRCC and the clinical repercussions of this combination is lacking. Thus far tumours with both biallelic inactivation of *VHL* and *ELOC* have not been found suggesting that the complete inactivation of one of these tumour suppressor genes is mutually exclusive with the other.

While *ELOC*-mutated RCC are rare, we show quantitatively the reduction in ELOC expression on the protein level in biallelic *ELOC* inactivated ccRCC samples, extending the work by Hakimi et al., which showed the reduction in ELOC expression on the transcriptomic level [38]. Notably, we identify a relatively high number of unique ELOC specific peptides resulting in a high sequence coverage and in particular a reliable quantitation accuracy. We determine that *ELOC* inactivated tumours have similar CAIX and VEGFA expression levels as tumours with the biallelic inactivation of *VHL* suggesting the downstream effects of *ELOC* inactivation are comparable to those of *VHL* inactivation in ccRCC. The inactivation of ELOC prevents the recruitment of pVHL to the E3-ubiquitin ligase complex, which, similarly to *VHL* inactivation, prevents the ubiquitination of HIF-α resulting in its stabilisation and the overexpression of downstream HIF targets. Our data suggests that while the tumour suppressor gene affected is different, i.e., *VHL* vs *ELOC*, the downstream signaling cascade of these tumours is extremely similar[10]. With comparable levels of VEGFA expression within these samples, the VEGFA targeting treatment strategies for ccRCC may also be effective in *ELOC*-mutated RCC.

ELOC-mutated RCC show overlapping features with current and emerging RCC entities such as ccRCC, renal cell carcinoma with leiomyomatous stroma (RCCLS) and clear cell papillary renal cell carcinoma (ccpRCC) [1]. In our cohort, *ELOC*-mutated RCC with biallelic *ELOC*-inactivation showed no specific histological architecture, but a very characteristic diffuse CK7 expression. This is in contrast to ccRCC with monoallelic *ELOC* deletion in which only one tumour possesses weak CK7 expression. A previous morphological assessment undertaken by Hakimi et al. showed thick fibromuscular bands with diffuse CAIX staining with CK7 positivity in 11 biallelically inactivated *ELOC*-mutated RCC [38]. However, this association is far from specific as both fibromuscular bands and CK7 positivity have been observed in tumours in the absence of *ELOC* mutations [41,42]. Further, *ELOC* mutations have also been identified in tumours without prominent fibromuscular bands and CK7 positivity [43,44].

Besides *ELOC*-inactivated RCC, renal tumours with leiomyomatous stroma [43,45] show morphologic (clear cells) and immunohistochemical (CA-IX and CD10 positivity) features, which are similar to those seen in ccRCC[46]. A closer examination of renal tumours with leiomyomatous stroma revealed intact *VHL* and four tumours possessing monosomy 8 including two harbouring an *ELOC* mutation suggesting some overlap with *ELOC*-mutated RCC[44,45]. Clear cell papillary RCC are characterised by the presence of CAIX positive tumours with strong CK7 staining but are also CD10 negative. Although we find strong CAIX and CK7 positivity in two of the *ELOC*-mutated RCC, we also observe a high number of CD10 positive tumours with only one showing evidence of papillary structure not too dissimilar from the observations amongst those tumours without *ELOC* mutations. Whether or not renal tumours with leiomyomatous stroma and *ELOC* inactivated RCC represent distinct tumour entities of ccRCC with seperate initiating molecular events is still under debate.

There have been attempts to classify *ELOC*-mutated RCC as a less aggressive tumour entity than *VHL*-inactivated ccRCC. However, Like Hakimi et al. [38] we find tumours with *ELOC*-mutation lack mutations in PBRM1, SETD2 and BAP1; the absence of these are indicative of a more aggressive form of ccRCC. Additionally, the vast majority of *ELOC*-mutated RCC are found to be low-stage and low-grade as is seen in both the Sato et al. [10] and DiNatale et al. [12] cohorts. Among our patients, two of the three *ELOC*-mutated cases possess pT stage 1b and ISUP grade 2 tumours. While one of these patients, who was elderly, passed away and contained a monoallelic *VHL* deletion, the other was still alive 19.5 years following their surgery possessing biallelic *ELOC*-inactivation and no *VHL* aberrations or mutations in *PBRM1*, *SETD2* and *BAP1*. The final patient with an *ELOC*-mutation possessed a pT stage 3a and ISUP grade 3 tumour, however, this patient also contained a hypermethylated promoter region leading to a monoallelic inactivation of *VHL*.

Conclusions

The combination of chromosomal *ELOC* deletion and *ELOC* mutations affecting key pVHL-ELOC binding domains will lead to the reduction of ELOC protein expression and result in similar molecular perturbations as in ccRCC containing a biallelic inactivation of *VHL*; functioning as a classical tumour suppressor gene. While monoallelic *VHL* and *ELOC* aberrations are not mutually exclusive, biallelic inactivation of either *VHL* or *ELOC* define different tumour subtypes. The combination of both in a single tumour has not been found pointing toward their mutually exclusivity.

Acknowledgments

Protein structure images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081). The proteomics work here is supported by the Lotte und Adolf Hotz-Sprenger Stiftung. The authors thank Susanne Dettwiler and Fabiola Prutek from the Tissue Biobank USZ for aiding the procurement of tissue material and IHC stainings.

Author contributions

HM, PS and NB supervised the project. HM obtained funding and was responsible for the project design. HM and BS reviewed IHC stainings. AAB and PS selected the cohort. AAB and DR obtained the proteomics and clinical data. AAB gathered the WES, epigenetic and CNV data. AAB carried out the data analysis and visualisation. AAB, PS, DR and HM drafted the manuscript. Review of the manuscript was carried out by all authors.

Bibliography

1. Moch, H., Cubilla, A. L., Humphrey, P. A., Reuter, V. E. & Ulbright, T. M. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours. *Eur. Urol.* **70**, 93–105 (2016).
2. Laguna, M. P. Re: international variations and trends in renal cell carcinoma incidence and mortality. *J. Urol.* **194**, 950–951 (2015).

3. Chen, M. *et al.* Genome-wide profiling of chromosomal alterations in renal cell carcinoma using high-density single nucleotide polymorphism arrays. *Int. J. Cancer* **125**, 2342–2348 (2009).
4. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
5. Maxwell, P. H. *et al.* The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271–275 (1999).
6. Semenza, G. L. HIF-1 mediates metabolic responses to intratumoral hypoxia and oncogenic mutations. *J. Clin. Invest.* **123**, 3664–3671 (2013).
7. Turajlic, S. *et al.* Deterministic evolutionary trajectories influence primary tumor growth: tracerx renal. *Cell* **173**, 595-610.e11 (2018).
8. Clark, D. J. *et al.* Integrated proteogenomic characterization of clear cell renal cell carcinoma. *Cell* **180**, 207 (2020).
9. Batavia, A. A., Schraml, P. & Moch, H. Clear cell renal cell carcinoma with wild-type von Hippel-Lindau gene: a non-existent or new tumour entity? *Histopathology* **74**, 60–67 (2019).
10. Sato, Y. *et al.* Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.* **45**, 860–867 (2013).
11. International Agency for Research on Cancer. *WHO Classification of Tumours ; Urinary and male genital tumours. 5th ed*. (2022).
12. DiNatale, R. G. *et al.* Putative Drivers of Aggressiveness in TCEB1-mutant Renal Cell Carcinoma: An Emerging Entity with Variable Clinical Course. *Eur. Urol. Focus* **7**, 381–389 (2021).
13. Moore, A. L. *et al.* Spatial distribution of private gene mutations in clear cell renal cell carcinoma. *Cancers (Basel)* **13**, (2021).
14. Rechsteiner, M. P. *et al.* VHL gene mutations and their effects on hypoxia inducible factor HIFα: identification of potential driver and passenger mutations. *Cancer Res.* **71**, 5500–5511 (2011).
15. Razafinjatovo, C. *et al.* Characterization of VHL missense mutations in sporadic clear cell renal cell carcinoma: hotspots, affected binding domains, functional impact on pVHL and therapeutic relevance. *BMC Cancer* **16**, 638 (2016).
16. Jung, H.-S., Lefferts, J. A. & Tsongalis, G. J. Utilization of the oncoscan microarray assay in cancer diagnostics. *Appl. Cancer Res.* **37**, 1 (2017).
17. Bolck, H. A. *et al.* Tracing Clonal Dynamics Reveals that Two- and Three-dimensional Patient-derived Cell Models Capture Tumor Heterogeneity of Clear Cell Renal Cell Carcinoma. *Eur. Urol. Focus* **7**, 152–162 (2021).
18. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci USA* **107**, 16910–16915 (2010).
19. Ricketts, C. J. *et al.* Methylation profiling and evaluation of demethylating therapy in renal cell carcinoma. *Clin. Epigenetics* **5**, 16 (2013).
20. Singer, J. *et al.* NGS-pipe: a flexible, easily extendable and highly configurable framework for NGS analysis. *Bioinformatics* **34**, 107–108 (2018).
21. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

22. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
23. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* (2013).
24. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
25. Broadinstitute. Picard. *Picard* https://broadinstitute.github.io/picard/.
26. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
27. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).
28. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
29. Takagi, Y., Pause, A., Conaway, R. C. & Conaway, J. W. Identification of elongin C sequences required for interaction with the von Hippel-Lindau tumor suppressor protein. *J. Biol. Chem.* **272**, 27444–27449 (1997).
30. Rodrigues, J. P. G. L. M., Teixeira, J. M. C., Trellet, M. & Bonvin, A. M. J. J. pdb-tools: a swiss army knife for molecular structures. [version 1; peer review: 2 approved]. *F1000Res.* **7**, 1961 (2018).
31. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
32. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).
33. Escher, C. *et al.* Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **12**, 1111–1121 (2012).
34. Bruderer, R. *et al.* Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Mol. Cell. Proteomics* **16**, 2296–2309 (2017).
35. Hortin, G. L., Sviridov, D. & Anderson, N. L. High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance. *Clin. Chem.* **54**, 1608–1616 (2008).
36. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
37. Stebbins, C. E., Kaelin, W. G. & Pavletich, N. P. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science* **284**, 455–461 (1999).
38. Hakimi, A. A. et al. TCEB1-mutated renal cell carcinoma: a distinct genomic and morphological subtype. Mod. Pathol. 28, 845–853 (2015).
39. Pfister, S. M. *et al.* A Summary of the Inaugural WHO Classification of Pediatric Tumors: Transitioning from the Optical into the Molecular Era. *Cancer Discov.* **12**, 331–355 (2022).

40. Tabaro, F. *et al.* VHLdb: A database of von Hippel-Lindau protein interactors and mutations. *Sci. Rep.* **6**, 31128 (2016).
41. Favazza, L. *et al.* Renal cell tumors with clear cell histology and intact VHL and chromosome 3p: a histological review of tumors from the Cancer Genome Atlas database. *Mod. Pathol.* **30**, 1603–1612 (2017).
42. Hes, O., Compérat, E. M. & Rioux-Leclercq, N. Clear cell papillary renal cell carcinoma, renal angiomyoadenomatous tumor, and renal cell carcinoma with leiomyomatous stroma relationship of 3 types of renal tumors: a review. *Ann. Diagn. Pathol.* **21**, 59–64 (2016).
43. Parilla, M. *et al.* Genetic underpinnings of renal cell carcinoma with leiomyomatous stroma. *Am. J. Surg. Pathol.* **43**, 1135–1144 (2019).
44. Shah, R. B. *et al.* "Renal Cell Carcinoma With Leiomyomatous Stroma" Harbor Somatic Mutations of TSC1, TSC2, MTOR, and/or ELOC (TCEB1): Clinicopathologic and Molecular Characterization of 18 Sporadic Tumors Supports a Distinct Entity. *Am. J. Surg. Pathol.* **44**, 571–581 (2020).
45. Petersson, F. *et al.* Renal Cell Carcinoma With Leiomyomatous Stroma: A Group of Tumors With Indistinguishable Histopathologic Features, But 2 Distinct Genetic Profiles: Next-Generation Sequencing Analysis of 6 Cases Negative for Aberrations Related to the VHL gene. *Appl. Immunohistochem. Mol. Morphol.* **26**, 192–197 (2018).
46. Udager, A. M. & Mehra, R. Morphologic, molecular, and taxonomic evolution of renal cell carcinoma: A conceptual perspective with emphasis on updates to the 2016 world health organization classification. *Arch. Pathol. Lab. Med.* **140**, 1026–1037 (2016).

Consequences of *ELOC* alterations in renal cancer

Aashil A. Batavia[1,2,3], Dorothea Rutishauser[1], Bettina Sobottka[1], Peter Schraml[1], Niko Beerenwinkel[2,3] and Holger Moch[1]*

[1]*Department of Pathology and Molecular Pathology, University and University Hospital Zurich, Schmelzbergstrasse 12, 8091 Zurich,Switzerland*

[2]*Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, 4058 Basel*

[3]*SIB Swiss Institute of Bioinformatics, Mattenstrasse 26, 4058 Basel, Switzerland*

* Corresponding author

Running title: RCC with *ELOC* and *VHL* alterations

Conflicts of interest: The authors declare no potential conflicts of interest

Supplementary data

Figures S1-S2
Table S1-S2

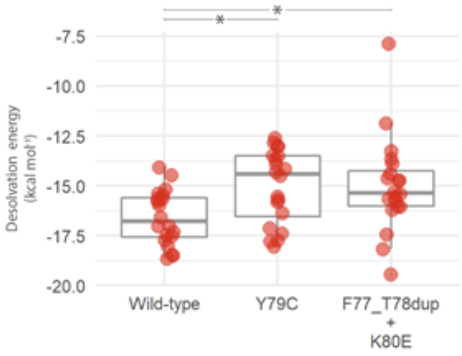

Figure S1: The desolvation energies of wild-type, Y79C and F77_T78dup + K80E models. A significant increase in desolvation energy is observed in both mutant structures in comparison to the wild-type ELOC structure (* = $q \leq 0.05$; n = 20 for each group)

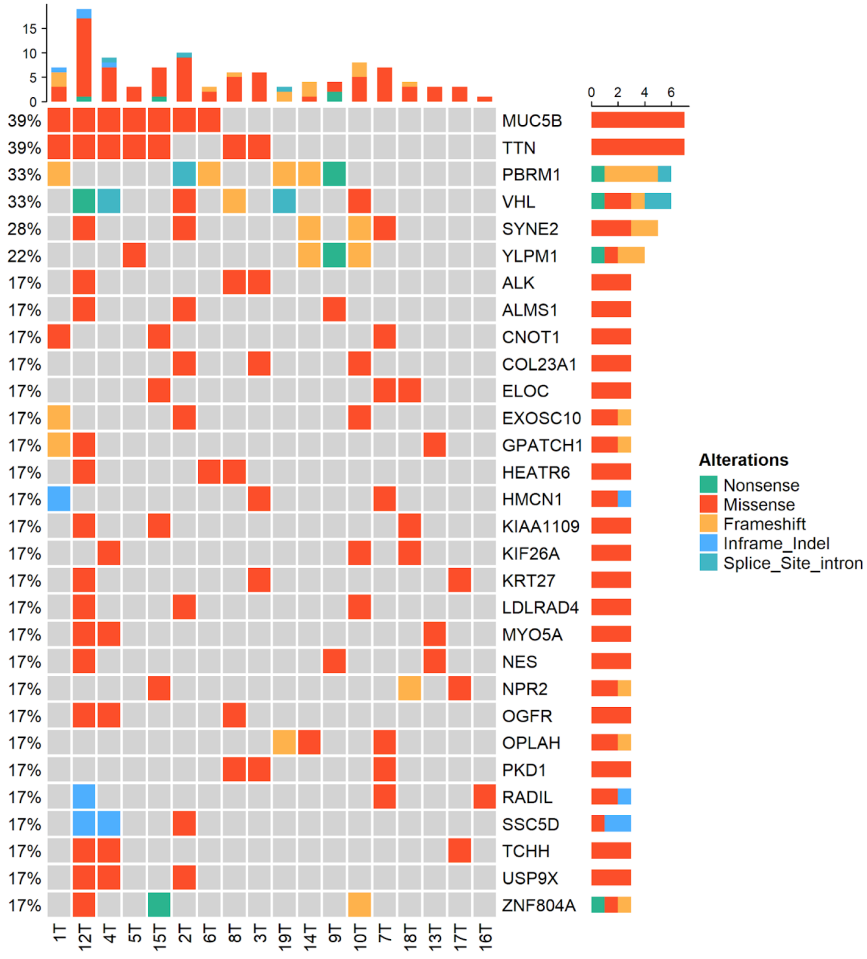

Figure S2: Mutation summary of genes which have been found to be altered in a minimum of three out of the 18 *ELOC* deleted samples.

Table S1: Antibodies used for the immunohistochemical assessment of the FFPE sections.

Protein	Clone	Dilution	Species	Manufacturer
CAIX	polyclonal	1:3000	Rabbit Polyclonal	Abcam Limited
CK7	OV-TL 12/30	1:100	Mouse Monoclonal	Dako A/S
CD10	56C6	1:25	Mouse Monoclonal	Novocastra Laboratories Ltd.
TFE3	MRQ-37	Prediluted	Rabbit Monoclonal	Cell Marque Lifescreen Ltd.
Pmel17	HMB-45	1:50	Mouse Monoclonal	DAKO A/S
Cytokeratin	34BetaE12	1:40	Mouse Monoclonal	DAKO A/S

Table S2: Properties of ELOC structures containing both the F77_T78dup and K80E events as well as structures containing each event individually. The means of each property across models for each structure is listed on the left. All q-values for the pairwise comparisons for each measured property is also shown (right)

Means	
$\mu^{WT:BSA}$	2217.22 Å^2
$\mu^{WT:VdW}$	-73.60 kcal mol^{-1}
$\mu^{WT:Desol}$	-16.79 kcal mol^{-1}
$\mu^{K80E:BSA}$	2228.42 Å^2
$\mu^{K80E:VdW}$	-72.89 kcal mol^{-1}
$\mu^{K80E:Desol}$	-17.07 kcal mol^{-1}
$\mu^{F77\_T78dup:BSA}$	2106.31 Å^2
$\mu^{F77\_T78dup:VdW}$	-67.80 kcal mol^{-1}
$\mu^{F77\_T78dup:Desol}$	-14.81 kcal mol^{-1}
$\mu^{F77\_T78dup+K80E:BSA}$	2112.48 Å^2
$\mu^{F77\_T78dup+K80E:VdW}$	-64.05 kcal mol^{-1}
$\mu^{F77\_T78dup+K80E:Desol}$	-15.35 kcal mol^{-1}

VdW	WT	K80E	F77_T78dup
WT			
K80E	n.s		
F77_T78dup	8.19×10^{-6}	4.66×10^{-4}	
F77_T78dup+K80E	3.89×10^{-9}	6.73×10^{-7}	n.s

BSA	WT	K80E	F77_T78dup
WT			
K80E	n.s		
F77_T78dup	5.76×10^{-6}	6.26×10^{-7}	
F77_T78dup+K80E	3.69×10^{-6}	5.54×10^{-7}	n.s

Desol	WT	K80E	F77_T78dup
WT			
K80E	n.s		
F77_T78dup	0.001	3.28×10^{-4}	
F77_T78dup+K80E	0.02	0.006	n.s

III

Wild-type von Hippel-Lindau ccRCC

Whilst ELOC RCC accounts for a small number of ^{wt}*VHL* ccRCC tumours, there remains a significant portion for which the aetiology and pathways for progression are unknown. In this chapter, I focus on those tumours diagnosed as ccRCC with no *VHL* inactivations. ^{wt}*VHL* ccRCC are often disregarded in large cohort ccRCC studies contributing to a knowledge gap in scientific literature, however, the few mentions of ^{wt}*VHL* ccRCC point toward a more aggressive tumour population[8,9]. I identify ^{wt}*VHL* ccRCC in two independent cohorts taken from The Cancer Genome Atlas and the University Hospital Zurich's Tissue Biobank. Using data assessing multiple omic layers allows us to identify processes central for ^{wt}*VHL* ccRCC function and progression, even in the low sample sizes expected here. What follows is a manuscript presenting my work. My contribution was as follows: identifying the cohorts, obtaining, integrating and visualising all the data, all statistical testing, production of all visualisations and writing the manuscript.

[8] Dagher, J. *et al.* Wild-type VHL Clear Cell Renal Cell Carcinomas Are a Distinct Clinical and Histologic Entity: A 10-Year Follow-up. *Eur. Urol. Focus* **1**, 284–290 (2016).

[9] Turajlic, S. *et al.* Deterministic evolutionary trajectories influence primary tumor growth: tracerx renal. *Cell* **173**, 595-610.e11 (2018).

Wild-type *von Hippel-Lindau* clear cell renal cell carcinoma confer a poorer prognosis due to increases in proliferative and migratory properties

Aashil A. Batavia[1,2,3], Dorothea Rutishauser[1], Peter Schraml[1], Niko Beerenwinkel[2,3] and Holger Moch[1]*

[1]*Department of Pathology and Molecular Pathology, University and University Hospital Zurich, Schmelzbergstrasse 12, 8091 Zurich,Switzerland*

[2]*Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, 4058 Basel*

[3]*SIB Swiss Institute of Bioinformatics, Mattenstrasse 26, 4058 Basel, Switzerland*

* Corresponding author

Abstract

Biallelic inactivation of the *von Hippel-Lindau* (*VHL*) gene is the molecular hallmark of clear cell renal cell carcinoma (ccRCC). However, 5-12% of ccRCC possess no *VHL* aberrations yet present with identical morphological phenotypes. We call these tumours wild-type *VHL* ccRCC (^{wt}*VHL* ccRCC). Here, using both publicly available data and our own cohort from the University Hospital Zurich (USZ), we characterise ^{wt}*VHL* ccRCC using genetic, epigenetic, transcriptomic, proteomic and clinical data in order to identify distinguishing pathological processes in comparison to *VHL* inactivated ccRCC, as well as indicators of their aetiology. From TCGA public data we obtain 8 ^{wt}*VHL* ccRCC (Wt), 144 biallelic *VHL* inactivated ccRCC (Bidel), and 89 monoallelic VHL inactivated ccRCC (Monodel). From our cohort at the USZ, we identify 11 Wt, 84 Bidel, and 32 Monodel tumours. We find Wt tumours confer poorer survival in comparison to Bidel and Monodel tumours. Wt tumours have very heterogeneous mutational landscapes and do not possess the mutations frequently found in classical ccRCC. Genes promoting epithelial-mesenchymal transition (EMT) are enriched within Wt samples in both the transcriptome and proteome. Using a directed functional interaction network, we integrate genomic, epigenomic and transcriptomic data in order to identify mediator genes orchestrating Wt-specific processes, identifying factors orchestrating cell division and promoting proliferation including *CCNB1*. We also identify potential biomarkers for ^{wt}*VHL* ccRCC e.g. IGF2 combining both transcriptomic and proteomic data. Finally, we describe HMGA1 as a key regulator in ^{wt}*VHL* ccRCC promoting both proliferation and migration in Wt tumours. Although Wt samples possess considerable overlap with Bidel samples, the upregulation of EMT and cell proliferation distinguishes them as a more aggressive ccRCC subgroup.

Introduction

Clear cell renal cell carcinoma (ccRCC) forms approximately 80% of all renal cell carcinoma (RCC), making them the most common histological subtype of RCC[1]. The molecular hallmark of ccRCC is the biallelic inactivation of the von Hippel-Lindau gene (*VHL*) gene. This biallelic inactivation can occur via a combination of chromosome 3p deletion, *VHL* promoter hypermethylation and/or inactivating *VHL* mutations. Protein VHL (pVHL), the protein product of *VHL*, is a key component of the E3 ubiquitin ligase complex essential for the ubiquitin-mediated degradation of HIF-α[2]. The absence of pVHL causes an accumulation of HIF-α resulting in the overexpression of hypoxia associated genes such as *CA9* and *VEGFA*, the latter of which is the target of antiangiogenic RCC therapies[3]. Several large cohort studies have described in great detail the complex molecular landscape of ccRCC[4–6] with significant intratumor heterogeneity[7–10] and single or combinatorial events leading to tumourigenesis[6]. Notably, alongside *VHL*, the frequent pathogenomic chromosome 3p (chr3p) deletion causes the allelic imbalance of a further 3 frequently mutated genes in ccRCC: *PBRM1*, *BAP1* and *SETD2*, propagating tumour development. Concurrently with chr3p deletion, simultaneous chromosome 5q amplification has been observed with high frequency[11]. Other frequent chromosome aberrations include losses of chromosome 1, 6q, 8p, 9 and 14q and gains of 7q, 12 and 20 demonstrating the genomic instability present in ccRCC.

As aberrations of *VHL* are observed in ~90% of ccRCC[12] and are considered truncal with respect to tumour evolution[6], the assessments of the molecular features of these tumours often disregard the approximate 10% of cases possessing no VHL inactivations. We call ccRCC with no *VHL* inactivations wild-type *VHL* ccRCC (*^{wt}VHL* ccRCC). The diagnosis and treatment of ccRCC is driven by its characteristic clear cell histology[3]. *^{wt}VHL* ccRCC are histologically identical to ccRCC with *VHL* inactivation and are therefore diagnosed and treated in an identical manner. The absence of a comprehensive analysis of the molecular characteristics of *^{wt}VHL* ccRCC allows the risk of putting these patients through unnecessary and ineffective treatments while also potentially missing strategies from which they may benefit. The limited efforts to study *^{wt}VHL* ccRCC suggest these tumours possess a more aggressive phenotype conferring poorer prognosis[13]. Increased rates of metastasis, sarcomatoid component, lymphocyte infiltration and higher tumour grades have all been associated with *^{wt}VHL* ccRCC[6,13]. Furthermore, an assessment of the tumour evolution of ccRCC identified the presence of a distinct evolutionary trajectory for *^{wt}VHL* ccRCC associated with higher proliferation rates and increased genomic instability[6]. However, a thorough assessment of the molecular landscape within the *^{wt}VHL* ccRCC identified in their cohort was not conducted.

A subgroup of *^{wt}VHL* ccRCC was found to carry biallelic inactivations of the *ELOC* gene[5]. *ELOC*'s protein product, Elongin C, is a binding partner of pVHL within the E3 ubiquitin ligase complex. Within *ELOC* RCC, monoallelic deletions of *ELOC* are accompanied by mutations disrupting its interaction with pVHL[5,14,15](Batavia et al; Chapter 2) potentially preventing the recruitment of pVHL to the E3 ubiquitin ligase complex. This disruption in ELOC-pVHL binding together with a reduction in Elongin C expression, elevated HIF-α expression and similar VEGFA and CAIX expression to classical ccRCC suggest a similar path to tumourigenesis in ELOC RCC[5] (Batavia et al; Chapter 2). However, while the biallelic inactivation of *ELOC* provides an explanation for tumourigenesis in a small fraction of *^{wt}VHL* ccRCC (only 20 ELOC RCC with clear cell histology have been identified

worldwide), there remains a significant portion of ^{wt}*VHL* ccRCC for which no explanation is offered as to their origin and increased aggressiveness.

Here we aim to address the lack of a molecular phenotypic assessment of ^{wt}*VHL* ccRCC by using both a publicly available multi-omic dataset and our own patient cohort from the Tissue Biobank of the University Hospital Zurich (USZ). We identify 8 ^{wt}*VHL* ccRCC from TCGA and 11 within the USZ. Given the rarity of ^{wt}*VHL* ccRCC, the sample sizes found here fall within the expected frequencies of these tumours. We find ^{wt}*VHL* ccRCC confer poorer survival rates compared to those tumours with *VHL* inactivation, potentially due to the upregulation of factors promoting EMT and cell proliferation. We also identify candidates to act as ^{wt}*VHL* ccRCC biomarkers in order to aid the identification of these rare malignancies.

Methods

Acquiring ccRCC data from TCGA

The KIRC dataset from TCGA contains 537 ccRCC primary tumour samples. We obtained tumour and normal mutation (Mutation annotation files), CNV (GISTIC2[16]), methylation (HumanMethylation450), mRNA expression, miRNA expression and clinical data for these samples using the Broad Institute's GDAC FireBrowse platform[17] and TCGABiolinks[18].

Determining *VHL* status in ccRCC samples from TCGA

Identification of ^{wt}*VHL* ccRCC required the sample specific assessment of mutations, copy number aberrations and *VHL* promoter methylation. The KIRC dataset is composed of 537 samples of which 242 samples contain the datasets needed to assess the three modes of *VHL* inactivation forming the cohort used here. Mutation annotation files (MAF) are combined into a single gene × sample matrix. We consider missense, nonsense, and splice-site mutations along with in frame deletions and frameshift indels as inactivating mutations. Copy number data was accessed in the form of GISTIC output files containing gene-wise copy number changes per sample. Illumina Human Methylation 450 array data was obtained from TCGABiolinks[18]. This dataset consists of probe β-values spanning the genome including promoter regions, CpG islands and key differentially methylated sites in multiple tumours in comparison to normal samples. The gene × sample matrices for both mutations and CNVs were used to determine the *VHL* status with respect to *VHL* mutations and chromosomal deletions. To assess the promoter hypermethylation of *VHL* we first identify a 636 base pair (bp) CpG promoter island beginning 226 bp upstream of *VHL's* first exon[19]. We identify probes attributed to *VHL* targeting this CpG island for which β-values are provided. A β-value for a given site is the ratio of the methylated probe intensity and the sum of the methylated and unmethylated probe intensities. The median β-value was taken as a representation of the methylation status of a sample's *VHL* promoter region. 160 matched normal samples were present in the methylation dataset for which median β-values were also calculated. A threshold of 0.058 was determined to call the methylation status of each sample since:

$$Threshold = (\sigma \times 3.09) + \bar{x}$$

where σ and $\bar{x}$ are the standard deviation and mean of the median β-values for normal samples while 3.09 is the t-value corresponding to a p-value of 0.001. We considered a sample unmethylated when

the median β-value is less than 0.058. We apply a Gaussian mixture model to the data to confirm the *VHL* methylation status of the samples considered non-methylated (Figure S1).

With each of the three *VHL* inactivation methods assessed we divide the 242 samples across three categories: Biallelically *VHL* inactivated ccRCC (Bidel), Monoallelic *VHL* inactivated ccRCC (Monodel) and ^{wt}*VHL* ccRCC samples (Wt). One ELOC inactivated ccRCC was not considered in subsequent analysis.

TCGA clinical data and survival comparisons

From the clinical data available from TCGA we consider the "Years to Birth", "Vital Status", "Overall Survival (OS)", "Days to death", "Days to last follow up", "Pathology T-stage", "Gender" and "Race". For the survival statistics we take the "Days to death" or "Days to last follow up" for each sample given the vital status of the patient to determine overall survival. We apply a Cox proportional hazards model to assess survival between the groups separated by their *VHL* aberration status. We also conduct a multivariate assessment of survival accounting for age, pathological T-stage, gender, race and ethnicity. The pathological T-stage was reduced to four categories, collapsing pT2a and pT2b into pT2 and pT3a, pT3b, pT3c into pT3.

Differential mRNA and miRNA expression

Messenger RNA (mRNA) counts were obtained from TCGA in order to determine differentially expressed genes using EdgeR[20]. Of the 241 samples from the TCGA, mRNA expression data was not available for one of the Bidel samples leaving 240 samples from which differentially mRNA expression can be assessed. We also gather mRNA expression data from 72 normal samples. Differentially expressed genes are determined between the three groups (Bidel, Monodel, Wt) and against normal samples. We apply a quasi-likelihood F-test[20] to determine differentially expressed genes providing a more robust and reliable error rate control; necessary when low sample sizes are present. Comparisons between tumours groups and also against normal samples allows us to identify genes perturbed and contributing to tumour-specific processes. We call genes differentially expressed if they possess a q-value < 0.05 and an absolute Log Fold Change (LFC) of more than 1, the latter is not considered for the enrichment analysis. Gene Set Enrichment Analysis was carried out using FGSEA via ClusterProfiler[21] and the hallmark gene set from MSigDB[22]. As we use the QLF test to determine differentially expressed genes we rank genes by LFC during the enrichment analysis filtering out those which possess q-values less than 0.05. Differential expression of miRNA was determined similarly to mRNA. 120 of the 241 samples contained miRNA data (n = 75 Bidel, n = 6 Wt and n = 39 Monodel) and we determined differential expression between each group and against all 71 normal samples. For visualisation, we convert normalised expression values for each sample to Z-scores using the mean and standard deviation of normal samples.

Identifying differentially methylated sites

Differential methylation analysis was carried out using TCGABiolinks. We obtain genome-wide methylation data from the Illumina Human Methylation 450 platform of the 241 samples from the TCGA-KIRC dataset in which the *VHL* status could be determined successfully. We consider all probes within Chromosomes 1-22 removing those with "NA" values. We are left with 378,586 probes

from which we determine differentially methylated sites between the Wt and Bidel groups. A Wilcoxon test is applied to identify differential methylation applying the Benjamini-Hochberg adjustment to correct for the FDR. We consider probes with |mean difference| > 0.15 and a q-value of less than 0.05 to be differentially methylated.

Identification of mediator genes via network based multi-omic integration.

Mediator genes orchestrating downstream differential expression given the upstream aberrations are identified via the Integration of mutation, CNV, methylation and mRNA expression data using NetICS[23]. The notion of directionality is provided using a directed network available in NetICS. For the Wt and Bidel groups all mutations (derived from the MAFs obtained for all samples), CNV (derived from the gene-wise CNV calls from GISTIC2) and significantly methylated genes for each sample are used as the aberrations. Sample specific differentially methylated genes were obtained for all genes present in the network. We apply a Wilcoxon rank sum test with BH correction comparing probe values for each gene in each tumour sample to probe values within normal samples. Those genes with a q-value less than 0.05 and an absolute β-value mean difference of more than 0.15 are selected for each sample. The output of NetICS is a gene list ranked by their associated mediator score representing the sum of the per-sample ranks. For tumour versus tumour comparisons, we normalise the mediator score by the sample size to obtain the average rank per-sample and identify the change in rank between the top 1000 ranked mediators for each group.

Cohort selection and *VHL* status of ccRCC from the USZ Tissue Biobank

Our previously described cohort of 464 ccRCC diagnosed at the University Hospital Zurich (USZ) are also used here (Batavia et al; Chapter 2). In brief, 464 formalin-fixed and paraffin-embedded (FFPE) ccRCC samples from the USZ's Department of Pathology and Molecular Pathology Tissue Biobank ranging from 1993-2019 are considered. Genomic DNA was extracted from normal and tumour FFPE samples with a minimum of 70% tumour cells using 0.6 mm punch needles. The *VHL* status was determined using Sanger sequencing (identifying *VHL* mutations), OncoScan® CNV FFPE Microarrays (identifying *VHL* deletions) and Bisulfite sequencing (identifying *VHL* promoter hypermethylation). Full details of the methods are described in Batavia et al; Chapter 2. Three samples were found to be of the rare ELOC RCC subtype and therefore removed from this study. In total, we possess 127 tumours from as many patients for which the *VHL* status was successfully determined for this study. As with the samples derived from TCGA, samples originating from the USZ are then placed into three categories: 'Bidel' containing those with biallelic *VHL* inactivations, 'Monodel' containing those with a monoallelic *VHL* inactivation and 'Wt' containing those with no *VHL* inactivations. Given the heterogeneity present in ccRCC and the high frequency with which *VHL* mutations are seen in ccRCC, we resequence all tumours found to possess no mode of *VHL* inactivation using an alternate tumour block; thereby, ensuring their categorization as ^{wt}*VHL* ccRCC.

Our retrospective study fulfilled the legal conditions according to the Swiss Law "Humanforschungsgesetz (HFG)", which, in exceptional cases, allows the use of biomaterial and patient data for research purposes without informed consent, if i) it is impossible or disproportionately difficult to obtain patient consent; ii) there is no documented refusal; iii) research interests prevail the individual interest of a patient. Law abidance of this study was reviewed and approved by the ethics commission of the Canton Zurich (BASEC_2019-01959).

Data acquisition and processing of HRM mass spectrometry proteomics

Mass spectrometry was conducted on 120 FFPE samples containing 60 tumour-normal pairs also used in Batavia et. al (Chapter 2). 0.6 mm punches were digested with trypsin. Peptides are spiked with index Retention Time (iRT) calibration peptides and then subject to HRM data-independent acquisition (DIA). The DIA method consisted of one full range MS1 scan and 29 DIA segments (MS2) and was adopted from Bruderer et al.[24]. A sample-specific spectral library was generated using shotgun LC-MS/MS on a Thermo Q-Exactive HF-X instrument. Detailed specification of the technologies, reagents and their dilution are given in Batavia et al.(Chapter 2). HMR mass spectrometry data were analysed using SpectronautTM Pulsar software (Biognosys AG). The false discovery rate (FDR) on the peptide and protein level was set to 1% and the data was filtered using row-based extraction. The data were normalised using global normalisation. For testing of differential protein abundance, MS1 and MS2 protein intensities were analysed using a two sample Student's t-test with FDR correction[25,26]. Proteins with a fold change greater than 1.5 ($|LFC| > 0.58$) and a q-value < 0.05 are considered differentially expressed. Enrichment analysis is conducted in a similar manner to the mRNA data from TCGA mentioned above.

Data Accessibility

The OncoScan SNP-array data generated for this study are publicly available in GEO via the accession number GSE201277. The mass spectrometry data is available on ProteomeXchange via PRIDE with the accession number PXD033291. The accessions of the samples used can be found in Additional File 1.

Results

From the 537 ccRCC samples in TCGA, we obtain 417 mutation profiles, 528 copy number profiles and 320 epigenetic profiles. We identify 242 samples possessing all three datasets to produce a cohort of ccRCC samples in which we can accurately call *VHL* status (Figure 1A). 128 samples contain *VHL* mutations considered to be inactivating (see methods) and 224 samples contain chromosomal *VHL* deletions. 27 tumours are found to be hypermethylated (median β-value > 0.058; Figure S1). The samples are then split into three groups: A) Biallelically deleted ccRCC (referred to as Bidel; N = 144) containing those tumours with *VHL* mutations and chromosomal deletions (n = 119), *VHL* chromosome deletion and promoter hypermethylation (n = 23) and tumours possessing *VHL* mutations, chromosomal deletion and promoter hypermethylation (n = 2). B) Monoallelically deleted ccRCC (referred to as Monodel; N = 89) containing those with solely a *VHL* mutation (n = 7), *VHL* chromosomal deletion (n = 80) or a hypermethylated *VHL* promoter region (n = 2). C) ^{wt}*VHL* ccRCC (referred to as Wt; N = 8) are those that do not possess any aberration in *VHL* (Figure 1B). One sample contained both an *ELOC* Mutation and *ELOC* deletion, we considered this tumour to be of the newly described ELOC RCC subtype and removed[5,14] (Batavia et al; Chapter 2). Of the 127 samples from the USZ, 84 are in the Bidel group, 32 in the Monodel group and 11 in the Wt group (Figure S2).

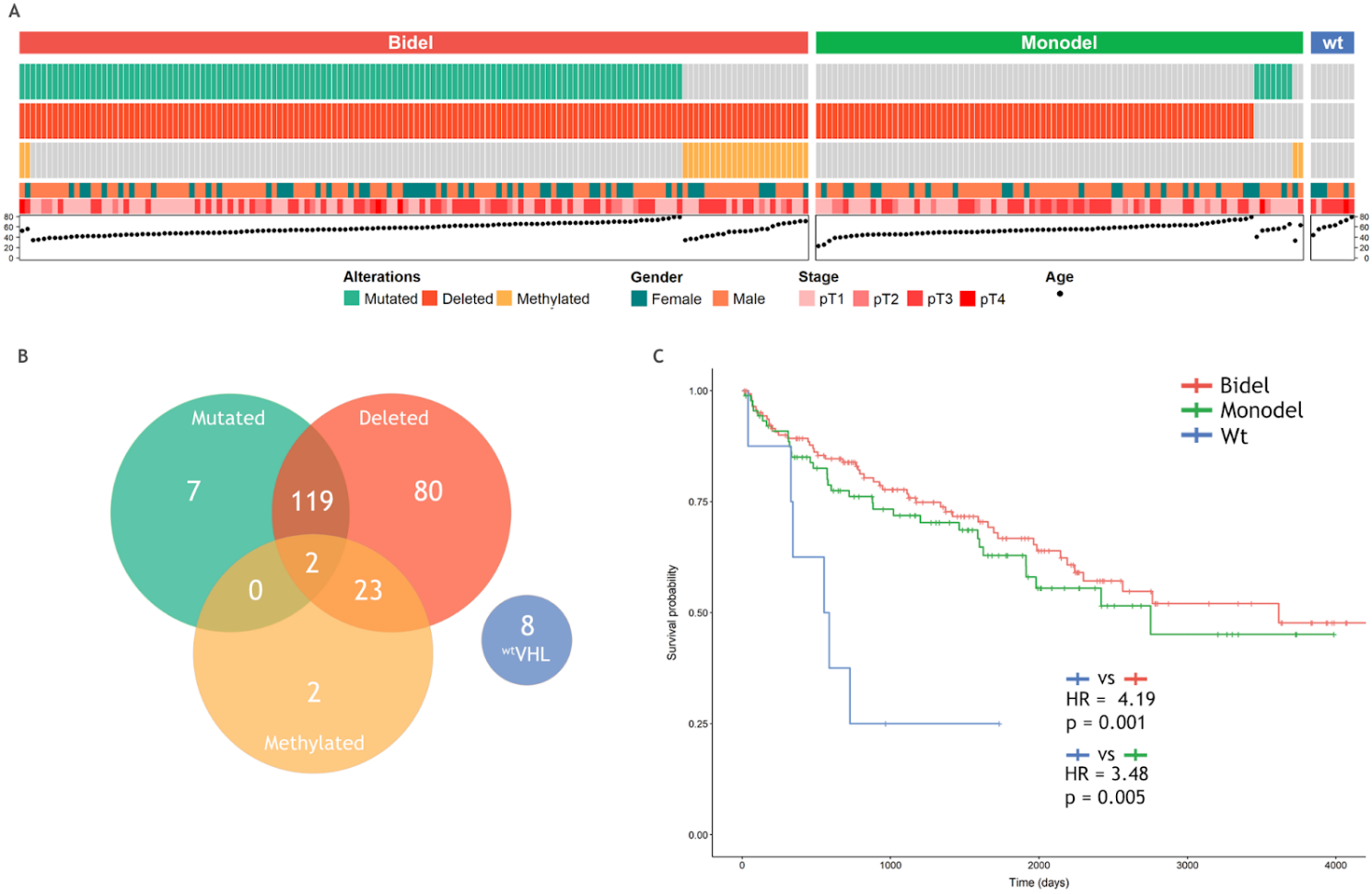

Figure 1: A) Summary of the *VHL* aberrations identified in the TCGA cohort along with the pT stage and age of each patient. Samples are divided into three groups according to their *VHL* status. **B)** Counts of each mode of *VHL* inactivation and the combinations of those aberrations. 8 samples were identified as ^{wt}*VHL* ccRCC. **C)** A Cox proportional hazards model is fitted showing differences in survival between the three groups. ^{wt}*VHL* ccRCC confers a poorer prognosis than those tumours with either a monoallelic or biallelic *VHL* inactivation. ^{wt}*VHL* status was also significantly associated with poorer survival in a multivariate analysis considering age and pathological stage (p = 0.04).

Wild-type *VHL* status confers a poorer prognosis in ccRCC

Having separated tumours by *VHL* status we assess the clinical composition of each group and assess the prognostic differences between them. The age distributions are similar between all three groups in TCGA cohort with the majority of cases residing within the 60-80 age group (Bidel = 52.1%, Monodel = 51.7%, Wt = 62.5%); this is also true for the Bidel and Monodel groups from the USZ (Bidel = 60.7%, Monodel = 56.3%). Within the USZ Wt group, 5 of the 11 (45.45%) patients were diagnosed between the ages of 40-59 while the remaining 6 ranged between 60-70. Incidences of renal tumours are higher in males than females[1]. This is reflected in both the USZ and TCGA cohorts where males are more prevalent than females. The distribution of the tumours with respect to the pathological T stage was also similar between the Bidel and Monodel groups. The majority of the tumours were stage pT1 and pT3 in both cohorts (Figure 1 and S2). Tumours at stage pT2 were present at low frequency, although pT4 were by far the least prevalent. There is a higher proportion of

late stage tumours in the Wt group obtained from TCGA with 75% being pT3 and 12.5% being pT4 tumours. There are no pT2 tumours in the Wt group and only one pT1 tumour.

Comparisons of the overall survival between the three TCGA groups show patients belonging to the Wt group have a significantly worse prognosis compared to those patients in both the Bidel group ($p = 0.001$; hazard ratio = 4.19) and the Monodel group ($p = 0.005$; hazard ratio = 3.48; Figure 1C). There was no significant difference between the patients within the Bidel and Monodel groups. As there are a larger proportion of advanced stage tumours in the Wt group, we carry out a multivariate analysis using all clinical parameters found to significantly impact patient survival which include *VHL* status, pathological T stage and age. *VHL* status is still found to significantly impact patient survival with wild-type *VHL* tumours conferring worse survival in comparison to the Bidel group ($p = 0.04$, Hazard Ratio = 2.52).

^{wt}*VHL* ccRCC have heterogeneous genomic landscapes distinct from classical ccRCC

The mutational landscape of classical *VHL* inactivated ccRCC is well characterised. Upon assessing the variants identified in our groups, we find that the Wt samples do not possess mutations in genes frequently appearing in ccRCC and found in both the Bidel and Monodel samples (Figure S3). *PBRM1*, *SETD2* and *BAP1*, which are amongst the most commonly aberrated genes in ccRCC, are mutated with a frequency of 41%, 11% and 10% in Bidel samples and 30%, 13% and 7% in Monodel samples. A solitary mutation was identified in TP53 when considering the commonly aberrated ccRCC genes in Wt samples. The intragroup assessment of the Wt samples revealed very heterogeneous mutational landscapes with few mutated genes being shared between them. The mean number of mutated genes identified in the Wt group was 43.6 but only 9 genes were found to be mutated in more than one sample; *CHD7*, *CLCN2*, *CSPG4*, *TRIOBP*, *PCDH15*, *DST*, *HGC6.3* and *DPCR1* were mutated in 2 samples and *MUC4* was mutated in 3. Contrary to the mutational landscape we observe few differences in the frequency of copy number aberrations (CNAs) between the Bidel and Wt groups. Notably, a larger proportion of the Wt group derived from within the USZ experience chr1p deletions (Figure S4). The Wt samples from the USZ lacked chr5q amplifications known to co-occur with chr3p deletion via chromothripsis and following mitotic catastrophe in classical ccRCC samples[11].

^{wt}*VHL* ccRCC specifically overexpress EMT and cell proliferation markers

Differential mRNA expression identified 4,655 genes differentially expressed in Bidel samples in comparison and 4,220 in Wt samples in comparison to normal (q-value < 0.05, |LFC| > 1). Gene set enrichment analysis shows a considerable overlap between the enriched gene sets in both the Bidel and Wt groups (Figure S5). Notably, hypoxia associated genes are enriched in both groups (q-value = 3.97×10^{-4} and 1.95×10^{-7} for Bidel and Wt respectively). This is expected in the case of the Bidel group as the inactivation of *VHL* results in the overactivity of HIF-α, however, the enrichment in the case of Wt samples indicates an alternative mechanism, independent of *VHL*, activating the hypoxia response in these tumours. Assessing the core enrichment of the hypoxia gene sets in both groups, i.e. the genes forming the leading-edge subset in the hypoxia gene set, found 5 genes specific for the Wt group: *ENO3*, *IL6*, *SRPX*, *TGFB3*, *KDELR3*. Notably we find a significantly lower VEGFA expression in Wt samples in comparison to Bidel ($q = 0.01$, LFC = -1.18).

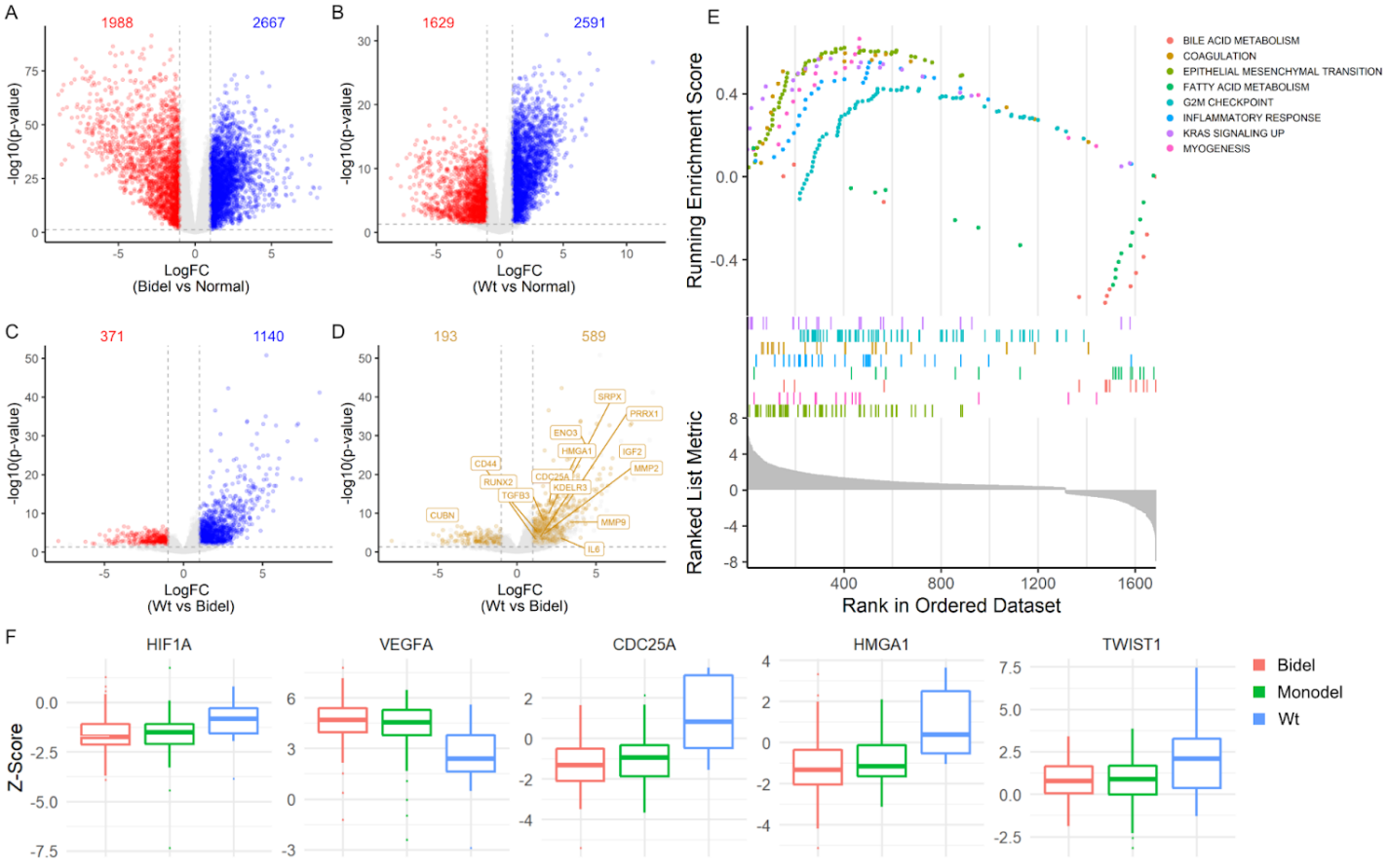

Figure 2: Differential mRNA expression between **A)** Bidel and normal, **B)** Wt and normal and **C)** Wt and Bidel groups. Genes over and under expressed are shown in blue and red respectively. Genes with an absolute LFC > 1 and a q-value < 0.05 are shown here. Genes highlighted in **D)** are those both overexpressed and underexpressed in Wt versus Bidel and normal samples. **E)** Gene set enrichment analysis of the 1686 genes specifically over- and underexpressed in Wt using the Mutsig2 Hallmark gene sets showing only those significantly enriched (q-value < 0.05). **F)** Messenger mRNA expression of selected genes across the three groups. Z-scores are produced using normal samples. *VEGFA* expression is significantly higher in all three samples versus normal but also significantly lower in Wt samples in comparison to the Bidel group. *CDC25A* and *HMGA1* are specifically overexpressed in Wt samples. *TWIST1* is significantly higher in the Wt group in comparison to Bidel (q-value < 0.05, |LFC| > 1). There is no significant difference in HIF-α expression.

Enrichment analysis of genes significantly (q-value < 0.05) over and underexpressed in the Wt group in comparison to both the Bidel and normal samples (n = 1,686) revealed the enrichment of genes associated with the G2M checkpoint pushing cells through cell division (q-value = 2.80×10^{-2}), those upregulated by KRAS activation (q-value = 2.80×10^{-2}), those involved in the inflammatory response (q-value = 2.10×10^{-2}), myogenesis (q-value = 1.80×10^{-2}), coagulation (q-value = 2.59×10^{-2}) and most significantly EMT (q-value = 3.13×10^{-6}) (Figure 2E). Amongst the core enrichment of the EMT gene set are *PRRX1* and *RUNX2* both of which are considered key transcription factors associated with driving this transition[27–30]. Furthermore, *PRRX1* expression in the Wt group experiences a LFC in expression greater than 4 against normal samples (adjusted p-value = 4.36×10^{-11} ; LFC = 3.09) and more than 3 versus the Bidel group (adjusted p-value = 9.26×10^{-11}; LFC = 3.09). The core enrichment of the G2M transition hallmark specific for the Wt group is made of 36 genes. Of these 36 genes, *CDC25A* possesses the most significant change in expression within Wt

samples when compared to the Bidel group. Further, we observe that *CDC25A* is specifically overexpressed in Wt samples (q-value = 4.79×10^{-5} and LFC = 1.21 vs normal; q-value = 2.66×10^{-11} and LFC = 2.02 vs Bidel) while within the Bidel group there is a slight decrease in expression versus normal tissue (q-value = 3.35×10^{-12} and LFC = -0.84 ; Figure S6). *CDC25A* is one of three main *CDC25* isoforms along with *CDC25B* and *CDC25C* [31]. CDC25A is a phosphatase thought to activate CDK2-cyclin E complexes playing a key role in cell cycle progression[32]. *HMGA1* expression was also identified amongst the leading edge subset of the G2M transition hallmark overexpressed in Wt samples specifically. *HMGA1* has been associated with the regulation of *CDC25A* via binding to its promoter region[33] and was found to be epigenetically altered in Bidel samples in comparison to Wt samples.

Epigenetic alterations and gene expression changes in ^{wt}*VHL* ccRCC

With the emergence of the CpG island methylator phenotype (CIMP) which is known to be associated with early onset and poor survival in RCC[34,35], we compare the difference in methylation levels between the three groups and found no difference between them (Figure S7). Epigenetic alterations of DNA are a known means of regulating gene expression. We identify 573 probes to be differentially methylated between Wt and Bidel samples. Our interest lies with epigenetic alterations resulting in changes in mRNA expression, therefore, we associate the differential methylation expression with genes differentially expressed in Wt samples in comparison to both normal and Bidel samples (Figure 3). 13 genes overexpressed in Wt samples were also found to contain hypomethylated sites within regulatory regions in comparison to the Bidel group, several of which promote and aid migration e.g. *TGF-β*[36–38] and *KRT17*[39]. Another one of these 13 genes was *HMGA1* which contained two significantly hypomethylated sites within its promoter region in Wt samples (q-value = 0.049 and 0.047), potentially being the cause of the expression difference observed in these tumours. Furthermore, CD44 which also contains a hypomethylated site within its promoter region (q-value = 0.048, mean difference = -0.18) is known to mediate migration and invasion of tumour cells via its interaction with matrix metalloproteinases (MMP), specifically *MMP9*[40,41] which was also specifically upregulated in the Wt group along with several other MMPs. *CD44* signalling has also been shown to increase the phosphorylation, and in turn activity, of *RUNX2*[42].

GIPC2 was found to contain a hypermethylated site (q-value = 0.048, mean difference = 0.26) and a reduction in gene expression specific to Wt samples (q-value = 8.40×10^{-4} and LFC = -2.47 vs normal; q-value = 6.02×10^{-5} and LFC = -2.31 vs Bidel). *GIPC2* has been described as a tumour suppressor, knockdown of which induces *HIF-a* upregulation and causes an increase in cell proliferation. Loss of *GIPC2* has been shown to be accompanied by loss of chromosome 1p, where it resides and hypermethylation of its promoter[43]. CUBN, another gene containing a hypermethylated site (meth: q-value = 0.049, mean difference = 0.32), has been shown to be a prognostic marker in ccRCC with low expression conferring a poorer survival as is seen in Wt samples[44,45]. As cubilin expression is largely specific to kidney tissue, the reduction in CUBN is suggestive of increased dedifferentiation within the Wt group supportive of more advanced EMT in these tumours.

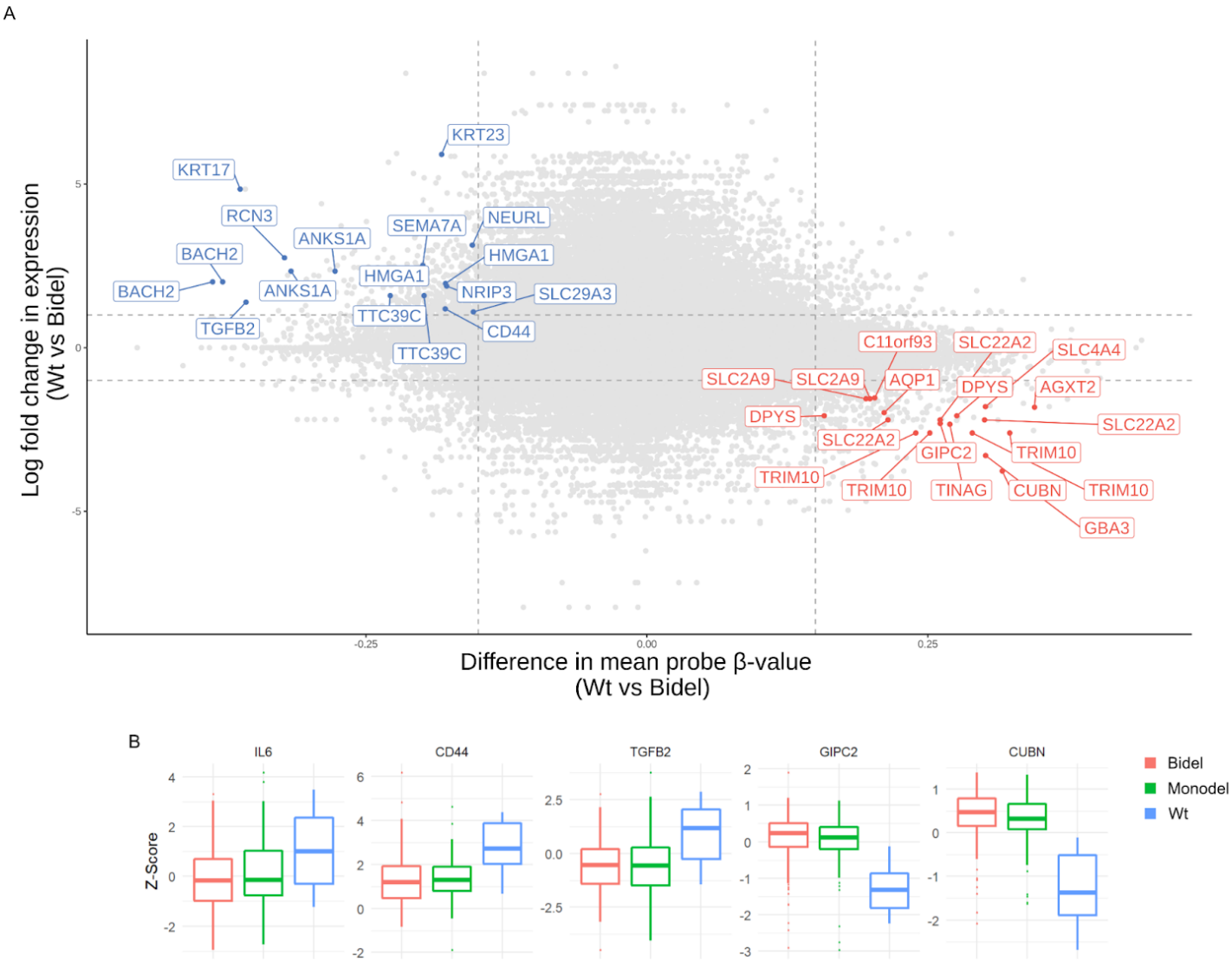

Figure 3: A) Association of epigenetic alterations to changes in the transcriptome. We identify the difference between the mean probe β-values and the log fold change in expression of the genes corresponding to each probe in the Wt group versus the Bidel group. Those probes hypomethylated in Wt versus Bidel and corresponding to genes specifically overexpressed in Wt (significantly overexpressed in Wt versus both the Bidel and normals samples) are shown in blue. Those probes hypermethylated in the Wt samples versus Bidel and corresponding to genes with reduced expression in the Wt group specifically are shown in red (differential methylation = |mean difference| > 0.15 & q-value < 0.05, differentially mRNA expression = |LFC| > 1 & q-value < 0.05). **B)** Messenger RNA expression of selected genes that are found to be both differentially methylated and differentially expressed in Wt samples in comparison to the Bidel samples.

Network based integration identifies key orchestrators of tumour processes

Via the integration of genomic, epigenomic and transcriptomic data through diffusion within a directed functional interaction network, we identify genes (termed mediators) which are hubs for orchestrating downstream differential expression changes given the upstream aberrations. This would also allow us to determine if the heterogeneous mutational landscape observed in *^{wt}VHL* ccRCC converges on specific processes. We apply NetICS using the aberrations identified in each group and the differentially expressed genes in each group in comparison to normal samples. There is considerable overlap between the top 1000 mediators identified for both Wt and Bidel samples (76.4% Wt mediators overlap with Bidel). Seven of the top 10 mediators in both groups are G-protein subunits essential for the initiation of intracellular signal transmission from extracellular stimuli emphasising the importance of tumour-ECM interaction in ccRCC. We identify several genes with a high average rank which have previously been associated with ccRCC processes e.g. *CTNNB1*[46], *PI3K*, *AKT1*[47] and *NFKB1*[48] (Additional File 2). However, In order to identify mediators central to orchestrating Wt specific downstream changes and therefore contributing to the more aggressive phenotype observed in these tumours, we run NetICS using those genes specifically over- or underexpressed in Wt samples in comparison to both normal and Bidel groups. We then identify those genes with the largest increase in mediator activity in comparison to bidel samples (Table S1). Amongst the highest ranked genes are those involved in both cell division and EMT, an example of this is the mediator role of CCNB1 (Figure S8). Fewer genetic and epigenetic aberrations are identified for Wt samples in genes upstream of CCNB1 in comparison to Bidel samples. Additionally while CCNB1 is itself overexpressed in Wt samples, several of its downstream interacting partners also experience significant increases in expression; most notably CDK1 which together with CCNB1 drive cells though G2/M progression and are essential during mitosis[49,50].

Proteomic assessment allows the identification of potential *^{wt}VHL* ccRCC biomarkers.

To validate findings from the assessment of TCGA data we turn to the proteomics assessment of our cohort from the USZ Tissue Biobank. HRM mass spectrometry data was obtained for all 11 ^{wt}VHL ccRCC and 43 Bidel samples from the USZ cohort with matched normal tissue. 7,343 proteins are identified within the entire cohort. Partial least squares discriminant analysis (PLS-DA) shows a clear separation between tumour and matched normal samples (Figure S9). We identify 2,566 differentially expressed proteins between the Wt group and matched normals and 2,675 between Bidel and matched normals. Enrichment analysis identified considerable overlap in enriched gene sets between the transcriptomic level assessment from TCGA samples and the proteomic level using samples from the USZ. Furthemore we find a similar enrichment of differentially expressed proteins in Wt samples as we do in Bidel samples when compared to matched normal tissue within known ccRCC (Figure S10). The enrichment of the *Mitotic spindle* (q-value = 0.01) and *Apical Junction* (q-value = 0.04) gene sets pertaining to genes involved in cell division and those required for cell adherence were amongst the 4 gene sets specifically enriched in the Wt samples.

We identify 1,174 differentially expressed proteins between the Wt and Bidel groups with 940 being overexpressed and 234 experiencing a reduction in expression in Wt samples (Figure 4A). As with the transcriptomic assessment of Wt tumours from TCGA we also identify a significant enrichment of EMT factors (q-value = 6.50×10^{-5}) and KRAS activation (q-value = 8.76×10^{-3}) in the Wt group in comparison to the Bidel group (Figure 4B). These properties further confirm the notion that Wt

tumours are more malignant and possess properties to further promote their survival and proliferation in comparison to the Bidel group, demonstrating that the upregulation of processes observed in the transcriptome is also identifiable in the Proteome. Similarly to the decrease in *VEGFA* mRNA expression within Wt samples from TCGA, we find a significant decrease in VEGFA protein expression in Wt samples from the USZ in comparison to the Bidel group ($q = 8.88 \times 10^{-5}$, LFC = 0.67).

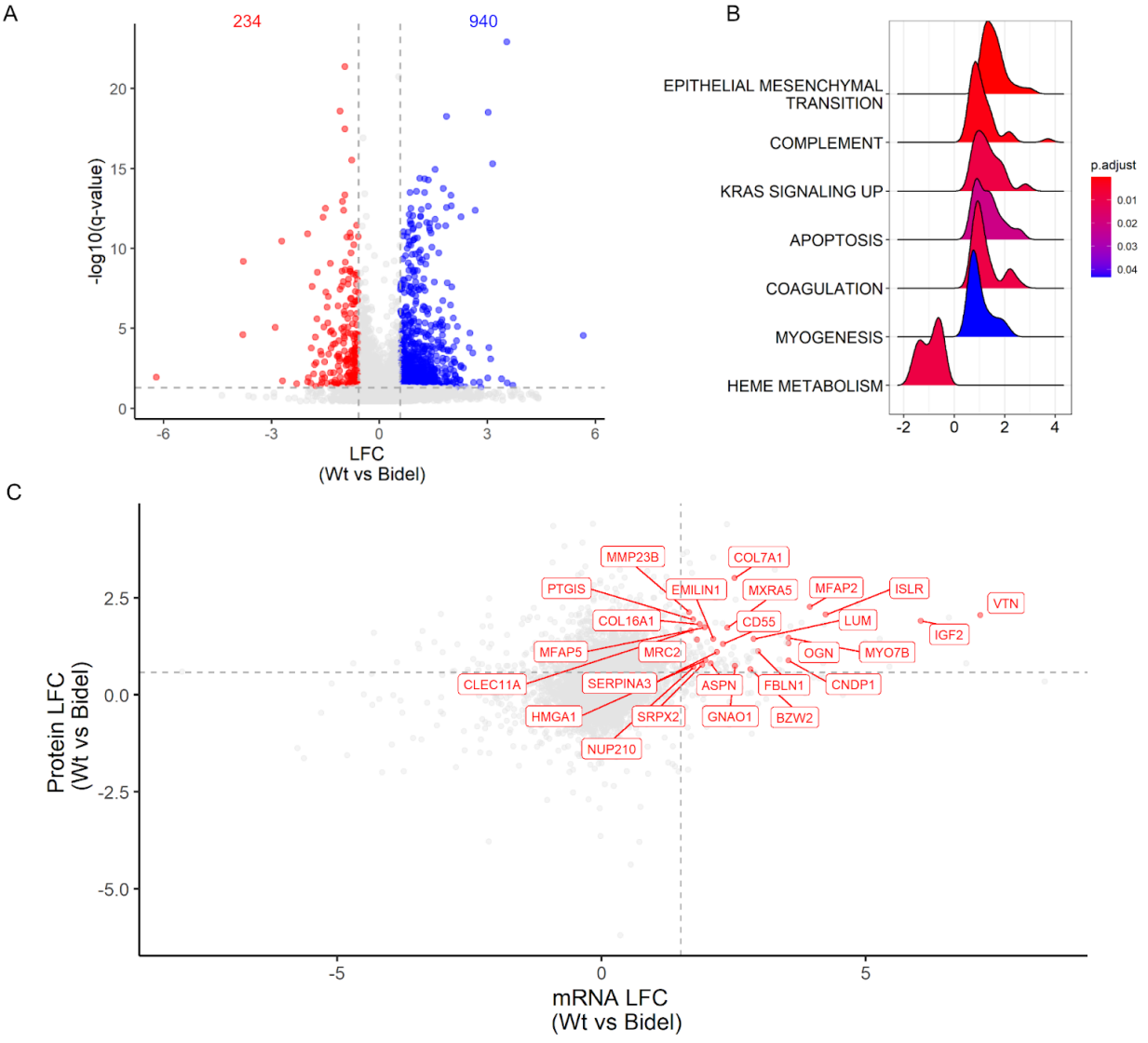

Figure 4: A) Differential proteomic analysis between Wt and Bidel samples identifies 940 proteins overexpressed (Blue) and 234 underexpressed (red) in Wt samples (q-value < 0.05, |L2R| > 0.58). **B)** Significantly enriched gene sets in Wt samples in comparison to the Bidel group. The enriched gene sets identified using the proteomics data from the USZ overlap with the enriched gene sets identified with Wt samples from the TCGA cohort using transcriptomic data. **C)** Identification of candidate biomarkers for ^{wt}*VHL* ccRCC. Highlighted genes are those that 1) are overexpressed in Wt samples across both the transcriptome in the TCGA cohort and the proteome in the USZ cohort (mRNA: q-value < 0.05 & LFC > 1.5 protein: q-value < 0.05 & FC > 1.5) and 2) remain after filtering out those genes overexpressed in the Bidel group within either the transcriptome (q-value < 0.05 & LFC > 0.5) or proteome (q-value < 0.05 & FC > 1.5) in comparison to normal samples.

As there are no histological differences between *^{Wt}VHL* ccRCC and *VHL* inactivated ccRCC, the identification of *^{Wt}VHL* ccRCC requires methylation, copy number and mutation data; given their rarity, this can be a costly and inefficient endeavour. The presence of a diagnostic biomarker to aid the identification of these tumours would greatly increase the efficiency with which these tumours are found. We identify genes for which both their transcript and protein product is overexpressed in Wt samples in comparison to Bidel tumours using more stringent criteria. We first identify genes experiencing both an overexpression of mRNA (q-value < 0.05 & LFC > 1.5) and protein (q-value < 0.05 & FC > 1.5) in Wt samples. From these genes we remove all those also found to be overexpressed in Bidel samples at either the transcriptomic or proteomic level in comparison to normal samples (mRNA thresholds: q-value < 0.05 & LFC > 0.5, protein thresholds: q-value < 0.05 & FC > 1.5). We find 26 potential markers specific for *^{wt}VHL* ccRCC in comparison to *VHL* inactivated ccRCC (Figure 4C). Of these 26 genes the largest fold changes are observed in VTN, IGF2, ISLR and MFAP2. VTN, MFAP2 and ISLR, often expressed in mesenchymal cells, all play a role in cell adhesion and interactions with the ECM[51–53]. IGF2, however, behaves as an autocrine growth factor promoting the transcription of *HIF-a*[54] which is suggestive of its potential role in causing the downstream changes resulting in the similar phenotypes as seen in Bidel samples which are hypoxia associated.

The role of HMGA1 in *^{wt}VHL* ccRCC

From these data, we identify High Mobility Group AT-Hook 1 (HMGA1) as playing a significant role in Wt samples. As mentioned earlier, HMGA1 is overexpressed on the transcriptome level in Wt samples from TCGA compared to both Bidel and normal samples. We find that the epigenetic differences in Wt samples may be the cause of this overexpression. Furthermore, we also find HMGA1 overexpression on the proteome level amongst the Wt group in comparison to the Bidel group using our samples from the USZ (q-value $=7.25 \times 10^{-4}$, LFC = 0.89). When we assess the mediator activity using NetICS we also find a large increase in mediator activity of HMGA1 within Wt samples with the average rank changing from 2498.5 in Bidel samples to 772.75 in Wt samples. Given these data, we assess the known downstream effects of HMGA1 and hypothesise the mechanism by which HMGA1 may be orchestrating the increase in cell migration and proliferation identified in Wt samples (Figure 5).

Expression of *CDC25A*[33], *Cyclin E1*[55] and the matrix metalloproteinases *MMP2* and *MMP9*[55] has been shown to be regulated by *HMGA1*. Along with *CDC25A* we also found *Cyclin E1* (*CCNE1*), *MMP2* and *MMP9* to be overexpressed (mRNA) in Wt samples in comparison to both Bidel and normal samples (q-value < 0.05, LFC > 1.5). From the differential miRNA analysis, we further identify *miR-221* and *miR-222* to be overexpressed specifically in Wt samples (q-value < 0.05, LFC > 1.5) (Figure 5). *CDC25A* has been shown to activate *CCNE1* during the cell cycle, amongst other cyclins, exemplifying the role of CDC25 phosphatases throughout the cell cycle[32]. The overexpression of *HMGA1* may cause the increase in *CDC25A* and *CCNE1* expression together with other factors, e.g. *CCNB1*, shown to be upregulated in Wt samples driving proliferation in these tumours. Additionally, by causing the overexpression of *miR-221* and *miR-222,* which targets TIMP3 mRNA's 3'-UTR preventing its regulation of *MMP2/9,* and by directly regulating the expression of *MMP2* and *MMP9*, *HMGA1* can also promote migratory properties within tumour cells.

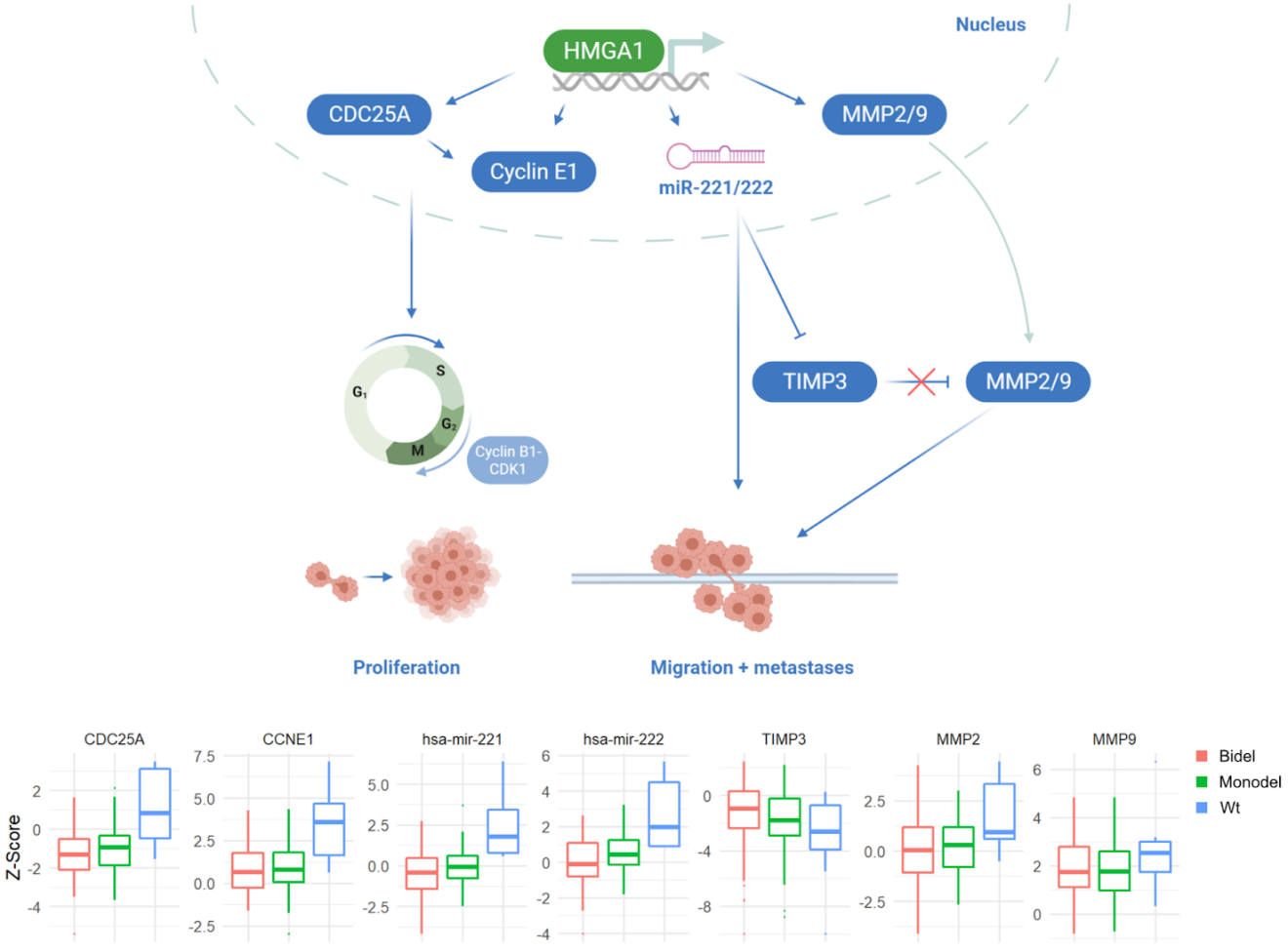

Figure 5: Schematic adapted from Fu et al. [55] showing the mechanism by which HMGA1 can promote tumour proliferation, migration and metastasis. HMGA1 regulates the expression of *CDC25A, Cyclin E1*, *MMP2*, *MMP9*, *miR-221* and *miR-222* all of which are specifically overexpressed in Wt samples. CDC25A activates Cyclin E1 necessary for G1/S transition which together with other factors overexpressed in Wt samples can increase tumour cell proliferation. Overexpression of *miR-221* and *miR-222* downregulates TIMP3 expressions which in turn results in the increased activity of MMP2 and MMP9. The mRNA expression of each gene across all groups is shown in the bottom panel. Changes in all factors bar TIMP3 are significantly different in Wt samples in comparison to normal ($q < 0.05$, $LFC > 1$)

Discussion

Here we describe the first multi-omic assessment of ^{wt}*VHL* ccRCC and determine the molecular basis for their aggressive phenotype using two independent cohorts. Mutation, CNV and methylation data are used to identify 8 ^{wt}*VHL* ccRCC in TCGA and 11 within our tissue biobank at the USZ; these are the sample sizes expected given the low incidence rates of ^{wt}*VHL* ccRCC. We find patients suffering from ^{wt}*VHL* tumours have a poorer prognosis than those with mono- or biallelic *VHL* inactivation. The genetic assessment of ^{wt}*VHL* ccRCC demonstrates a heterogeneous mutational landscape amongst which common ccRCC mutated genes are not found. Using both single and integrative omics analysis we find that factors promoting cell proliferation, EMT and cell invasion are enriched in and central to the molecular landscape observed specifically in ^{wt}*VHL* ccRCC. Using both the transcriptomic and proteomic data we are able to identify candidates to act as ^{wt}*VHL* ccRCC biomarkers increasing the

ease with which these rare tumours can be identified. Finally, we describe HMGA1 as a key regulator of multiple factors promoting proliferation and migration in *^{wt}VHL* ccRCC.

We find considerable overlap between the molecular landscapes of classical and *^{wt}VHL* ccRCC. The hypoxia pathway and the overactivation of *HIF-a* is central to tumourigenesis of classical ccRCC due to the inactivation of *VHL*[6,11]. Therefore, the enrichment of hypoxia associated genes in classical ccRCC was expected. However, Hypoxia enrichment amongst the *^{wt}VHL* ccRCC, which incur no changes in *HIF-a* mRNA expression, is suggestive of an alternative mechanism for the increased activity of *HIF-a* independent from the truncal inactivation of *VHL*. Pheochromocytoma and paraganglioma (PPGL) are a subtype of endocrine tumours in which the VHL/HIF axis is found to be frequently aberrated e.g *VHL* mutations and 3p deletions[56]. In PPGL, loss of *GIPC2* has been shown to activate *HIF-a* in tumours without *VHL* mutations and cause tumourigenesis. Loss of *GIPC2* can occur in PPGL via the loss of chromosome 1p or hypermethylation of its promoter region[43]. With the identification of a hypermethylated site together with the significant reduction of *GIPC2* mRNA expression in Wt samples, we hypothesise loss of *GIPC2* may lead to an increase in *HIF-a* activity in a portion of *^{wt}VHL* ccRCC. While *^{wt}VHL* ccRCC from the USZ contained a high proportion of Chr1p deletion we did not observe a significant reduction in protein expression of GIPC2 suggesting additional alternatives for HIF-α activity in these tumours.

Increased Ki67, a marker of cell proliferation, has previously been associated with *^{wt}VHL* ccRCC using immunohistochemistry[6]. Our analysis corroborates those findings as we identify cell cycle regulators such as CDC25A[31], Cyclins e.g. CCNE1 and Cyclin-dependent kinases e.g CDK1, being specifically overexpressed in *^{wt}VHL* ccRCC. In addition to these factors, our search for those genes broadly orchestrating the downstream changes specific to *^{wt}VHL* ccRCC given their genetic and epigenetic alterations further identified cell cycle regulators as being central to *^{wt}VHL* ccRCC specific processes e.g. CCNB1, AURKA and CSNK1D. The highly proliferative nature of *^{wt}VHL* ccRCC may provide an alternate explanation for the increase in hypoxia associated genes observed in these tumours. Increased proliferation and the activation of the hypoxia pathway often coincide, therefore, the increased activity of HIF-α may arise from a truly hypoxic tumour microenvironment developed further along the evolutionary timeline of *^{wt}VHL* ccRCC which in turn can further drive proliferation[57].

Along with an increased proliferative state, the upregulation of epithelial-mesenchymal transition and factors promoting and allowing cell invasion can contribute toward the poor prognosis attributed to Wt tumours. Enrichment of the expression of EMT factors were identified in both the transcriptome (TCGA) and proteome (USZ) of Wt samples. EMT is the process by which epithelial cells acquire mesenchymal phenotypes and increase factors allowing cell motility and invasion[30]. *TWIST1*, found to be significantly upregulated in Wt samples with a large log fold change (>3.5 mRNA), is a key transcription factor regulating EMT. The TWIST family of transcription factors are associated with the control of cell adhesion, cell migration and ECM degradation[58]. In addition to TWIST1 other transcription factors are specifically upregulated in Wt samples that increase invasive properties: *PRRX1*[59] and *RUNX2*[60]. PRRX1 can regulate EMT via several signalling pathways including TGF-β, all three isoforms of which are overexpressed in Wt samples[61]. Matrix metalloproteinases (MMPs) are a large family of endopeptidases that are necessary for the degradation of the extracellular matrix, promoting metastasis[62]. Amongst the seven MMPs found to be upregulated in Wt samples, MMP2 and

MMP9 have been shown to be regulated by both *RUNX2* and *HMGA1* while also being associated with metastasis and cell invasion[63,64].

Anti-angiogenic treatments targeting the tyrosine kinase receptor VEGFR such as Sunitinib and pazopanib are given as first-line therapies for RCC[3]. Here we find a reduction in VEGFA within both the transcriptome and proteome of *^{wt}VHL* ccRCC. Therefore, the targeted treatment against VEGFR may be less effective in patients possessing *^{wt}VHL* ccRCC in comparison to those with *VHL* inactivated tumours. Given the increased rates of proliferation and migration, therapies targeting these processes, for example, CDK4/6 inhibitors[65], may represent a more effective therapeutic option.

To aid the identification of *^{wt}VHL* ccRCC we identify potential candidates to act as biomarkers granting the ability to distinguish between *VHL* inactivated and *^{wt}VHL* ccRCC. HMGA1 regulation encompasses the processes we identify to be increased significantly in *^{wt}VHL* ccRCC[55] i.e. it promotes proliferative and migratory factors. We find it to be significantly overexpressed in both cohorts with a large increase in mediator activity within the TCGA cohort. In addition, *VTN*, *ISLR*, *MFAP2* and *IGF2* are amongst those that have the largest increases in mRNA expression within the TCGA samples and protein expression in the USZ Wt samples. VTN is an adhesive glycoprotein interacting with the ECM and was found to be upregulated in metastatic tumours[51,66]. ISLR, which is considered a marker for mesenchymal stem cells and EMT[53] also increased the migratory capabilities of tumour cells[67]; and *MFAP2* was found to be overexpressed in gastric cancers promoting migration and invasion[52]. IGF2 is a growth factor and downstream target of HIF-α, however, even with the stabilisation and activation of HIF-α in ccRCC, increased expression of *IGF2* is not found in *VHL* inactivated ccRCC. IGF2 has previously been implicated in several other tumour entities[68,69].

The identification of *^{wt}VHL* ccRCC requires the assessment of the *VHL* mutation, copy number and promoter methylation status. Therefore, *^{wt}VHL* ccRCC have largely been identified incidentally through large cohort studies and subsequently removed from further analyses as they aim to describe *VHL* inactivated ccRCC. The identification of a biomarker would allow the more efficient discovery of *^{wt}VHL* ccRCC without the need for large scale data generation in sizeable ccRCC cohorts, from which only a small number of *^{wt}VHL* ccRCC are expected to be identified. Further validation of these candidate markers is necessary in order to confidently apply them for the identification and selection of *^{wt}VHL* ccRCC. However, this endeavour would ultimately increase the ease with which *^{wt}VHL* ccRCC can be assessed with larger cohort sizes.

Conclusion

Here we present the first in-depth multi-omic assessment of *^{wt}VHL* ccRCC. With the upregulation of cell cycle regulators, EMT and factors promoting cell migration and invasion, patients possessing *^{wt}VHL* ccRCC experience poorer survival outcomes in comparison to those with *VHL* inactivations; supportive of a more aggressive ccRCC subgroup. The reduction in VEGFA expression in *^{wt}VHL* ccRCC in comparison to classical ccRCC may result in poorer response rates from first line ccRCC therapies. The validation of candidate biomarkers for *^{wt}VHL* ccRCC would simplify the identification of these tumours and allow for more effective targeted treatments.

Acknowledgements

The authors thank Susanne Dettwiler and Fabiola Prutek from the Tissue Biobank USZ for aiding the procurement of tissue material and IHC stainings. We also thank Adriana von Teichman for her routine *VHL* sequencing of ccRCC samples from the Tissue Biobank USZ. BioRender.com was used for the creation of schematic figures.

Author contributions

HM, PS and NB supervised the project. HM obtained funding. AAB and PS selected the cohort. DR obtained the clinical data. AAB and DR obtained the proteomics data. AAB gathered all experimental data and all public data. AAB carried out the data analysis and visualisation. AAB drafted the manuscript.

Bibliography

1. Wild, C. P., Weiderpass, E. & Stewart, B. W. *World Cancer Report: Cancer Research for Cancer Prevention*. (2020).
2. Maxwell, P. H. *et al.* The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271–275 (1999).
3. Escudier, B. *et al.* Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann. Oncol.* **30**, 706–720 (2019).
4. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
5. Sato, Y. *et al.* Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.* **45**, 860–867 (2013).
6. Turajlic, S. *et al.* Deterministic evolutionary trajectories influence primary tumor growth: tracerx renal. *Cell* **173**, 595-610.e11 (2018).
7. Martinez, P. *et al.* Parallel evolution of tumour subclones mimics diversity between tumours. *J. Pathol.* **230**, 356–364 (2013).
8. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46**, 225–233 (2014).
9. Moore, A. L. *et al.* Spatial distribution of private gene mutations in clear cell renal cell carcinoma. *Cancers (Basel)* **13**, (2021).
10. Dagher, J. *et al.* Multiple metastatic clones assessed by an integrative multiomics strategy in clear cell renal carcinoma: a case study. *J. Clin. Pathol.* (2021) doi:10.1136/jclinpath-2020-207326.
11. Mitchell, T. J. *et al.* Timing the landmark events in the evolution of clear cell renal cell cancer: tracerx renal. *Cell* **173**, 611-623.e17 (2018).
12. Batavia, A. A., Schraml, P. & Moch, H. Clear cell renal cell carcinoma with wild-type von Hippel-Lindau gene: a non-existent or new tumour entity? *Histopathology* **74**, 60–67 (2019).

13. Dagher, J. *et al.* Wild-type VHL Clear Cell Renal Cell Carcinomas Are a Distinct Clinical and Histologic Entity: A 10-Year Follow-up. *Eur. Urol. Focus* **1**, 284–290 (2016).
14. DiNatale, R. G. *et al.* Putative Drivers of Aggressiveness in TCEB1-mutant Renal Cell Carcinoma: An Emerging Entity with Variable Clinical Course. *Eur. Urol. Focus* **7**, 381–389 (2021).
15. Hakimi, A. A. *et al.* TCEB1-mutated Renal Cell Carcinoma: A Distinct Genomic and Morphologic Subtype. *Journal of Urology* **191**, E249–E249 (2014).
16. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
17. Broad Institute TCGA Genome Data Analysis Center. FireBrowse. http://firebrowse.org/.
18. Colaprico, A. *et al.* TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).
19. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci USA* **103**, 1412–1417 (2006).
20. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
21. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (N Y)* **2**, 100141 (2021).
22. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–15550 (2005).
23. Dimitrakopoulos, C. *et al.* Network-based integration of multi-omics data for prioritizing cancer genes. *Bioinformatics* **34**, 2441–2448 (2018).
24. Bruderer, R. *et al.* Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Mol. Cell. Proteomics* **16**, 2296–2309 (2017).
25. Huang, T. *et al.* Combining precursor and fragment information for improved detection of differential abundance in data independent acquisition. *Mol. Cell. Proteomics* **19**, 421–430 (2020).
26. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* **100**, 9440–9445 (2003).
27. Ocaña, O. H. *et al.* Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell* **22**, 709–724 (2012).
28. Takano, S. *et al.* Prrx1 isoform switching regulates pancreatic cancer invasion and metastatic colonization. *Genes Dev.* **30**, 233–247 (2016).
29. Pratap, J. *et al.* The Runx2 osteogenic transcription factor regulates matrix metalloproteinase 9 in bone metastatic cancer cells and controls cell invasion. *Mol. Cell. Biol.* **25**, 8581–8591 (2005).
30. Yang, J. *et al.* Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **21**, 341–352 (2020).

31. Boutros, R., Lobjois, V. & Ducommun, B. CDC25 phosphatases in cancer cells: key players? Good targets? *Nat. Rev. Cancer* **7**, 495–507 (2007).
32. Blomberg, I. & Hoffmann, I. Ectopic expression of Cdc25A accelerates the G(1)/S transition and leads to premature activation of cyclin E- and cyclin A-dependent kinases. *Mol. Cell. Biol.* **19**, 6183–6194 (1999).
33. Lau, K.-M. *et al.* Overexpression of HMGA1 deregulates tumor growth via cdc25A and alters migration/invasion through a cdc25A-independent pathway in medulloblastoma. *Acta Neuropathol.* **123**, 553–571 (2012).
34. Ricketts, C. J. *et al.* The cancer genome atlas comprehensive molecular characterization of renal cell carcinoma. *Cell Rep.* **23**, 313-326.e5 (2018).
35. Cancer Genome Atlas Research Network *et al.* Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N. Engl. J. Med.* **374**, 135–145 (2016).
36. Moore, L. D., Isayeva, T., Siegal, G. P. & Ponnazhagan, S. Silencing of transforming growth factor-beta1 in situ by RNA interference for breast cancer: implications for proliferation and migration in vitro and metastasis in vivo. *Clin. Cancer Res.* **14**, 4961–4970 (2008).
37. Liu, Z., Yi, L., Du, M., Gong, G. & Zhu, Y. Overexpression of TGF-β enhances the migration and invasive ability of ectopic endometrial cells via ERK/MAPK signaling pathway. *Exp. Ther. Med.* **17**, 4457–4464 (2019).
38. Melzer, C., von der Ohe, J., Hass, R. & Ungefroren, H. TGF-β-Dependent Growth Arrest and Cell Migration in Benign and Malignant Breast Epithelial Cells Are Antagonistically Controlled by Rac1 and Rac1b. *Int. J. Mol. Sci.* **18**, (2017).
39. Li, D. *et al.* KRT17 Functions as a Tumor Promoter and Regulates Proliferation, Migration and Invasion in Pancreatic Cancer via mTOR/S6k1 Pathway. *Cancer Manag. Res.* **12**, 2087–2095 (2020).
40. Gupta, A. *et al.* Promising noninvasive cellular phenotype in prostate cancer cells knockdown of matrix metalloproteinase 9. *ScientificWorldJournal* **2013**, 493689 (2013).
41. Senbanjo, L. T. & Chellaiah, M. A. CD44: A multifunctional cell surface adhesion receptor is a regulator of progression and metastasis of cancer cells. *Front. Cell Dev. Biol.* **5**, 18 (2017).
42. Gupta, A., Cao, W. & Chellaiah, M. A. Integrin αvβ3 and CD44 pathways in metastatic prostate cancer cells support osteoclastogenesis via a Runx2/Smad 5/receptor activator of NF-κB ligand signaling axis. *Mol. Cancer* **11**, 66 (2012).
43. Dong, Y. *et al.* GIPC2 is an endocrine-specific tumor suppressor gene for both sporadic and hereditary tumors of RET- and SDHB-, but not VHL-associated clusters of pheochromocytoma/paraganglioma. *Cell Death Dis.* **12**, 444 (2021).
44. Gremel, G. *et al.* A systematic search strategy identifies cubilin as independent prognostic marker for renal cell carcinoma. *BMC Cancer* **17**, 9 (2017).
45. Niinivirta, M. *et al.* Tumoral cubilin is a predictive marker for treatment of renal cancer patients with sunitinib and sorafenib. *J. Cancer Res. Clin. Oncol.* **143**, 961–970 (2017).
46. Kruck, S. *et al.* Impact of an altered Wnt1/β-catenin expression on clinicopathology and prognosis in clear cell renal cell carcinoma. *Int. J. Mol. Sci.* **14**, 10944–10957 (2013).

47. Tumkur Sitaram, R., Landström, M., Roos, G. & Ljungberg, B. Significance of PI3K signalling pathway in clear cell renal cell carcinoma in relation to VHL and HIF status. *J. Clin. Pathol.* **74**, 216–222 (2021).
48. An, J. & Rettig, M. B. Mechanism of von Hippel-Lindau protein-mediated suppression of nuclear factor kappa B activity. *Mol. Cell. Biol.* **25**, 7546–7556 (2005).
49. Wang, Z. *et al.* Cyclin B1/Cdk1 coordinates mitochondrial respiration for cell-cycle G2/M progression. *Dev. Cell* **29**, 217–232 (2014).
50. Clemm von Hohenberg, K. *et al.* Cyclin B/CDK1 and Cyclin A/CDK2 phosphorylate DENR to promote mitotic protein translation and faithful cell division. *Nat. Commun.* **13**, 668 (2022).
51. Burgos-Panadero, R., Noguera, I., Cañete, A., Navarro, S. & Noguera, R. Vitronectin as a molecular player of the tumor microenvironment in neuroblastoma. *BMC Cancer* **19**, 479 (2019).
52. Yao, L.-W. *et al.* MFAP2 is overexpressed in gastric cancer and promotes motility via the MFAP2/integrin α5β1/FAK/ERK pathway. *Oncogenesis* **9**, 17 (2020).
53. Maeda, K. *et al.* Identification of meflin as a potential marker for mesenchymal stromal cells. *Sci. Rep.* **6**, 22288 (2016).
54. Feldser, D. *et al.* Reciprocal positive regulation of hypoxia-inducible factor 1alpha and insulin-like growth factor 2. *Cancer Res.* **59**, 3915–3918 (1999).
55. Fu, F. *et al.* HMGA1 exacerbates tumor growth through regulating the cell cycle and accelerates migration/invasion via targeting miR-221/222 in cervical cancer. *Cell Death Dis.* **9**, 594 (2018).
56. Fishbein, L. *et al.* Comprehensive molecular characterization of pheochromocytoma and paraganglioma. *Cancer Cell* **31**, 181–193 (2017).
57. Koong, A. C. *et al.* Candidate genes for the hypoxic tumor phenotype. *Cancer Res.* **60**, 883–887 (2000).
58. Yang, J. *et al.* Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**, 927–939 (2004).
59. Wu, X. & Bao, H. Tumor suppressive microRNA-485-5p targets PRRX1 in human skin melanoma cells, regulating epithelial-mesenchymal transition and apoptosis. *Cell Biol. Int.* **45**, 1404–1414 (2021).
60. Niu, D.-F. *et al.* Transcription factor Runx2 is a regulator of epithelial-mesenchymal transition and invasion in thyroid carcinomas. *Lab. Invest.* **92**, 1181–1190 (2012).
61. Derynck, R., Turley, S. J. & Akhurst, R. J. TGFβ biology in cancer progression and immunotherapy. *Nat. Rev. Clin. Oncol.* **18**, 9–34 (2021).
62. Egeblad, M. & Werb, Z. New functions for the matrix metalloproteinases in cancer progression. *Nat. Rev. Cancer* **2**, 161–174 (2002).
63. Takeha, S. *et al.* Stromal expression of MMP-9 and urokinase receptor is inversely associated with liver metastasis and with infiltrating growth in human colorectal cancer: a novel approach from immune/inflammatory aspect. *Jpn J Cancer Res* **88**, 72–81 (1997).
64. Hillion, J. *et al.* Upregulation of MMP-2 by HMGA1 promotes transformation in undifferentiated, large-cell lung cancer. *Mol. Cancer Res.* **7**, 1803–1812 (2009).
65. Sager, R. A. *et al.* Therapeutic potential of CDK4/6 inhibitors in renal cell carcinoma. *Nat. Rev. Urol.* **19**, 305–320 (2022).
66. Bera, A. *et al.* Functional role of vitronectin in breast cancer. *PLoS ONE* **15**, e0242141 (2020).

67. Chi, C. *et al.* ISLR affects colon cancer progression by regulating the epithelial-mesenchymal transition signaling pathway. *Anticancer Drugs* **33**, e670–e679 (2022).
68. Zhao, R. *et al.* Loss of imprinting of the insulin-like growth factor II (IGF2) gene in esophageal normal and adenocarcinoma tissues. *Carcinogenesis* **30**, 2117–2122 (2009).
69. El Tayebi, H. M. *et al.* Expression of insulin-like growth factor-II, matrix metalloproteinases, and their tissue inhibitors as predictive markers in the peripheral blood of HCC patients. *Biomarkers* **16**, 346–354 (2011).

Wild-type *von Hippel-Lindau* clear cell renal cell carcinoma confer a poorer prognosis due to increases in proliferative and migratory properties

Aashil A. Batavia[1,2,3], Dorothea Rutishauser[1], Peter Schraml[1], Niko Beerenwinkel[2,3] and Holger Moch[1]*

[1]*Department of Pathology and Molecular Pathology, University and University Hospital Zurich, Schmelzbergstrasse 12, 8091 Zurich,Switzerland*

[2]*Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, 4058 Basel*

[3]*SIB Swiss Institute of Bioinformatics, Mattenstrasse 26, 4058 Basel, Switzerland*

* Corresponding author

Supplementary data

Figures S1-S10
Table S1-S2

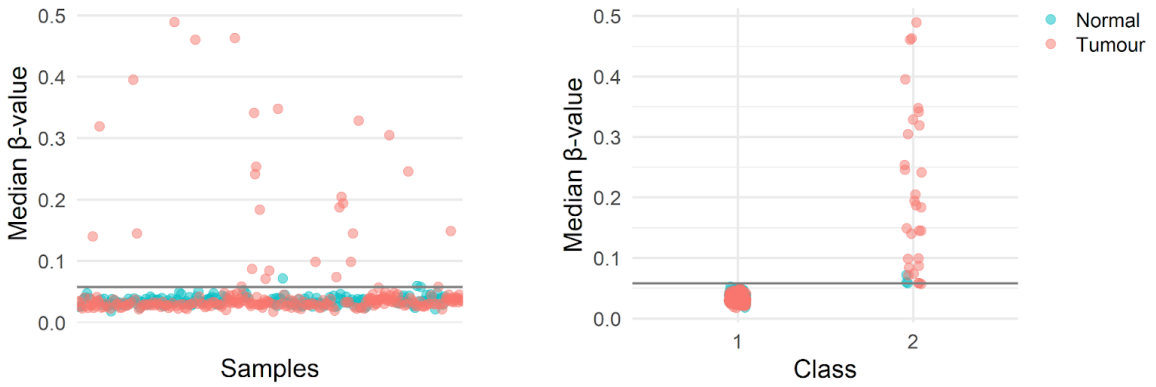

Figure S1: *VHL* methylation status was determined via the comparison to normal samples. We calculate a threshold of 0.058 corresponding to a p-value of 0.001 under which we classify a sample as being unmethylated (right panel). We use a gaussian mixture model to cluster samples in an unsupervised manner in order to validate the methylation calls and ensure those classified as unmethylated fall in the same cluster as normal samples.

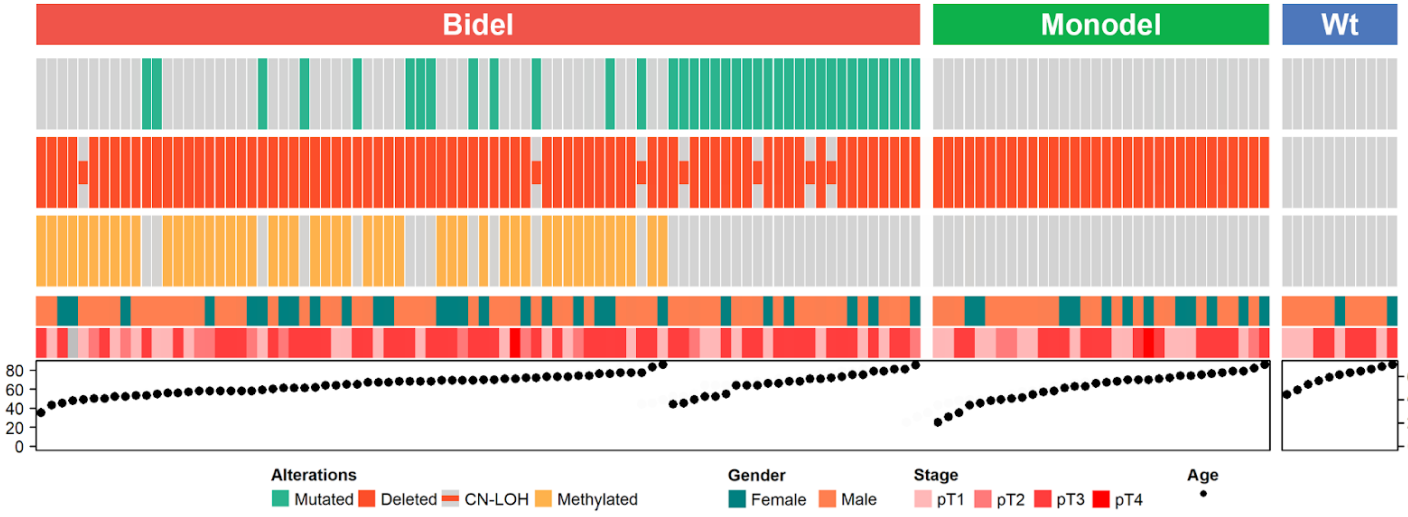

Figure S2: *VHL* aberrations and clinical properties of 127 ccRCC from the USZ renal biobank. We find 84 Biallelically inactivated *VHL* ccRCC, 32 monoallelically *VHL* inactivated ccRCC, and 11 ^{wt}*VHL* ccRCC possessing no *VHL* inactivations.

Figure S3: **A)** Summary of commonly mutated genes in ccRCC across the Bidel (n = 144), Monodel, (n = 89) and Wt (n = 8) groups. Aside from a TP53 mutation found in many cancer types, no common ccRCC are found within the Wt group. **B)** Summary of all mutations identified in Wt samples showing they possess heterogeneous mutational landscapes with few shared mutations.

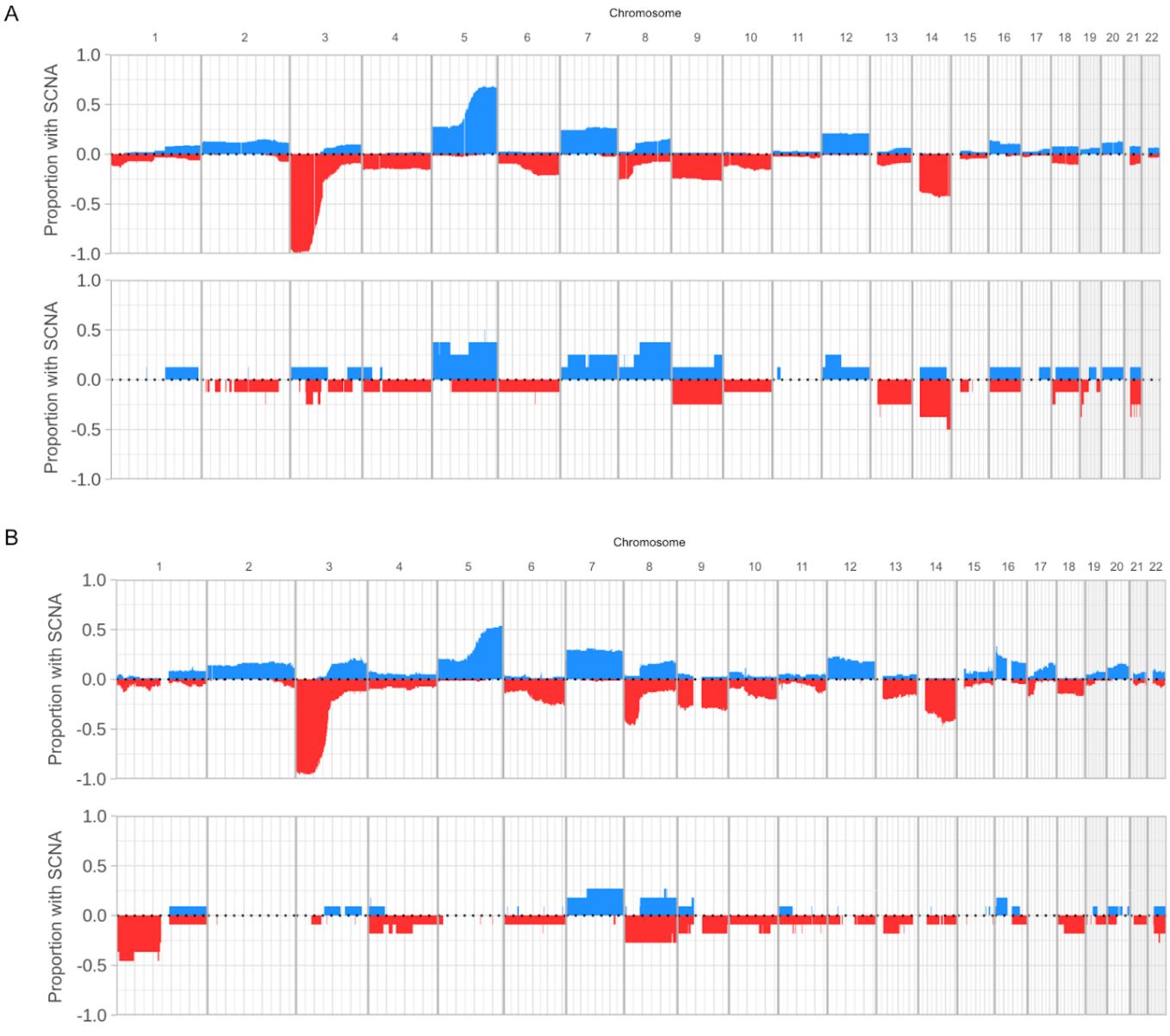

Figure S4: A) Frequency of copy number events of the Bidel samples (top panel; n = 144) and Wt samples (bottom panel; n = 8) from the TCGA cohort. **B)** Frequency of copy number events of the Bidel samples (top panel; n = 84) and Wt samples (bottom panel; n = 11) from the USZ cohort. A larger proportion of Wt samples experience chr1p deletions in the USZ cohort. USZ Wt samples also lack chr5q amplifications which are often found to co-occur with 3p deletions as seen in the Bidel group.

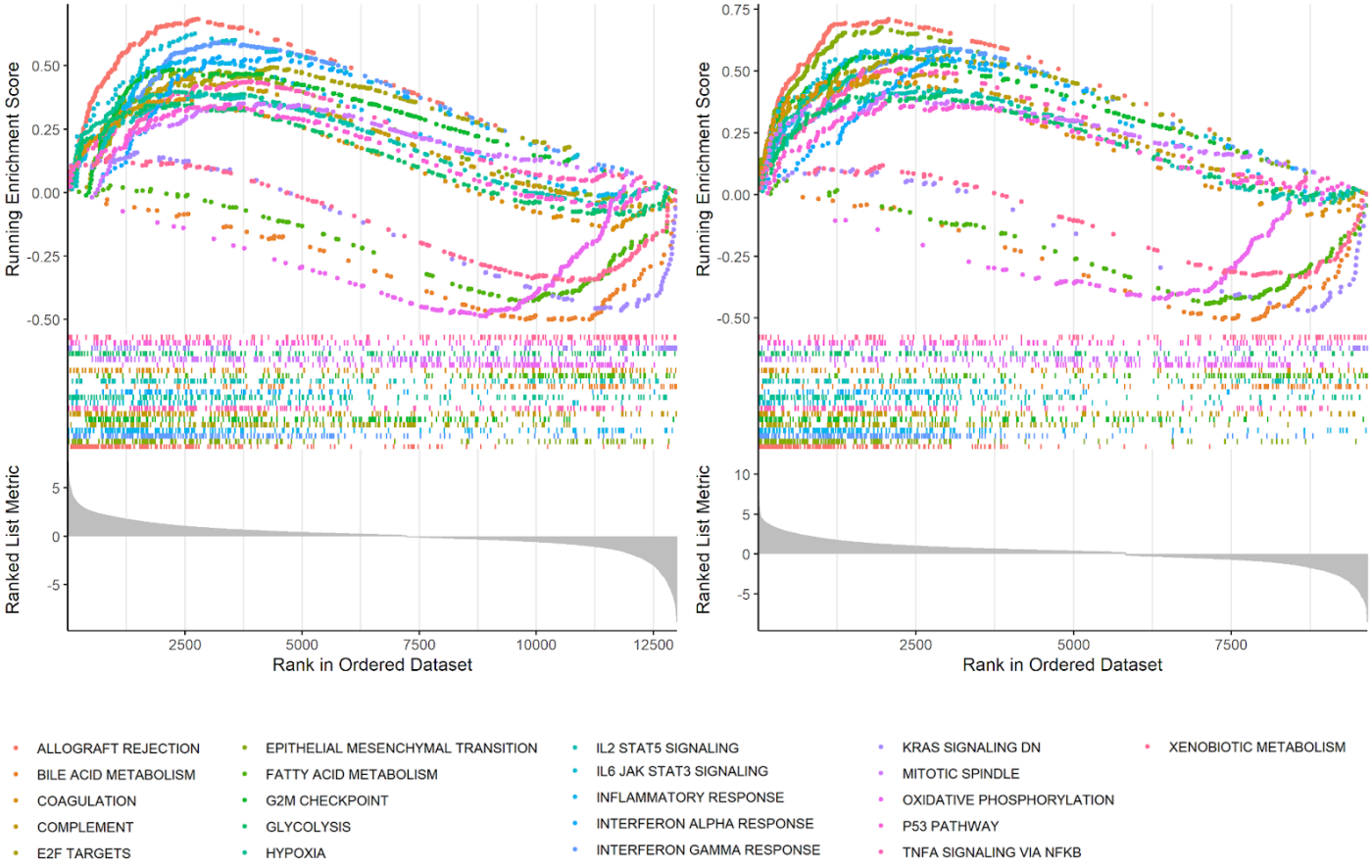

Figure S5: Enrichment plots for those gene sets enriched in both Bidel samples (A) and Wt samples (B). Gene set enrichment analysis is conducted using the Hallmark gene sets from MSigDB following differential mRNA expression analysis (Tumour vs Normal).

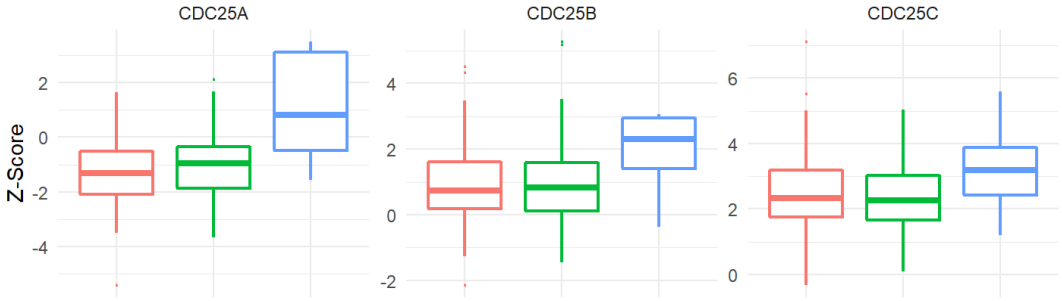

Figure S6: Expression of the CDC25 phosphatases across the three tumour groups using normal samples as the reference. CDC25A is specifically overexpressed in Wt samples in comparison to both normal and Bidel samples (q-value = 4.79 10-5 and LFC = 1.21 vs normal; q-value = 2.66 10-11 and LFC = 2.02 vs Bidel).

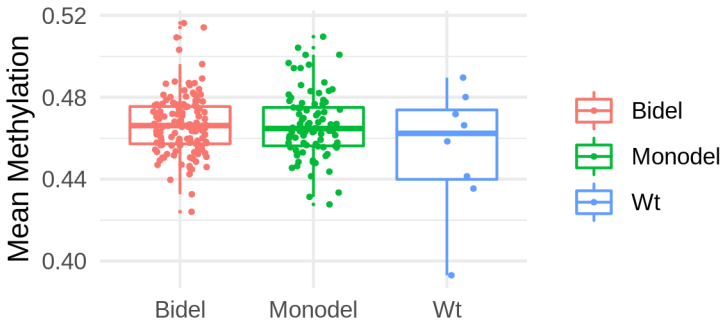

Figure S7: Mean methylation of each sample separated by *VHL* status. No significant difference was found between the groups.

Table S1: Genes with the largest increase in mediator activity in Wt samples in comparison to Bidel samples, orchestrating the Wt specific differential gene expression given the genetic and epigenetic aberration. The top 1000 mediators are considered from both groups.

Gene	Change in average rank*
NCBP2	-1018.99
AURKA	-988.21
TGFB3	-981.69
ADCY9	-902.88
CCNB1	-884.24
TAF1	-876.73
CSNK1D	-874.41
CARM1	-865.37
LAMA3	-862.00
ITGAE	-830.44
POLR2A	-775.63
CKAP5	-758.17
MICALL1	-732.86
IGF2	-722.48
FOSL1	-701.81
MMP9	-692.76
CDH17	-677.77
ATP1A1	-673.57
RBBP7	-673.45
MEF2A	-664.17

*negative changes in rank indicate a higher rank in Wt samples vs Bidel samples.

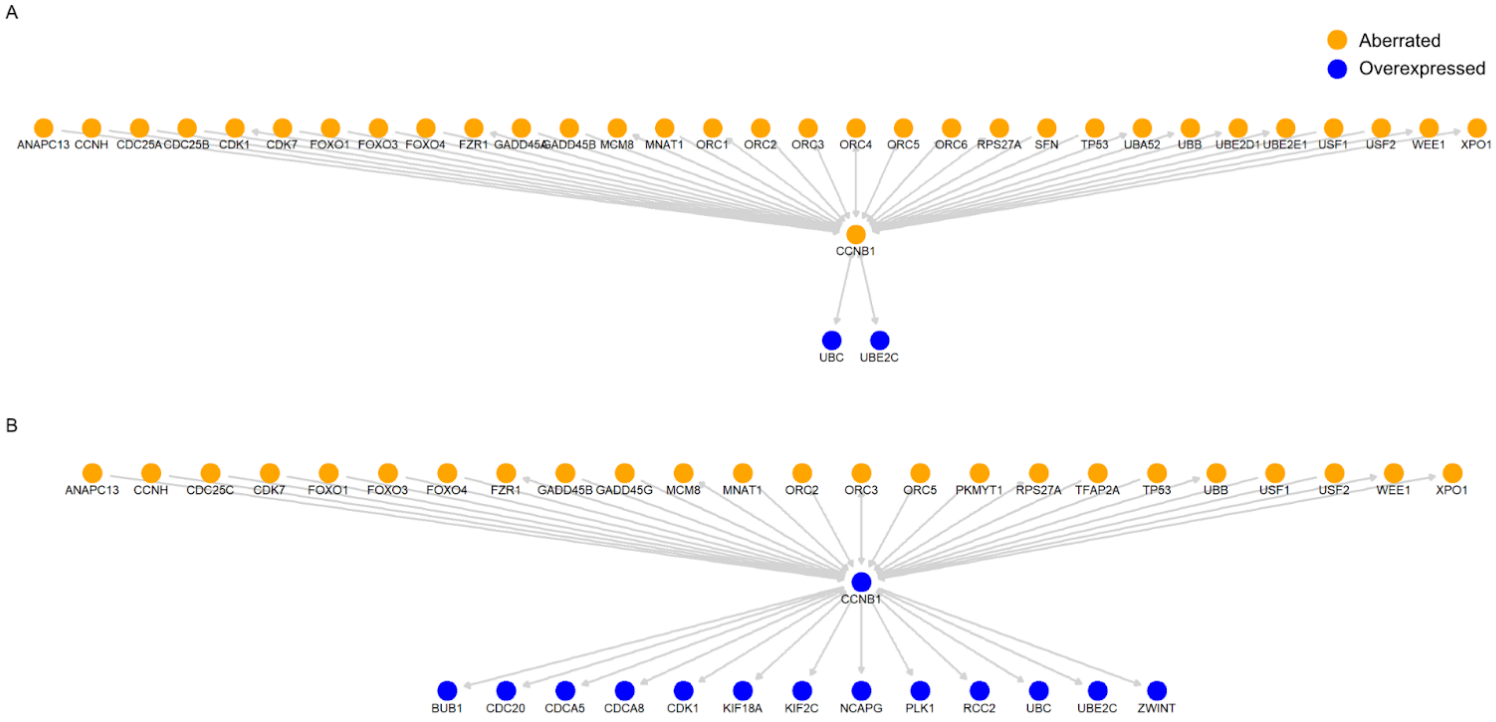

Figure S8: Change in mediator activity on CCNB1 from within Bidel samples **(A)** and Wt samples **(B)**. We observe fewer mutations directly upstream of CCNB1 and an increase in overexpressed genes directly downstream of CCNB1 including CDK1, BUB1, and CDC proteins all necessary for cell division.

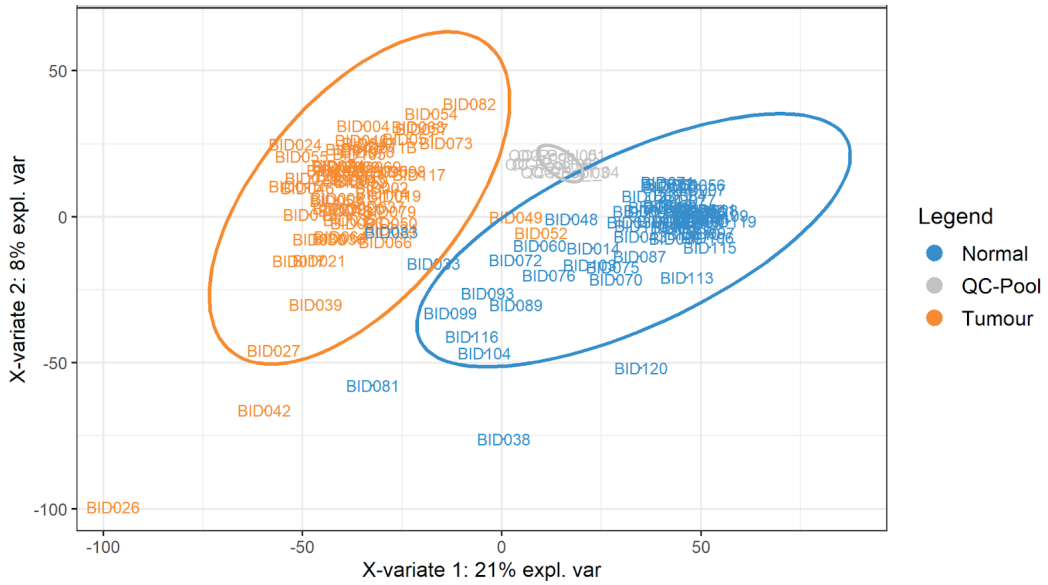

Figure S9: PLS-DA was performed using the protein intensities in each sample. We find clear separation between the Normal and Tumour with the QC-pool samples forming a cluster between the two as expected.

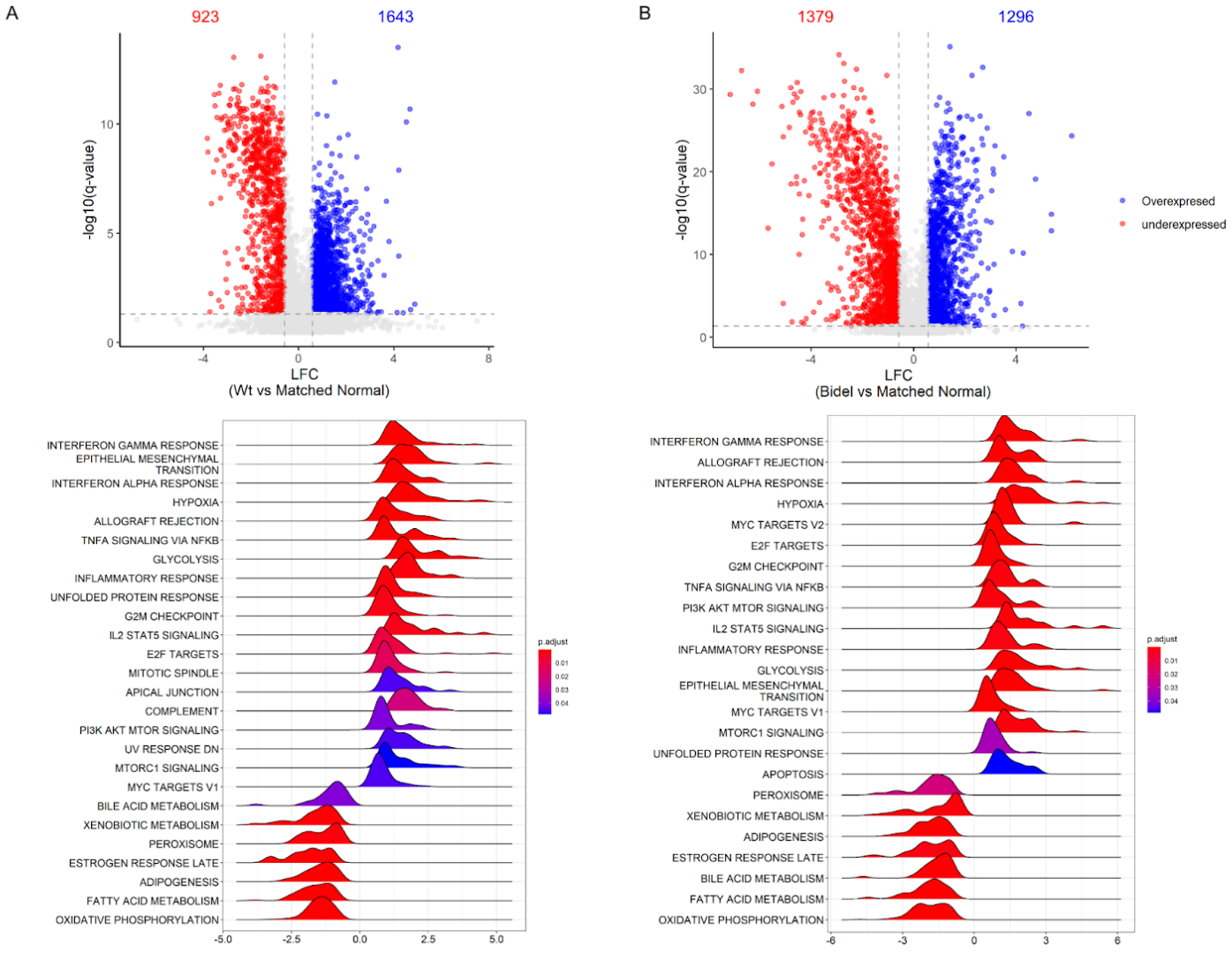

Figure S10: Differential protein expression and corresponding enriched gene sets for Wt samples **(A)** and Bidel samples **(B)** in comparison to matched normal samples from the USZ.

IV

Concluding remarks

In an era where multilayered high dimensional data is rapidly attainable, it is becoming increasingly vital to employ a multidisciplinary approach to answer complex biomedical questions. This dissertation presented an assessment of wild-type *von Hippel-Lindau* clear cell renal cell carcinoma (*^{wt}VHL* ccRCC), a rare and scarcely characterised tumour entity.

Clinically recognised renal cell carcinoma (RCC) subtypes as endorsed by the World Health Organisation (WHO) are diagnosed using histological criteria i.e predominant cytoplasmic observations. However, progressing into the modern era of molecular medicine, tumour diagnostics is now moving toward a classification system incorporating morphological, immunohistochemical and molecular features pertinent to specific entities[2]. The histological characteristic of clear cell renal cell carcinoma is, unsurprisingly, the presence of clear cell histology. As mentioned in Chapter **I** however, clear cell histology has also been observed in newer RCC subtypes alongside other histological and immunohistochemical characteristics that when missed can lead to misdiagnosis[3,4] e.g. clear cell papillary RCC (CCPRCC) and renal angiomyoadenomatous tumours (RAT). Therefore, the molecular assessment of these tumours is essential for their accurate classification. Clear cell renal cell carcinoma have been well charecterised[5–10]. The biallelic inactivation of the *VHL* gene is considered the molecular hallmark of ccRCC and is not seen in other molecularly distinguishable RCC entities[11–14]. Large scale integrated omic analyses of ccRCC, however, identify tumours with no *VHL* activation even after accounting for novel clear cell renal entities mentioned above[7]. These *^{wt}VHL* ccRCC, which are often exuded from any detailed molecular assessments of ccRCC, formed the subject matter for this dissertation and were assessed using cohorts from the USZ Tissue Biobank and The Cancer Genome Atlas (TCGA).

The aetiology of a small fraction of *^{wt}VHL* ccRCC is considered to be the biallelic inactivation of the *ELOC* gene (previously called *TCEB1*), first identified in ccRCC with no *VHL* inactivation by Sato et al[6]. ELOC is a binding partner of pVHL within the E3 ubiquitin ligase complex mediating the

activation of HIF-α. To date, 20 of these ELOC RCC with clear cell histology have been discovered worldwide[15]. In Chapter **II** the identification of 3 of these 20 rare tumour entities from the systematic assessment of 464 ccRCC from the USZ Tissue Biobank, and their intersection with ccRCC was described. Whole-exome sequencing and OncoScan SNP-arrays allowed the identification of tumours containing both *ELOC* mutations and chromosomal *ELOC* deletions causing the biallelic inactivation of *ELOC*. A novel variant of ELOC was identified within its *VHL* binding domain [16,17]. In order to confirm the disturbance in the interaction between ELOC and pVHL, High Ambiguity Driven Biomolecular Docking (HADDOCK)[18] was employed. Previous structural assessments of ELOC variants used static side chain changes in order to identify clashes[6,15]. In doing so, the macromolecular impact of these variants are not considered and therefore does not represent the true *in vivo* structure. Using HADDOCK allows the acquisition of pVHL-ELOC docking predictions within conditions comparable to the intracellular environment, resulting in a more accurate and valid description of the structural consequences of a variant. The novel F77_T78dup in ELOC prevented the interaction of residues that are known to disrupt pVHL-ELOC binding, significantly increased the energy required to form the complex and reduced the size of their interface.

Previously Hakimi et al. had found the loss of *ELOC* led to the reduction in ELOC mRNA expression[19]. Using mass spectrometry, I show that this change within the transcriptome is reflected in the proteome with a significant reduction in ELOC protein expression within ELOC RCC. Taking the HADDOCK results and proteomics data together, I demonstrate how mutations and chromosomal deletions are translated to prevent ELOC function within the proteome. As ELOC and pVHL are interacting partners within the E3 ubiquitin ligase complex targeting HIF-α, it is hypothesised that the events leading to tumourigenesis initiate similar downstream signalling cascades as see in classical ccRCC. Our finding that ELOC RCC possess similar CAIX and VEGFA protein expression compared to classical ccRCC support this notion and finds ELOC functions as a classical tumour suppressor gene to initiate tumourigenesis.

As of 2022 ELOC RCC are considered a separate RCC tumour entity by the WHO [20]. However, there remain ^{wt}*VHL* ccRCC with unknown aetiology. The manuscript presented in Chapter **III** focused on this set of tumours. With the use of Sanger sequencing, bisulfite sequencing and OncoScan SNP-arrays, *VHL* status is determined for 127 ccRCC from the USZ Tissue Biobank, of which 11 are ^{wt}*VHL* ccRCC. Genome-wide CNV and mass spectrometry data were generated for these samples. TCGA represents a valuable resource for the research community consisting of bulk multi-omics data for 33 different tumour types encompassing 11,000 patients. Using TCGA ccRCC dataset, *VHL* status was determined for an additional 242 ccRCC tumours. Within TCGA cohort 8 ^{wt}*VHL* ccRCC were identified for which clinical, genetic (mutations and CNVs), epigenetic (methylation) and transcriptomic (mRNA and miRNA) data was acquired. The assessment of multiple omic layers allows the reliable identification of processes which are central in governing the clinical phenotypes observed in a disease of interest[21]. This is especially true here when assessing ^{wt}*VHL* ccRCC given the small sample sizes.

The limited literature regarding ^{wt}*VHL* ccRCC indicated a more aggressive phenotype for those tumours with active *VHL*[22]. Applying cox-proportional hazard models we indeed find ^{wt}*VHL* ccRCC to confer poorer survival in comparison to *VHL* inactivated ccRCC. The more aggressive nature of

these tumours may be attributed to the upregulation of factors promoting cell proliferation, invasion and EMT in comparison to *VHL* inactivated ccRCC. These factors are identified via both single and integrative omics analysis. Network-based integration of the genetic, epigenetic and transcriptomic data provided a holistic assessment of the aberrations identified in ^{wt}*VHL* ccRCC[23]. This allowed the identification of key orchestrators of downstream differential expression, which in turn, cause the clinical phenotypes observed in ^{wt}*VHL* ccRCC.

We find the overexpression of hypoxia associated genes in ^{wt}*VHL* ccRCC. The observation of a significant reduction in *GIPC2* expression within ^{wt}*VHL* ccRCC obtained from TCGA indicates a separate, *VHL* independent, method for HIF-α over-activation in these samples. A similar phenomenon has been observed in Pheochromocytoma and paraganglioma (PPGL)[24]. A cluster of PPGL were found to obtain biallelic inactivation of *GIPC2* via chromosomal deletion and hypermethylation resulting in reduced expression of *GIPC2* and over-activation of HIF-α. This led the authors to conclude that *GIPC2* acts as a novel tumour suppressor gene and provides a molecular mechanism for the tumorigenesis of *SDHB* and *RET* associated PPGL but not in those with *VHL* inactivation. With ELOC-mutated RCC now being a considered a molecularly defined RCC subtype resulting in similar downstream perturbations as in classical ccRCC, *GIPC2* inactivation in ccRCC should be explored further as they have demonstrated the properties of classical tumour suppressor genes much like *ELOC* and *VHL*. *GIPC2* dependant upregulation of HIF-α may explain tumourigenesis in a portion of ^{wt}*VHL* ccRCC.

One of the key treatment strategies for ccRCC is the administration of antiangiogenic factors. These are effective due to the increases in VEGFA expression observed in ccRCC. In ^{wt}*VHL* ccRCC, however, the increase in proliferation factors along with a reduction in VEGFA expression in comparison to classical ccRCC suggests alternative treatment strategies may be beneficial in patients suffering from these tumours. CDK4/CDK6 inhibitors e.g Abemaciclib are being explored as treatment strategies in RCC [25]. These inhibitors target cell cycle progression at the G1/S phase and are approved for use in other cancer subtypes e.g. breast[26]. This therapeutic option can be used in conjunction with first line treatment strategies i.e. immune checkpoint inhibitors such as nivolumab and pembrolizumab in ccRCC. With the significantly poorer prognosis conferred by ^{wt}*VHL*, indicative of a more aggressive subgroup of ccRCC, noval treatment options targetting factors identified here e.g *HMGA1* can be explored in order to reduce mortality in these patients.

This assessment of ^{wt}*VHL* ccRCC was the first to use a multi-layered approach to specifically describe the molecular landscape within these tumours. This is potentially due to the rarity with which these tumours occur along and the need for mutation, copy number and methylation data for their identification. In order to provide a more efficient means of identifying ^{wt}*VHL* ccRCC, candidate biomarkers are proposed that may be used to distinguish them from *VHL* inactivated ccRCC e.g. HMGA1 and IGF2. In the near future, whether the overexpression of these markers is evident using immunohistochemistry and tissue microarrays, consisting of both ^{wt}*VHL* ccRCC and *VHL* inactivated ccRCC, will be assessed. If this is successful and specific overexpression of the markers can be seen in ^{wt}*VHL* ccRCC using a proposed candidate then further validation will be required on an additional independent ccRCC cohort. As ^{wt}*VHL* ccRCC are an extremely rare group of tumours a limitation of the work presented here are the sample sizes used and therefore the generalisability of the findings. This is especially pertinent to heterogeneous tumour types such as ccRCC. The use of two independent cohorts is a means by which I aimed to mitigate this limitation. However, the successful

validation of a biomarker would increase the ease with which these tumours can be found and therefore studied. Together with the rise of global biobanking efforts[27,28], a multi-institutional collaborative study would be preferential to assess these tumours with a larger cohort; providing greater insight as to their aetiology.

In conclusion, it is evident from the 2022 WHO guidelines that the molecular aberrations within rare renal tumours, which cannot be distinguished using morphology alone, are becoming significant contributors for classification schemes aiding diagnostics. While the weight placed on either molecular or morphological properties when making these classification is still debated, it is clear that the histological characteristics, molecular features and the conferred clinical phenotypes of tumours are essential for their classification. As of 2022 the biallelic inactivation of *ELOC* is considered a hallmark of a novel RCC tumour entity termed ELOC-mutated RCC. However, the inactivation of *ELOC* is akin to the biallelic inactivation of *VHL* with respect to their downstream ramifications. With the consideration of the various subtypes of RCC, there still remain ^{wt}*VHL* ccRCC with unknown aetiology. These ^{wt}*VHL* ccRCC describe a subgroup of aggressive tumours conferring a poorer survival potentially due to increases in invasive properties. The current first-line treatment strategies for ccRCC may be ineffective in ^{wt}*VHL* tumours given the downregulation of their targets e.g. *VEGFA*. The results here demonstrate that the molecular evaluation of ^{wt}*VHL* ccRCC on a larger scale would greatly benefit the identification of treatment options better suited to patients suffering from these tumours.

Bibliography

1. Moch, H. *et al.* The 2022 WHO Classification of Tumours of the Urinary System and Male Genital OrgansPart A: Renal, Penile and TesticularTumours. *(In press)*.
2. Pfister, S. M. *et al.* A Summary of the Inaugural WHO Classification of Pediatric Tumors: Transitioning from the Optical into the Molecular Era. *Cancer Discov.* **12**, 331–355 (2022).
3. Gill, S. *et al.* Incidence of Clear Cell Papillary Renal Cell Carcinoma in Low-Grade Renal Cell Carcinoma Cases: A 12-Year Retrospective Clinicopathologic Study From a Single Cancer Center. *Int. J. Surg. Pathol.* **24**, 207–212 (2016).
4. Zhou, H. *et al.* Clear cell papillary renal cell carcinoma is the fourth most common histologic type of renal cell carcinoma in 290 consecutive nephrectomies for renal cell carcinoma. *Hum. Pathol.* **45**, 59–64 (2014).
5. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
6. Sato, Y. *et al.* Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat. Genet.* **45**, 860–867 (2013).
7. Turajlic, S. *et al.* Deterministic evolutionary trajectories influence primary tumor growth: tracerx renal. *Cell* **173**, 595-610.e11 (2018).
8. Martinez, P. *et al.* Parallel evolution of tumour subclones mimics diversity between tumours. *J. Pathol.* **230**, 356–364 (2013).
9. Mitchell, T. J. *et al.* Timing the landmark events in the evolution of clear cell renal cell cancer: tracerx renal. *Cell* **173**, 611-623.e17 (2018).
10. Turajlic, S. *et al.* Tracking cancer evolution reveals constrained routes to metastases: tracerx renal. *Cell* **173**, 581-594.e12 (2018).
11. Ricketts, C. J. *et al.* The cancer genome atlas comprehensive molecular characterization of renal cell carcinoma. *Cell Rep.* **23**, 313-326.e5 (2018).

12. Wolfe, A., Dobin, S. M., Grossmann, P., Michal, M. & Donner, L. R. Clonal trisomies 7,10 and 12, normal 3p and absence of VHL gene mutation in a clear cell tubulopapillary carcinoma of the kidney. *Virchows Arch.* **459**, 457–463 (2011).
13. Michal, M. *et al.* Renal angiomyoadenomatous tumor: morphologic, immunohistochemical, and molecular genetic study of a distinct entity. *Virchows Arch.* **454**, 89–99 (2009).
14. Deml, K.-F. *et al.* Clear cell papillary renal cell carcinoma and renal angiomyoadenomatous tumor: two variants of a morphologic, immunohistochemical, and genetic distinct entity of renal cell carcinoma. *Am. J. Surg. Pathol.* **39**, 889–901 (2015).
15. DiNatale, R. G. *et al.* Putative Drivers of Aggressiveness in TCEB1-mutant Renal Cell Carcinoma: An Emerging Entity with Variable Clinical Course. *Eur. Urol. Focus* **7**, 381–389 (2021).
16. Takagi, Y., Pause, A., Conaway, R. C. & Conaway, J. W. Identification of elongin C sequences required for interaction with the von Hippel-Lindau tumor suppressor protein. *J. Biol. Chem.* **272**, 27444–27449 (1997).
17. Stebbins, C. E., Kaelin, W. G. & Pavletich, N. P. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science* **284**, 455–461 (1999).
18. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).
19. Hakimi, A. A. *et al.* TCEB1-mutated Renal Cell Carcinoma: A Distinct Genomic and Morphologic Subtype. *Journal of Urology* **191**, E249–E249 (2014).
20. International Agency for Research on Cancer. *WHO Classification of Tumours ; Urinary and male genital tumours. 5th ed*. (2022).
21. Chen, R. *et al.* Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* **148**, 1293–1307 (2012).
22. Dagher, J. *et al.* Wild-type VHL Clear Cell Renal Cell Carcinomas Are a Distinct Clinical and Histologic Entity: A 10-Year Follow-up. *Eur. Urol. Focus* **1**, 284–290 (2016).
23. Dimitrakopoulos, C. *et al.* Network-based integration of multi-omics data for prioritizing cancer genes. *Bioinformatics* **34**, 2441–2448 (2018).
24. Dong, Y. *et al.* GIPC2 is an endocrine-specific tumor suppressor gene for both sporadic and hereditary tumors of RET- and SDHB-, but not VHL-associated clusters of pheochromocytoma/paraganglioma. *Cell Death Dis.* **12**, 444 (2021).
25. A Study of Abemaciclib in Combination With Sunitinib in Metastatic Renal Cell Carcinoma - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT03905889.
26. Turner, N. C. *et al.* Palbociclib in Hormone-Receptor-Positive Advanced Breast Cancer. *N. Engl. J. Med.* **373**, 209–219 (2015).
27. Bolck, H. A. *et al.* Cancer sample biobanking at the next level: combining tissue with living cell repositories to promote precision medicine. *Front. Cell Dev. Biol.* **7**, 246 (2019).
28. Sudlow, C. *et al.* UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).

Acknowledgments

First and foremost, I would like to thank my doctoral supervisors Holger Moch, Peter Scharml and Niko Beerenwinkel for providing me with the opportunity to carry out my work in their research groups and allowing me to pursue research interest outside of renal tumour biology. I would also like to thank Mitch Levesque for agreeing to be on my doctoral committee and for providing valuable feedback during my committee meetings. Throughout my doctoral studies, I have been fortunate enough to collaborate with many great academics on multiple projects on the topics of tumour biology, stress, angiogenesis, and virology; I would like to thank them all. Thanks to all the past and present members of the Computational Biology Group. It was always a pleasure coming up to Basel to see and speak to you all. Our conversations not only about our work and research interests but about all sorts of weird and wonderful things are ones I remember fondly. I would specifically like to thank Jack Kuipers for providing me with great little pep talks and support when needed; they provided me with valuable feedback and a sense of reassurance.

As someone who started her doctoral work at the same time as me, Sivia Angori and I have been through it all together. I could not have imagined someone more supportive to go on this journey with. I would like to thank her, as well as all the people in the Molecular Tumour Pathology and Translational Cancer Research group including Harini Lakshminarayanan, Hella Bolck, Dorothea Rutishauser, Ailsa Christiansen, Amir Banaei-Esfahani, Tulay Karakulak, Adriana von Teichman, Katharina Muehlbauer, Francesca Buja and Susanne Dettwiler. Another special mention to Fabiola Prutek who, through our shared sense of humour, became one of my closest and most supportive friends. It was a pleasure to share PATH E with the Weber group over all these years. They have some fantastic characters I am also proud to call my friends; specifically, the filterless Jasna Jetzer whose confidence, work ethic, and kind heart is something to be admired, Hanna Honcharova who not only looked over several histological slides for me but was also my partner in crime when it came to heading out on a Friday (or a random Tuesday) night, Anne-Laure Leblond who would leave you feeling amazing about life and the world following every conversation, and finally Marc Healy where our shared interest in football and science meant our quick catch-ups could last hours. I would also

like to thank Abdullah Kahraman and Phil Cheng who as clinical/translational bioinformaticians themselves within the USZ would provide me with valuable advice when I faced either technical or analytical issues. Working with them was always a pleasure, especially when organising the Clinics meets Bioinformatics/Data Science symposiums.

I owe a great deal to Ruben Casanova, Karina Salina and Cheng Guang (David) Wu. They were truly my family here in Zurich, feeding me a countless number of times and providing great advice and support throughout my doctoral studies. The more recent evenings spent with Ruben, Karina, and their beautiful baby boy Lukas have been been a welcome distraction during the most stressful of times. I would like to also thank all the people at my football and badminton clubs, those Tuesday, Wednesday, and Thursday nights, although broke me, were a joy. Specifically thanks to Patrick Toggweiler, Stefano Carta, and Gerhard Schratt, I always learnt something new over our post-game dinner and drinks undoing all the hard work we had done in the two hours prior. A big thanks to Johannes Bohacek, we were equally excited by the fact our scientific collaboration originated on the football pitch. It has been a pleasure to work with him. Thanks to Stefan Herklotz and Emmanual Jay who have become good friends of mine off the badminton court. I must thank my flatmates Oliver Sturman and Day-Z the dog, it was never boring at home after a long day at work. I must not forget to thank Meneka Rupasinghe or I will never hear the end of it. It feels like I have known her for a lifetime. Her upbeat personality and spontaneous nature means she is always a joy to be around. Her response to reading these few lines thanking her was "*Truer words have not been written*", which exemplifies her character and why we are such good friends. Our trip to Dublin with Mattia Privitera is one that I will not forget anytime soon!

My childhood friends are still among my closest friends today and they have supported me throughout my doctoral work. I have to thank David Ekunno; from our days in the nursery, eating sand I presume, to nights out in Vegas just a few years ago, I would not be the person I am today without his friendship. Massive thanks also to Gavin Pang, Husni Abboushi, Jack Mouslay, Neil Okeke, and Dapo Adedipe. Whenever I go back to London, they provide a great respite from my work while we all regress to being 16 again.

I am extremely thankful to Sabrina Tevere. She has provided a great deal of support for the past few years, especially over the pandemic and toward the latter stages of my doctorate. Her support, understanding, and caring nature has been incredibly touching and I am fortunate to have her by my side. *Grazie Bella!*

Finally, I must acknowledge and thank my incredible family, specifically my Mum and Dad. The sacrifices they have made and the challenges they have overcome to get me to a place where I can chase my dreams is something I am incredibly grateful for. I owe them everything. To list all the other members of my family will be the length of a dissertation in itself, however, I am incredibly grateful to them all for not only what they have done for me but also in support of my parents while I have been away.

Thank you all.

Curriculum Vitæ

Aashil A Batavia

Education

Aug 2017 – Jun 2022 **Dr. sc.** Computational Biology and Translational Cancer Research — ETH Zurich & USZ, Zurich
Dissertation: *Translational multi-omic assessment of clear cell renal cell carcinoma*

Sep 2015 – Jun 2016 **M.Sc. (Distinction)** Bioinformatics and Systems Biology — University of Manchester, Manchester
Thesis 1: *Applying an in silico fluctuation test to estimate mutation rates using Illumina sequencing data*
Thesis 2: *Insights into Prpf8 structure and function via the analsysis of its human variants*

Sep 2011 – Jun 2014 **B.Sc. (Hons)** Biomedical Sciences — University of Manchester, Manchester
Thesis: *Using in silico experimental evolution to understand mutation rate plasticity, evolvability and robustness*

Positions of responsibility & volunteering

Sep 2019 – Present **PhD Student Representative + Member** — Swiss Institute of Bioinformatics

Sep 2021 - Present **Organiser and Session Chair** — Clinics Meets Data Science, University Hospital Zurich

Jan 2018 - Present **Co-Founder** — Bioinformatics@USZ, University Hospital Zurich

Nov 2019 - Apr 2021 **Organiser and Session Chair** — Clinics Meets Bioinformatics, University Hospital Zurich

Sep 2015 – Jul 2016 **Co-Founder + Treasurer** — UoM Bioinformatics Society, University of Manchester

Jul 2012 – Aug 2012 **Games Maker** — London 2012 Olympics, LOCOG

Aug 2009 – Jan 2012 **Board Member** — Live, Learn, Love

Notable Presentations

Presentation + Poster, **[BC]² Basel Computational Biology Conference**, Basel, Switzerland, September 2021

Poster, **ISMB/ECCB 2021**, virtual, July 2021

Poster, **Keystone Symposia - Precision Oncology: Translating Discovery to the Clinic**, virtual, June 2021

Presentation + Poster, **Deutsche Gesellschaft für Pathologie e.V.**, Berlin, Germany, June 2020

Presentation, **ISCB RSG Turkey: Webinar Series**, virtual, May 2020, available online

Presentation + Poster, **Basel Life 2019**, Basel, Switzerland, September 2019

Poster, **ISMB/ECCB 2019**, Basel, Switzerland, July 2019

Poster, **14th Charles Rodolphe Brupbacher Symposium**, Zurich, Switzerland, July 2019

Poster, **17th Day of Clinical Research**, Zurich, Switzerland, April 2018

Employment History

Jan 2017 – Jun 2022 **Staff researcher** — University Hospital Zurich, Zurich

Jul 2014 – Jun 2015 **Artist Assistant** — Foster + Partners, London

Jul 2013 – Sep 2013 **Photographer** — Foster + Partners, London

Sep 2011 – Jun 2012 **Waiting on Staff** — Manchester United Ltd.

Jul 2011 – Sep 2011 **Executive Recruitment Consultant** — BPO Recruit

Publications

Authors who equally contributed to a publication are marked with a †

Moore AL†, **Batavia AA**†, Kuipers J, Singer J, Burcklen E, Schraml P, et al. *Spatial distribution of private gene mutations in clear cell renal cell carcinoma.* Cancers (Basel). 2021 Apr 30;13(9)

Kuipers J†, **Batavia AA**†, Jablonski KP†, Bayer F, Borgsmüller N, Dondi A, et al. *Within-patient genetic diversity of SARS-CoV-2.* BioRxiv. 2020 Oct 12;

Ohashi R, Angori S, **Batavia AA**, Rupp NJ, Ajioka Y, Schraml P, et al. *Loss of CDKN1A mRNA and Protein Expression Are Independent Predictors of Poor Outcome in Chromophobe Renal Cell Carcinoma Patients.* Cancers (Basel). 2020 Feb 17;12(2).

Ohashi R, Schraml P, Angori S, **Batavia AA**, Rupp NJ, Ohe C, et al. *Classic chromophobe renal cell carcinoma incur a larger number of chromosomal losses than seen in the eosinophilic subtype.* Cancers (Basel). 2019 Oct 3;11(10).

Ohashi R, Schraml P, **Batavia A**, Angori S, Simmler P, Rupp N, et al. *Allele loss and reduced expression of CYCLOPS genes is a characteristic feature of chromophobe renal cell carcinoma.* Transl Oncol. 2019 Sep;12(9):1131–7.

Wu C, Mairinger F, Casanova R, **Batavia AA**, Leblond A-L, Soltermann A. *Prognostic immune cell profiling of malignant pleural effusion patients by computerized immunohistochemical and transcriptional analysis.* Cancers (Basel). 2019 Dec 5;11(12).

Batavia AA, Schraml P, Moch H. *Clear cell renal cell carcinoma with wild-type von Hippel-Lindau gene: a non-existent or new tumour entity?* Histopathology. 2019 Jan;74(1):60–7.