Genomic Analysis of VHL Syndrome Renal Cancer - A Biological System for Probing Tumor Evolution

BY

ASIA DAWN MITCHELL

A DISSERTATION

Presented to
The Department of Molecular and Medical Genetics and the
Oregon Health & Science University
School of Medicine

In partial fulfillment, of
the requirements for the degree of
Doctor of Philosophy

November 2018

Table of Contents

List of Abbreviations

AA	Amino Acid
aCGH	array Comparative Genomic Hybridization
ALL	Acute Lymphoblastic Leukemias
APOBEC	Apolipoprotein B mRNA editing Enzyme, Catalytic Polypeptide-like
BAF	B-Allele Frequency
BAM	Binary Alignment Map
BAP1	BRCA1 Associated Protein 1
BED	Browser Extensible Data
bp	Basepairs
C	Cytosine
CADD	Combined Annotation Dependent Depletion
CAN	Copy Number Alteration
CBS	Circular Binary Segmentation
CCF	Cancer Cell Fraction
ccRCC	Clear-cell Renal Cell Carcinoma
Chr.	Chromosome
cm	Centimeters
CNR	Copy Number Ratio
COSMIC	Catalog of Somatic Mutations in Cancer
CpG	*5'—C—phosphate—G—3'*
CRC	Colorectal Cancer
dbSNP	Database of Single Nucleotide Polymorphisms
DNA	Deoxyribose Nucleic Acid
FAP	Familial Adenomatous Polyposis
GATK	Genome Analysis Toolkit
GC	Guanine-Cytosine
Gr.	Grade
GWAS	Genome Wide Association Study
HIF	Hypoxia Inducible Factor
iCAGES	integrated CAncer GEnome Score
ICGC	International Cancer Genome Consortium
ID	Identifier
IGV	Integrated Genome Viewer
ITH	Intratumor Heterogeneity
Kb	Kilo-basepairs
KIRC	Kidney Renal Clear cell carcinoma
LOF	Loss of Function
LOH	Loss of Heterozygosity

LP-WGS	Low Pass Whole Genome Sequencing
Mb	Mega-basepairs
MRCA	Most Recent Common Ancestor
mTOR	mechanistic Target Of Rapamycin
Nuc.	Nucleotide
OHSU	Oregon Health & Science University
OSCS	Osteopathia Striata Congenita with Cranial Sclerosis
PBRM1	Polybromo 1
PCR	Polymerase Chain Reaction
PDX	Patient Derived Xenograft
Pos.	Position
qPCR	Quantitative Polymerase Chain Reaction
SAM	Sequence Alignment Map
SCC	Squamous Cell Carcinoma
SCNA	Somatic Copy Number Alteration
SE	Standard Error
SETD2	SET Domain Containing 2
SIFT	Sorting Intolerant From Tolerant
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
sSNV	Somatic Single Nucleotide Variant
T	Thymine
TCGA	The Cancer Genome Atlas
TERT	TElomerase Reverse Transcriptase
TRACERx	TRAcking Cancer Evolution through therapy (Rx)
UV	UltraViolet
VAF	Variant Allele Frequency
VHL	Von Hippel-Lindau
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

List of Figures

List of Tables

Dedication

In memory of my charismatic, kind, and valiant cousin David Dupree Williams.

Acknowledgments

First and foremost, I would like to thank each patient who consented to our studies. Without their willingness to support research, we would lack the resources necessary to improve our foundational knowledge of cancer genetics. I would also like to thank our collaborators at the National Cancer Institute: Marston Linehan M.D., Cathy Vocke, and Christopher Ricketts Ph.D., for providing all patient samples, clinical data, and for being available to answer questions when needed. This work was initiated by Suzi Fei Ph.D., who has continued to serve as a wealth of knowledge and support during my graduate studies. Christopher Boniface provided essential laboratory support by preparing sequencing libraries for the exome resequencing and low-pass whole-genome sequencing experiments. I would also like to extend gratitude to my DAC members: Shannon McWeeney Ph.D., Mu-Shui Dai M.D. Ph.D., Andrew Adey Ph.D., and Brian O'Roak Ph.D., for providing direction in my research progression by providing necessary criticism, outside perspective, and alternative approaches to work through experimental challenges. None of this work would have been possible without the continued support of my PI, Paul Spellman Ph.D., who was always willing to let me explore different approaches. I would like to acknowledge my sources of funding that in part, supported my graduate training: NIH Ruth L Kirschstein T32 PMCB Training Grant #5T32GM71338-95 and Nicholas L. Tartar Research Fellowship #6123711F.

I must extend immeasurable gratitude and love to my friends and family for their unconditional support through the years. Specifically, thank you Justin for your patience and for growing with me. Last but not least, I thank my parents and my grandma Lenia for keeping me grounded in my heritage and providing security so that I always knew where I came from and where I could ascend.

Abstract

The process by which a normal cell is transformed in to a cancer cell and the steps that follow to drive malignant disease are of interest to both cancer biology and cancer medicine. Determining the cumulative genomic changes and their functional outcomes can provide insight into the evolution of a tumor. Challenges exist in obtaining empirical evidence for tumor evolution, largely due to the heterogeneity of tumors. Tumors from different primary tissues of origin exhibit variance in their genomic and functional architectures. Even tumors of the same subtype vary widely due to differences between each individual and their host environment. Intratumor heterogeneity is a product of tumor evolution, thus regions from the same tumor can also have genotypic variance. Many approaches in studying tumor evolution exist to partially mitigate some of this variance and to amplify the signal of tumor evolution. Our approach is to profile tumors that arise from the same host environment – the same person and tissue of origin. This characterization of malignancy include synchronous cancer in which multiple primary tumors are coexisting at the same time, and metachronous cancer in which multiple incidences of the same tumor type are separated by six or months. Synchronous and metachronous tumors are rare in sporadic cancer types. However, they are both hallmarks of the heritable cancer syndrome VHL Syndrome. In performing genomic analyses of nearly 200 synchronous and metachronous clear-cell renal cell carcinomas (ccRCCs) from over 30 individuals with VHL Syndrome, several notable genetic alterations gave insight into evolutionary trajectories and constraints of these tumors. These data identified driver mutations in a subset of high and low-grade ccRCCs and determined that clonal somatic mutations in specific cancer genes influence the progression of Fuhrman nuclear grade under a constrained host environment. Specifically, clonal inactivating *BAP1* and *PBRM1* mutations were observed to be determinants of high and low Fuhrman nuclear grade, respectfully. Additionally, linear evolution was observed at *SETD2* in a single high grade ccRCC

which contradicts previous observations of branched evolution at *SETD2* in sporadic ccRCC. Finally, observations of recurrent loss of whole chromosome 3 in two different patients, suggest that a combination of genotype and environment can apply selective constraints for specific somatic copy number alterations. These results extend our understanding of the evolutionary architecture and trajectories of tumors including how early genetic alterations can shape future phenotypes that provide prognostic value and provide evidence that the host environment can constrain evolution of the somatic tumor genome. This work also substantiates continued genomic studies of synchronous and metachronous ccRCCs from VHL Syndrome patients and provides framework for this biological system in future tumor evolution studies.

Chapter 1. Introduction

It is understood that cancer is the result of cumulative changes to the genomic and functional architecture of cells. Based on decades of research the community has arrived at a Darwinian model of tumor evolution where the fitness of cells in the tumor is a function of their evolved genetics in combination with their interaction with the host environment. As in most examples of Darwinian selection, the two major evolutionary processes are: (1) the positive selection of genetic alterations that are advantageous to the tumor; and (2) the negative selection of cells carrying deleterious genetic events through purifying selection. Cancer is largely driven by positive selection, such as the selection for nonsynonymous variants in driver genes (Martincorena et al., 2017). Purifying selection may not be a major feature of tumor evolution. However, purifying selection is a product of immunoediting where a competent host immune system acts as an extrinsic tumor suppressor system, reducing the cancer cell population (Shankaran et al., 2001).

Every alteration to the tumor genome does not always encourage growth or tumorigenesis. Are purely stochastic processes at play or are there underlying mechanisms that constrain the genetic landscape of the tumor? We and others have become deeply interested in how a tumor genome evolves in the context of a constitutive genome and a specific environment (that may or may not be static). Here, I summarize our current understanding of how the germline context can constrain tumor evolution, how high-resolution data improves tumor evolution models, and how current models of tumor evolution can be integrated to represent the complexity of cancer. I first discuss foundational models of tumor evolution informed by epidemiological data. Next, I discuss variation at the germline and somatic levels as it relates to tumorigenesis. I then dissect current models of tumor evolution, achieved through large-scale

genomic studies. Last, I discuss the constraints applied at the germline and somatic level and their ultimate consequences on tumor evolution.

1.1. 20th Century Models of Tumorigenesis

Models of tumorigenesis began from epidemiology in the 1950s and then evolved to cover genetic mechanisms in the 1970s, finally, adding molecular mechanisms in the 1990s. From these foundational principles, we have a clear theoretical and mechanistic schema for tumor evolution. This work now needs to be placed into context of genetic and environmental constraints.

Initial models of tumorigenesis relied heavily on the relationship between cancer incidence and mortality with age. In the early 1950s Nordling observed that the frequency of cancer was increasing to the sixth power of age, and therefore hypothesized that a single cell would require at least seven mutations for carcinogenesis (Nordling, 1953) and that these mutations should be successive. Building on Nordling's model of tumorigenesis, Armitage and Doll (Armitage and Doll, 1954) introduced the idea that tumorigenesis occurred in stages. They hypothesized that the number of stages was tissue or cancer type specific and that more than one mutation could occur within a single stage. Such simple models generalized tumorigenesis into a specific number of events. The Armitage-Doll multi-stage model was later reduced to a two-stage model of carcinogenesis that incorporated clonal expansion of cells with a selective advantage as the first-stage (Armitage and Doll, 1957). Clonal expansion results in the production of a uniform population of cells. The models proposed by Nordling, Armitage, and Doll, provided a framework for thinking about how multiple mutations and clonal expansion with positive selection could lead to cancer.

Nearly two-decades later, Peter Nowell (Nowell, 1976) proposed a model that included successive clonal evolution, inviting stochastic processes to tumor evolution models. Nowell's model suggested that sequential expansion of subclonal populations within the tumor were driven by genetic variation and positive selection. Clonal expansion rates were considered due to the application of Knudson's "two-hit hypothesis" (Knudson, 1971) model of tumor suppressor activation to adult tumors. Later, Moolgavkar (Moolgavkar, 1978) suggested that clonal expansion consisted of two rate-limiting steps (1) the mutation in a normal cell forming an intermediate, and (2) the transition from the intermediate to a cancer cell.

In the late 1980's, the rise in molecular genetic technologies directed focus towards particular genes for single tumor types. During this period, Volgelstein and colleagues (Fearon and Vogelstein, 1990; Vogelstein et al., 1988) introduced a four-stage model for colorectal cancer (CRC). The four stages proposed were: (1) normal mucosa transformation to aberrant crypt, (2) transition to early adenoma, (3) advancement to late adenoma, and (4) progression from adenoma to carcinoma. It was described that the mutation of four cancer genes in order: *APC*, then *KRAS*, then *DCC/SMAD4,* then *TP53*, drove progression between the four-stages of CRC evolution. This model became the driving paradigm for clonal Darwinian selection in cancer development for the next two decades (Fearon, 2011). Only recently are we appreciating that it is almost certainly more complex.

1.2 The genetic context

As we refine our models of tumor evolution, it is critical to consider that the development of a cancer, as an evolutionary process, is strongly influenced by the host genotype. Currently, this is informed by enormous genome wide association studies (GWAS) of tens to hundreds of thousands of participants. The germline genome and the somatic genome both contribute to the

genetics of a tumor and, together, are capable of forming complex interactions. The germline genome sets a base genetic environment that is homogeneous for all cells within the tumor. The somatic genome is ever changing, and is subject to Darwinian evolution, while under constraint by the germline genome (Swanton, 2015).

1.2.1. Normal germline variation

Germline genomic variation among individuals drives variation in phenotypes, including future disease risk. Germline variants must be compatible with organismal viability so they are unlikely to be disruptive in a diploid state. The effect of germline variants is context dependent, so a phenotype may only manifest in certain tissue types at certain stages of life. It is estimated that about 80% of *de novo* germline mutations are heterozygous and the remaining mutation fraction are homozygous but not deleterious (Xue et al., 2012). Heterozygous germline mutations are of particular interest in cancer studies because their phenotype has the potential to be unmasked by somatic copy-number alterations (SCNAs), and thus can constrain future somatic events.

1.2.2. Highly penetrant pathogenic germline variants

There are over 200 Mendelian heritable cancer syndromes which cause 5 – 10% of all cancer cases. These diseases are due to highly penetrant germline variants of genes that contribute directly to tumorigenesis (Glaire et al., 2017; Nagy et al., 2004). Nearly all hereditary cancer syndromes have an autosomal dominant mode of inheritance, meaning they are heterozygous at their causative locus. Phenotypes vary across cancer syndromes, but generally the following are more frequently observed in comparison to sporadic disease: earlier age of onset, increased incidence of multiple tumors, and rapid growth. Of importance, inherited cancer syndromes typically show distinct differences from most sporadic forms of the same disease.

Whether this is because there are distinct constraints in tumor development, the ordering of the drivers is distinct, or they are different diseases is of great interest.

1.2.3. Low penetrance pathogenic variants modulate cancer risk

Highly disruptive variants tend to be rare in populations and are quickly eliminated before reaching significant population frequencies. However, later age of onset can allow highly disruptive variants to persist at high frequencies within certain populations. Heritable and familial cancers occur at an earlier age than sporadic cancers. However, most incidences of heritable cancer still occur after the age of reproductive maturity (Brandt et al., 2008) allowing for germline mutations associated with cancer syndromes to persist in populations. BRCA1 and BRAC2 mutations are associated with hereditary breast and ovarian cancer (King et al., 2003). For example, in the general population the prevalence of having a BRCA1/2 mutation is 0.2% – 0.3%. The prevalence of having a BRCA1/2 mutation increases to 2.5% within Ashkenazi Jewish population (PDQ Cancer Genetics Editorial Board, 2018).

Slightly disruptive variants are able to persist across many individuals in a population and contribute to risk of disease. Hundreds of low penetrance cancer susceptibility variants have been identified through large GWAS. A GWAS of breast cancer cases have identified 94 susceptibility loci explaining 16% of the familial risk of breast cancer (Michailidou et al., 2015). A meta-analysis across six GWAS identified seven new risk loci for renal cell carcinomas (RCC), totaling 13 reported risk loci which explains about 10% of the familial risk of RCC (Scelo et al., 2017). Prostate cancers have over 100 susceptibility loci identified explaining one-third of the familial risk of disease (Eeles et al., 2017). The number of GWAS hits is proportional to the number of prostate tumors studied compared to other tumor types. Of note, the Oncoarray Consortium has genotyped nearly 450,000 tumor samples and as the largest cancer GWAS, aims to uncover more cancer

susceptibility loci (Amos et al., 2017). These studies rely on large cohorts of cases and controls in order to achieve the required power to detect variants of moderate to low effect size.

In some cases, germline variants offer protection against cancer by establishing an unfavorable genetic context that constrains biological and mutational processes critical to initiating tumorigenesis. The *WTX* gene encodes a repressor of WNT signaling and germline *WTX* mutations cause skeletal dysplasia in females (osteopathia striata congenita with cranial sclerosis, or OSCS). Conversely, *WTX* germline mutations are lethal in males. The *WTX* gene is somatically activated in 11-29% of Wilms tumors, yet OSCS patients do not get cancer. This suggests there exist temporal and spatial constraints on action of WTX during tumorigenesis (Jenkins et al., 2009). Another example of a germline constraint against cancer exist in Trisomy 21 patients. The overall cancer prevalence in Trisomy 21 is low – 1/10 of the general population rate (Yang et al., 2002). There are many tumor suppressors residing on chromosome 21, including the *DSCR1* gene that codes for an angiogenesis suppressor. One theory is that having an additional wild-type copy of multiple tumor suppressor genes offers a protective advantage against cancer by reducing risk of inactivating tumor suppressor genes.

1.2.4. The somatic genome

Somatic genomic variation is common and continuous. Each cell division creates a chance of acquiring new genetic alterations and somatic variants that persist must not be lethal to the cell. A few studies have aimed to quantify and characterize somatic variation rates in non-cancer cells. At estimates of 2.8×10^{-7} bp per division in fibroblasts (Milholland et al., 2017), $2\text{-}6 \times 10^{-6}$ bp per division in epidermis (Martincorena et al., 2015), and 4.4×10^{-6} bp per division in colonic epithelium (Nicholson et al., 2018); somatic mutation rates are one to two orders of magnitude higher than germline mutation rate estimates. As a cell accumulates somatic variants, there is

increased risk of acquiring one or more somatic variants providing a selective growth advantage and thus, creating the genetic context for tumorigenesis.

The Cancer Genome Atlas (TCGA) and related projects have significantly shaped our understanding of the somatic genomic features across many tumor types. There is enormous variation in the number of somatic mutations in a cancer genome ranging from 300 for lesions found in children to 3,000 in typical adult epithelial tumors to 30,000 in hypermutated tumors (Kandoth et al., 2013). The majority of somatic mutations do not contribute to the fitness of the tumor and are referred to as passenger mutations.

Driver mutations are those that provide an evolutionary advantage to the tumor. They can activate oncogenes or inactivate tumor suppressors. It has been estimated that a typical tumor carries 2-8 mutations in known driver genes (Vogelstein et al., 2013), although this number is uncertain (Martincorena et al., 2017). TCGA's pan-cancer analysis across twelve tumor types (Kandoth et al., 2013) and a later pan-cancer study of 21 tumor types by The Broad Institute (Lawrence et al., 2014) have identified a total of 160 driver genes across cancers. Reanalysis of TCGA data using different statistical approaches has identified 26 additional driver genes (Carter et al., 2017; Dunford et al., 2017). However, we have yet to obtain enough deeply sequenced cancer genomes to confidently identify all possible cancer driver genes.

Some cancer types and subtypes are characterized by higher mutation frequency in certain cancer driver genes which help explain subtype-specific tumor evolution. In breast cancer, *TP53* is the most frequently mutated gene with aberrations appearing in nearly 30% of all breast tumors. However, *TP53* mutations are much higher in the triple negative subtype, occurring in over 80% of these tumors (Bertheau et al., 2013). Point mutations in *TP53* are enriched in luminal tumors, while basal tumors carry insertion or deletion mutations. It is likely that the type of

mutation in *TP53* results in different functional consequences from the p53 protein, which may also dependent on the molecular subtype of the breast tumor (Nicholson *et al.*, 2018). In low-grade gliomas, prognosis is segregated by mutations in certain genes. Favorable prognosis is observed in tumors with IDH mutations, whereas poor prognosis is observed in tumors with *TP53* and *ATRX* mutations (Cancer Genome Atlas Research Network et al., 2015).

Cancer mutations are not solely single nucleotide events. Insertion and deletion (indels) mutations can consistent of any number of nucleotides. Indel discovery is challenged by constraints in sequencing experiments and downstream processing of mapping and aligning reads. Indels reported in COSMIC were more commonly somatic rather than germline mutations (Yang et al., 2010). Oncogenes exhibited a near eight-fold enrichment for tri-nucleotide indels compared to tumor suppressor genes (Yang et al., 2010). These observations largely suggest that indels are under positive selection within the somatic tumor genome as driver mutations in tumor suppressor genes. A recent study (Ye et al., 2016) reanalyzed 8,000 cases from TCGA and identified complex indels in cancer driver genes in 3.5% of cases that were previously missed or incorrectly annotated.

A high fraction of somatic variants are copy-number alterations (SCNAs). These are likely a result of genetic processes in cancer such as chromothripsis and chromosomal aneuploidy. SCNAs can be as large as an entire chromosome or as small as a single gene. Due to the large size of most SCNAs, they affect a larger fraction of the genome than any other type of somatic genetic alteration (Beroukhim et al., 2010). Specifically, aneuploidy, whole-chromosome and arm-level SCNAs are predominant alterations and observed in 90% of solid tumors (Weaver and Cleveland, 2006). Large genomic copy-number studies in cancer have determined the average tumor has about 24 gain and 18 loss events (Beroukhim et al., 2010; Zack et al., 2013). In these studies, the vast majority of frequent SCNAs in cancer do not harbor known oncogenes or tumor suppressors.

SCNAs beginning or ending with a telomere tend to be longer than internal SCNAs. Additionally, SCNAs are more likely to end or break within a centromere than expected given the length of the centromere (Zack et al., 2013). This is likely an effect of the mechanism for acquiring an SCNA.

Some have determined that the three-dimensional architecture of chromatin (Fudenberg et al., 2011) and other secondary structures of genomes (De and Michor, 2011) shape the distribution of SCNAs. This suggests that SCNA can be spatially regulated. Aneuploidy is a consequence of chromosomal instability (CIN), an increased rate of gain and loss of whole chromosomes. Mechanisms of CIN are mostly due to errors in mitosis. For example, prior to metaphase chromosomes can be misaligned or mis-attached. If the spindle assembly checkpoint fails to recognize either error, then the defective chromosomes will lag and it will be randomly segregated into either daughter cell (Sansregret and Swanton, 2017). One daughter cell would gain an additional chromosome, and the other would lose that chromosome. A recent study has observed inherent bias among individual chromosomes in mis-segregation events and aneuploidy rates (Worrall et al., 2018). Genome doubling events, observed in nearly 30% of advanced stage tumors, cause multipolar spindles to form and increase the risk of unbalanced chromosomal segregation (Bielski et al., 2018). They are one of the earliest mutational events and most often lead to later SCNA events.

1.2.5. Mutation Signatures

Mutations are dependent on the context of the sequence and the environment, which can constrain tumor evolution. Many features of a genomic locus can affect its mutagenicity, such as nucleosome occupancy, replication timing, recombination rate, retentiveness, transcription, or GC-content. Microsatellites are more frequently mutated due to their repetitive sequence and low-fidelity within repetitive regions by DNA replication polymerases (Brinkmann et al., 1998).

CpG dinucleotides are mutated 10-18 times more often than all other dinucleotides in the cancer genome (Rubin and Green, 2009). There is a two-fold increase in transition mutations versus transversions (Gojobori et al., 1982). Transcribed regions are mutated more frequently than non-transcribed regions (Park et al., 2012). DNA damage mutations can occur on the antisense strand during transcription because it is left exposed as a single strand, while the sense strand bound to transcription machinery is protected from damage.

Many of these mutagenic processes are so distinctive that they leave behind a particular mutation signature in the somatic genome. Over 80 mutation signatures have been identified in cancer and some are associated with clinical and biological features of tumors (Alexandrov and Stratton, 2014; Alexandrov et al., 2018, 2013, 2015). Kataegis is a mutational process that results in hypermutation of C > T and C > G within a localized region. It is observed in about 55% of breast tumors and associates with cases that are later onset and have better prognosis (D'Antonio et al., 2016). An APOBEC, family of cytidine deaminases, signature results in C > T and C > G mutations and is thought to contribute to kataegis (Roberts et al., 2013). An elevation in C > T mutations is associated with an "age" signature, which is the most common signature. Most tumors studied exhibit more than one mutational signature suggesting there to be more than one mutational process during the lifetime of a tumor (Gerstung et al., 2017).

1.3. Emerging Models of Tumor evolution

Recent models of tumor evolution build on most of the historical models previously discussed, describe cancer as an adaptive process. It is a process of replication, mutation and selection each with a small effect on cell growth (Beerenwinkel et al., 2007). Genotypic and phenotypic heterogeneity is fundamental to the Darwinian evolution of cancer. Advancements in genome sequencing technologies and tumor sample collection have provided higher resolution

data to generate improved models of tumor evolution. We can quantify the heterogeneity of the tumor in relation to its spatial and temporal environment in some cases.

1.3.1. The Branched Architecture of Tumor Evolution

The ordered relationship of genetic lesions between a clone and its subclones and the relationship of genetic lesions within a clone provide insight into the clonal architecture of tumors. Earlier models of tumor evolution lacked consensus on the importance of temporal order of mutations. Perhaps, due to an underappreciation and misunderstanding of intratumor heterogeneity (ITH). Phylogenetic trees are often used to illustrate ITH, where the trunk represents mutations observed in all cells and the branches represent the mutations present in subsets of the tumor. The more branches present in a tumor's phylogenetic tree, the more heterogeneity there is.

As clones diverge from a common ancestor, tumor evolution becomes increasingly more branched and ITH increases. In evolution speciation, some lineages are dead ends. Not all subclones have equal fitness as the environment changes and thus will be out-competed by other subclones. To observe the clonal architecture of a tumor, high resolution data is required. Single-cell sequencing, multi-regional sequencing, ultra-deep sequencing of bulk tumor, or sequencing of serial samples could all provide sufficient data.

Branched evolution has been observed in the many tumor types including lung (Zhang et al., 2014, 2017), clear-cell renal cell carcinoma (Gerlinger et al., 2012, 2014), breast (Yates et al., 2015), colorectal (Suzuki et al., 2017), and even some pediatric cancers (Anderson et al., 2011). In branched evolution, clones diverge from a most recent common ancestor (MRCA) and evolve in parallel, usually with minimal clonal competition. There is not always a relationship between the size of the observed clonal and subclonal fractions because mutational processes differ during the

lifetime of a tumor. Conversely, tumors driven by similar mutational processes do share similar evolution architectures.

Pediatric tumors carry far fewer mutations than tumors from adults. It might be argued that they undergo less evolution and yet branched evolution is observed (Anderson et al., 2011). Pediatric acute lymphoblastic leukemias (ALL) positive for the initiating *ETV6-RUNX1* gene fusion event were profiled at the single-cell level at 8 genomic targets to determine clonal architecture. The genetic architecture observed was diverse, subclones demonstrated branched evolution, and recurrent CNAs were not acquired in a particular order – that is there was no difference in oncogenic "potency" across these secondary CNAs. Some CNAs present in multiple subclones could arise independently, suggesting that some mutations were not a result of selection but were due to locus-specific genomic instability (Anderson et al., 2011).

1.3.2. Clonal Sweeping – Establishing the Dominant Clonal Genotype

Clonal evolution suggests that a series of clonal expansions grows to dominate the neoplasm, the replacements are called selective sweeps. Linear evolution, where driver mutations are acquired in a step-wise fashion, requires the presence of selective sweeps. The time to the next driver mutation must be longer than the time required for a clone to sweep through the neoplasm in order for evolution to occur. Clonal interference occurs when a second mutation occurs in a competitor clone and results in restricted expansion of both clones due to mutual competition. This is likely a common occurrence due to the large number of cells and high mutation rate observed in most tumors. Serial sampling studies could address this further. Several studies utilizing serial sampling techniques (Maley et al., 2006; Suzuki et al., 2015; Walter et al., 2012) determined that parallel clonal expansions occur before competition exists and that certain subclones will begin to dominate within the tumor. Clonal sweeping and interference is less

common in branched evolution because multiple clones have increased fitness and expand simultaneously (Davis et al., 2017).

Linear and branched evolution, while competition models, are not necessarily mutually exclusive. Consider that a tumor acquires many driver mutations and each driver mutation is under selection individually and some are under selection collectively. For example, reconsider Fearon and Vogelstein's early model of CRC (Fearon and Vogelstein, 1990). They argued that progress of CRC involved a step-wise or linear series of specific mutational events. This model still stands strong today. However, branched evolutionary architecture has also been observed in CRC (Suzuki et al., 2017). Suzuki et al. observed the genetic alterations from Fearon and Vogelstein's linear model of CRC (*APC*, *TP53*, and *KRAS* mutations) as truncal events of branched evolutionary trees. In this example, a single tumor is exhibiting both linear and branched architecture during discrete moments of its evolution. Its evolution is first linear then later becomes branched, presumably after undergoing many more cell divisions and mutational events.

Driver mutations are observed more frequently in tumors than would be expected given the normal background mutation rate. Driver mutations are associated with clonal expansion, occurring within the clonal fraction as well as being responsible subclonal lineages in later tumorigenesis (Gerstung et al., 2017). Driver mutations are found in regions of a gene that affects its function, frequently occurring within the coding regions of the genome, and are a type of mutation that alters the function of a gene (non-synonymous, frameshift, nonsense) (Lawrence et al., 2014). Most noncoding driver mutations are found within promoter regions and regulate the expression of genes important to tumorigenesis (Weinhold et al., 2014). An example are *TERT* promoter mutations. They are frequently found in bladder and central nervous systems cancers and increase expression of telomerase (Vinagre et al., 2013).

Bozic et al. studied glioblastomas and pancreatic tumors, and estimated that driver mutations gave an average fitness advantage of only 0.4% (Bozic et al., 2010). Bulk sequencing studies underestimate the clonal prominence of driver gene mutations. To better measure the effect of a mutant clone on selection, a longitudinal study with clonal size estimated at multiple time points is needed.

1.3.3. Punctuated vs Gradual Evolution

It is debated if tumors evolve gradually through a sequence of genetic changes and clonal expansions, or if the evolution is due to large punctuated changes possibly caused by a catastrophic event that causes multiple genetic lesions (such as chromothripsis). Punctuated evolution describes tumorigenesis stages as short bursts of events followed by clonal expansion. Events are discrete and with order, represented by bursts of phenotypic change and complex chromosomal rearrangements. On the other hand, gradual events are slow and with a consistent rate. Is punctuated evolution the dominant effect or does gradual evolution dominate? Studies answering this question, focus on genomic rearrangements such insertions, deletions, and copy-number alterations. This can change over time and, within or between clones. Punctuated and gradual evolution are not mutually exclusive (Figure 1). Single-cell studies along with serial sampling are needed to discern this.

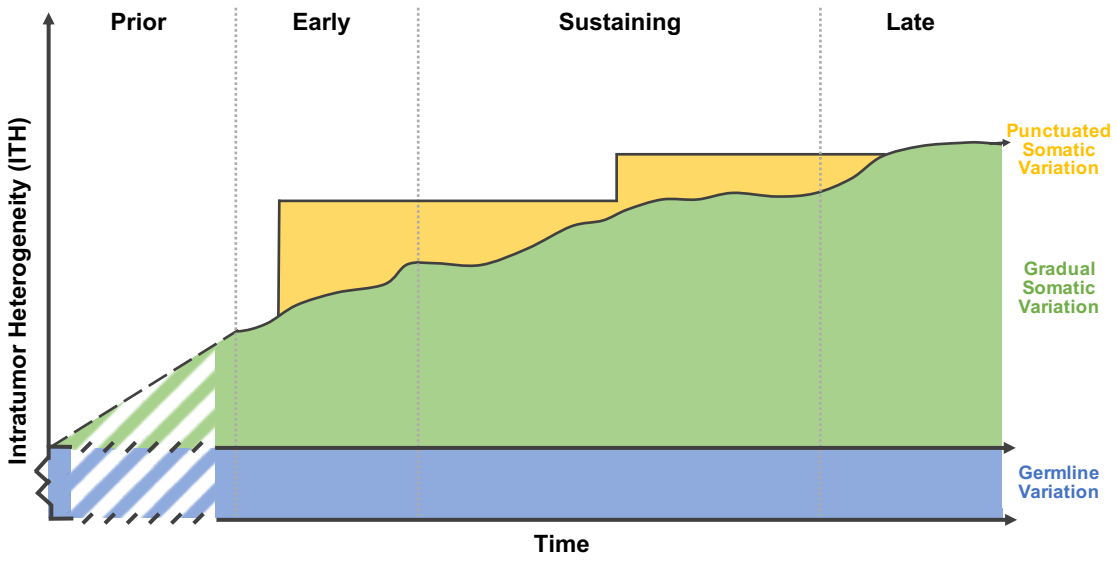


Figure 1: Mutational diversity landscape is shaped by punctuated and gradual somatic variation. Germline variation (blue) is constant and while it does not directly contribute to intratumor heterogeneity, it provides the context for subsequent somatic variation. Somatic variation can be categorized as punctuated or gradual. Gradual somatic variation (green) begins in early human development and has a relatively constant rate. However, disruption to biological processes that drive tumorigenesis can alter the mutation rate over time. Punctuated somatic variation (yellow) increases intratumor heterogeneity in large bursts within a short period of time. Clonal mutations will occur early in tumor development and will be present in all cells of the tumor. Subclonal mutations occur later in development and increase intratumor heterogeneity.

Chromosomal complexity, chromoplexy, is a hallmark of punctuated evolution. In a study of 57 prostate tumors, complex structural variation caused by breakpoint fusion bridge cycles was observed in successive cell divisions (Baca et al., 2013). Each tumor possessed multiple independent complex genomic rearrangements determined to occur in an ordered chain (Baca et al., 2013). The degree of chromoplexy altered phenotypes and simultaneously disrupted multiple known cancer driver genes. In prostate cancer, disruption of distinct cancer genes by chromoplexy was identified as an early stage followed by sustaining subclonal growth where deletion bridges were present. Deletion bridges are deletions that span two rearrangements and adjacent deletion bridges link multiple rearrangements in a chained fashion. This suggests that there must be multiple stages of chromoplexy in prostate tumors (Baca et al., 2013). Punctuated evolution was also observed in initiating events of triple negative breast cancer (Gao et al., 2016).

Certain mutation signatures like smoking and APOBEC, as well as genome doubling contribute to processes that support punctuated evolution. Exposures such as UV and smoking result in a higher accumulation of mutations, as observed in skin and lung cancers. Although the mutation rate is increased, the higher lifetime risk of lung cancers with increased pack-years and the decrease in lifetime risk upon quitting smoking supports gradual evolution (Blackford et al., 2009; Le Calvez et al., 2005). Primary driver genes of squamous cell carcinoma (SCC) of the skin including *NOTCH1*, *TP53*, and *FGFR3* were enriched for mutations and clone size in normal eyelid epidermis from a multiregional sequencing study (Martincorena et al., 2015). However, primary driver genes of basal cell carcinoma and melanoma carried few mutations. This represents positive selection for SCC driver mutations in normal eyelid epidermis. Interestingly, differential selection was observed in *NOTCH* genes. One individual carried a higher propensity for *NOTCH2* mutations across eyelid epidermis regions while another individual had a higher propensity for *NOTCH1* mutations (Martincorena et al., 2015).

From the analysis of whole genome sequence of synchronous tumors, the number of mutations positively correlated with age, which is representative of gradual evolution (Fei et al., 2016). In a multi-regional sequencing study of glioblastomas, there was a positive correlation between age and the number of clonal mutations, yet there was no correlation between age and the number subclonal mutations (Kim et al., 2015). This observation strongly suggests that mutational processes driving tumorigenesis at early versus later stages are not always the similar. Mutation timing was profiled in the International Cancer Genome Consortium's (ICGC) study of 2,658 tumors and their data showed that some mutation processes change activity over the lifetime of the tumor (Gerstung et al., 2017). For example, UV related signatures decreased in activity in later mutational stages, although, overall the average variation in mutation signatures throughout the lifespan was a modest 30%, it is still a topic requiring further studies.

1.4. Selective Pressures and Constraints

Germline and somatic genomes interact to drive tumorigenesis and thus produce modest, but real constraints on tumor evolution. Many studies within individual tumor types have identified germline pressures towards specific somatic copy number alterations driving tumor progression. Allele specific gains of the T91A allele of *AURKA* (Ewart-Toland et al., 2003; Hienonen et al., 2006) and losses of the A1176C allele of *PTPRJ* (Ruivenkamp et al., 2003) have been observed in human CRC. The rs6983267 SNP on 8q24 shows allele-specific imbalance in multiple genome wide association studies of CRC (Tuupanen et al., 2008). Breast tumors with germline *BRCA1* mutations have more frequent losses on 4p, 4q, 5q, Xp, and Xq, and gains of 10p and 16q compared to breast tumors without *BRCA1* mutations (Jönsson et al., 2005; Stefansson et al., 2009). Dworkin et al. studied aCGH copy-number profiles of 305 independent SCCs from 181 patients and identified 9 loci with allelic imbalance. One loci, 8q24, contained multiple SNPs

previously associated with an increased cancer risk and one SNP, rs13281615, showed evidence of allele-specific imbalance (Dworkin et al., 2010).

A recent integrative approach was applied to studying the interaction between the germline and somatic genomes within the collection of TCGA cases. Carter et al. identified 916 germline markers with strong associations with the tumor tissue type, and 62 associations between germline markers and protein-altering somatic mutations in cancer driver genes (Carter et al., 2017). Additionally, melanomas with germline *MC1R* mutations have a higher frequency of somatic *BRAF* mutations, and RAS/MAPK signaling oncogene that is mutated in two-thirds of melanomas (Landi et al., 2006; Maldonado et al., 2003). These interactions are the sorts of constraints that now need to be studied.

Another example of a constraint on tumor development is variation in essential drivers across different tissues. For example, germline mutations in *APC*, a tumor suppressor and regulator of cell-adhesion, cause Familial Adenomatous Polyposis (FAP) Syndrome (Crabtree et al., 2003) and result in colorectal tumors. Somatic mutation of *APC* is an early driver of sporadic CRC and is observed in over 76% of these tumors (Lawrence et al., 2014). von Hippel-Lindau (VHL) Syndrome is caused by a germline mutation in *VHL*, a tumor suppressor and negative regulator of angiogenesis. Loss-of-function mutations result in a 100% lifetime risk of developing multiple clear-cell renal cell carcinomas (ccRCC) (Prowse et al., 1997). *VHL* is a clear driver of sporadic clear-cell renal cell carcinomas, as it is altered in nearly two-thirds of these tumors (The Cancer Genome Atlas Research Network, 2013). Interestingly, *VHL* mutations are rare in other tumor types suggesting that *VHL* plays a pivotal role in ccRCC development (Kandoth et al., 2013).

Somatic mutations are random searches for solutions for fitness advantages given the constraints of the static germline genome and complex microenvironment. Resource, size, and

space limitations exist in the microenvironment. The doubling time of cancer cells in vitro, one to two days, is faster than the doubling time of tumors in vivo (60-200 days) (Chojniak and Younes, 2003; Furukawa et al., 2001; Nerli et al., 2014). Studies of circulating tumor DNA have determined *in vivo* doubling time of about two-weeks for clones containing driver mutations in early-stage primary breast cancer (Garcia-Murillas et al., 2015). One hypothesis is that the majority of cancer cells die before they can divide, likely due to the addition somatic mutations that are not sustainable given the constraints of the germline and microenvironment. The somatic "search" is much like a guess and check method. Some guesses may be cell lethal or may reduce fitness and be outcompeted by stronger clones. Or, perhaps the division rates between in vitro and in vivo conditions differ but cell viability remains similar.

Exogenous exposures introduce a source of selection to tumor evolution. Many chemotherapies, carcinogenic chemicals, and radiation exposures are genotoxic. Surviving cells must have mutations that provide protection against cell death and these mutations likely increase the malignant potential of these cells.

1.4.1. Endogenous Environment

Studying the environmental constraints of tumor evolution poses many challenges due to the complexity of the environment. One experimental approach is the application of patient derived xenograft (PDX) models. To create a PDX model, a collection of suspended-single cancer cells or a fragment of a tumor biopsy are removed from their natural environment and grafted into an immunocompromised mouse. These studies often lead to tumors that do not fully reproduce the features and heterogeneity observed in the original tumor sample. Eirew et al. examined how PDX models of breast cancer reflect the clonal composition of patient tumors and the clonal dynamics that exist when transplanting tumors. Tumors with the largest changes in

clonal structure became less heterogeneous in subsequent passages due to dominance of a few clones or reduction of subclonality (Eirew et al., 2015). Kreso et al. performed a similar study using a PDX model of CRC to understand if cells within a single genetic clone are functionally equivalent. They observed strong clonal selection after implantation, but maintenance of clonal architecture through serial passages (Kreso et al., 2013). Kreso and Eirew are similar studies but come to opposite conclusions, possibly due to the biological difference between breast tumors and CRC, differences in these specific tumors, or differences in the design of each study. The reproducibility of Eirew's study provides strong conviction that clonal dynamics are not stochastic and are largely influenced by the genetic makeup of clones.

1.5. VHL Syndrome – A biological model for probing tumor evolution

We can assume that the germline genome is largely constant over the lifespan of an individual. To fully understand the germline constraints on tumor evolution, one would need to study multiple tumors arising from an isogenic environment. To measure the influence of the host environment on tumor evolution we can consider VHL Syndrome.

1.5.1 Clinical Presentation of VHL Syndrome

VHL Syndrome is an autosomal dominant Mendelian disorder associated with germline mutations in the gene *VHL*. It has an incidence of 1 in 35,000 in the US and exhibits near complete penetrance by the age of 65 (Varshney et al., 2017). Most VHL Syndrome cases are inherited and 20% are *de novo*. The most significant phenotype is the formation of multiple benign and malignant tumors. Diagnosis is established if there is a family history of VHL Syndrome and a characteristic tumor presents, or if two or more characteristic tumors present.

Phenotypic heterogeneity in VHL Disease was first studied in 1991 (Neumann and Wiestler, 1991). There are two broad types of VHL Syndrome that differ in their functional

outcome of the *VHL* mutation (Glenn *et al.*, 1991; Gnarra *et al.*, 1994), thus affecting the phenotypic presentation. Type 1 individuals present with pheochromocytomas while Type 2 individuals do not. Both VHL Syndrome subtypes can present with clear-cell renal cell carcinomas (ccRCC), the phenotype that is the focus of this dissertation. Renal tumors are the only malignant tumor observed in VHL Syndrome and thus pose a significant health concern. Both subtypes can also present with hemangioblastomas of the central nervous system or the retina, pancreatic cysts, and neuroendocrine tumors (Varshney et al., 2017). Hemangioblastomas are vascular tumors that can grow large in size (3 – 30 mm in diameter) and are usually well tolerated unless their size causes significant pressure on the cerebrum (Slater et al., 2003). Renal tumors and hemangioblastomas of the central nervous system are the greatest risk to mortality in VHL Syndrome cases (Schmid et al., 2014).

1.5.2. Surveillance and Treatment

Given the multisystem involvement and the complete penetrance in VHL Syndrome, lifelong clinical screening is standard. Coordinated diagnostics, genetic registry, and recurrent screenings over the last 30 years have increased the life span of individuals with VHL Syndrome to at least 50 years of age (Maddock et al., 1996). Clinical surveillance is often coordinated across several medical specialties and screening measures are stratified by age (Lattouf et al., 2016; Schmid et al., 2014). Annual physical and ophthalmic screenings along with blood plasma molecular diagnostics for cerebral and retinal hemangioblastomas, and neuroendocrine tumors are suggested as early as age one to five years. Diagnostic imaging including abdominal MRI and CT scans are recommended annually beginning at age sixteen to detect renal tumors and reduce metastatic risk (Lattouf et al., 2016). A watchful waiting method is used for ccRCCs under three centimeters in the largest diameter because heritable and sporadic ccRCCs of this size have nearly

zero metastatic potential (Duffey et al., 2004; Walther et al., 1999). Routine surveillance and diagnostics are continued on the patient while all renal tumors remain under three centimeters in any diameter. Once any single tumor grows beyond this size, all visible tumors are resected at once using nephron sparing surgery and CT-guided focal ablation as preferred methods to radical (total) nephrectomy (Schmid et al., 2014). Preferred resection methods are minimally invasive and preserve as much normal kidney as possible. The unique high-level of surveillance in this disease provides a large volume of clinical data that can be used for collecting phenotypes across many timepoints during the evolution of a ccRCC.

1.6. VHL is a Negative Regulator of Hypoxia

The *VHL* gene is located on 3p25. It encodes the tumor suppressor VHL protein (pVHL) which functions as an E3 ubiquitin ligase complex (Gossage et al., 2015) to specifically target hypoxia inducible factor (HIF) for proteasomal degradation. HIF is a heterodimeric transcription factor involved in the regulation of angiogenesis, a necessary process for tumor growth (Benita et al., 2009). It is composed of one of three alpha subunits (HIF1α, HIF2α, or HIF3α) and HIF1β. pVHL is known to interact with both HIF1α and HIF2α. HIF1α is ubiquitously expressed while HIF2α expression is limited to endothelial, lung, renal, and hepatic tissues; but these two alpha subunits display different functionalities (Raval et al., 2005).

It is the disruption of the HIF2α and pVHL interaction that has been observed in renal carcinoma progression. Mouse and cell line models of ccRCC have determined active HIF2α is necessary for tumor growth under biallelic *VHL* inactivation (Chen et al., 2016; Kondo et al., 2003; Zimmer et al., 2004). Under normoxic conditions, pVHL complexes with adapter proteins elongin C and elongin B to ultimately bind cullin-2 (Pause et al., 1997) and complete the ubiquitin-ligase complex. This complex recognizes and binds hydroxylated HIF2α, targeting it for degradation and

preventing the formation of HIF heterodimers. Oxygen is a cofactor necessary for the HIF2α hydroxylation. Thus, under hypoxia HIF-alpha is not degraded and can complex with HIFβ. The HIF complex binds hypoxia-response elements to activate the transcription of pro-angiogenesis factors such as *VEGF*, erythropoietin, and PDGFβ. Somatic *VHL* mutations often fall with the elongin C and B or HIFα interaction domains (Gossage et al., 2015; Hoffman et al., 2001).

1.7. Clear Cell Renal Cell Carcinoma

In 2017, there were 65,000 estimated new cases of renal and pelvic cancers, and approximately 14,490 deaths (American Cancer Society, 2018). Approximately 75% of all renal cancers are of the clear-cell subtype. The rate of renal cancer diagnosis has steadily increased since 1975, while the number of deaths has remained constant. However, from 2005 to 2014 the rate of new diagnosis of renal and pelvic cancers has remained lower than the death rate (0.7% and 0.9% respectively). The five-year relative survival rate has increased over the last four decades. While, these trends suggest there to be improvements in early detection and treatment, it is still of public health concern to understand the features of tumors that determine prognosis.

1.7.1. Genetic Features of ccRCC

In 2013, TCGA published results of the genetic landscape from the largest cohort of ccRCC. 91% of tumors had loss of the p-arm of chromosome 3, and 9% of these tumors also had loss of 3q (whole loss of chromosome 3). Four of the five most frequently mutated ccRCC tumor suppressor genes are located on 3p, including VHL (The Cancer Genome Atlas Research Network, 2013). With the exception of VHL, these 3p driver genes are all chromatin remodelers or modifiers. Loss-of-function mutations or gene silencing through methylation in VHL occur in over 60% of ccRCC tumors and are not indicative of driving any other type of malignant disease.

VHL is rarely mutated in tumor types that are not observed in VHL Syndrome (The Cancer Genome Atlas Research Network et al., 2013). It is the most frequent driver of ccRCC and follows a traditional tumor suppressor mechanism of activation. Multiregional studies have identified a somatic VHL mutation to be clonal (Gerlinger et al., 2012, 2014), and thus occur early. Loss of 3p is also clonal but theorized to occur after the VHL mutation. Copy neutral LOH has also been observed in a minority of ccRCCs (The Cancer Genome Atlas Research Network, 2013).

Common co-morbidities for ccRCC include smoking, obesity, hypertension, chronic renal disease, and viral hepatitis. Interestingly, there exist a 2:1 male-bias in ccRCC incidence that is unexplained by disease risk factors. A recent study (Ricketts and Linehan, 2015) recently stratified mutation profiles of 628 ccRCC, across 3 datasets, by sex to determine the root of this sex-bias. Of the genes residing on the X chromosome, the mutation rate in *KDM5C*, a significant driver of ccRCC, was higher in males. Other ccRCC drivers, including *PBRM1* and *BAP1*, exhibited a sex-bias in mutation rate. *PBRM1* mutations were more frequent in males, while *BAP1* mutations were more frequent in females. *PBRM1* and *BAP1* are both located on chromosome 3p, within a LOH region observed in over 90-percent of ccRCCs. Additionally, *KDM5C*, *PBRM1* and *BAP1* are all chromatin remodelers or modifiers. Therefore, part of the sex-bias observed in renal cancers can be attributed to genetic factors. Other factors of this sex-bias might include hormonal differences between males and females.

Approximately 2% of ccRCC tumors have germline loss-of-function VHL mutations, resulting in VHL Syndrome. Fundamental biology and pathology of the tumors are similar, thus there is no reason to suggest they should be treated differently. Due to high surveillance and recurrent resection surgeries in VHL Syndrome, ccRCCs from these individuals tend to be biopsied at a lower grade and stage. Additionally, VHL Syndrome individuals acquire tumors at a much

younger age and the age signature observed in heritable ccRCCs is less pronounced (Fei et al., 2016).

1.7.2. Molecular Features of ccRCC

Evolution is dependent on prior steps that constrain future evolutionary routes. Early constraints can drive convergence towards specific phenotypes. A multiregional sequencing study of the clonal architecture of somatic mutations across a pair of synchronous renal tumors identified convergence towards mTOR pathway dysregulation in both tumors (Fisher et al., 2014). Consistent with these observations, PI(3)K/AKT/MTOR pathway is dysregulated in about 23% sporadic ccRCCs (The Cancer Genome Atlas Research Network, 2013). However, Fisher et al. did not find evidence for branched evolution, the evolution identified within their previous work in sporadic ccRCC (Gerlinger et al., 2012, 2014). Additionally, BAP1-mutant ccRCCs exhibit overexpression of mTOR pathway components (Peña-Llopis et al., 2012). mTORC1 has been observed to interact with HIF2α and the activation of mTOR and HIF pathways are positively correlated (Elorza et al., 2012). The connection between mTOR and HIF2α is interesting because HIF signaling is also aberrant in ccRCC (Chen et al., 2016; Shen et al., 2011). As a result, HIF and mTOR pathways are targets for therapeutic intervention in ccRCC (Chen et al., 2016; Cho et al., 2016; Pantuck et al., 2007).

1.7.3. Pathobiology features of ccRCC

Several renal cell types may have the capacity to give rise to ccRCC including proximal tubular epithelial cells and other epithelial nephrons (Droz et al., 1990; Paraf et al., 2000). The name “clear-cell” is derived from the histological features; Specifically, the cytoplasmic lipid and glycogen deposits which dissolve during histological processing and leave the appearance of clear cytoplasm (Frew and Moch, 2015). The high lipid content of these tumors gives a yellow

appearance at the surface. Some ccRCCs can be cystic due to necrosis or neoplastic cysts (Eble and Bonsib, 1998). Cystic ccRCCs contain a reduction of solid tumor tissue and may represent an alternate pathway in ccRCC initiation (Montani et al., 2010).

Of all pathological features of ccRCCs, the Fuhrman grade provides the most effective metric of metastatic potential. The Fuhrman nuclear grading system is used on all renal tumors to categorize by the size and shape of the nucleus, and the nucleolar prominence (Fuhrman et al., 1982). The majority of VHL Syndrome ccRCC tumors are detected at Fuhrman grade 2 and are three centimeters or less in the largest dimension. However, if monitoring is delayed, tumors may advance to higher Fuhrman grades and larger size. Tumors where any single dimension is under 5.5 cm diameter are considered the optimum actionable size for high survival, regardless of patient age (Ficarra et al., 2005). High Fuhrman grade correlates with increased ability for metastasis and lower survival (Nishikimi et al., 2011; Sukov et al., 2012).

1.8 Genomic Analysis of VHL Syndrome ccRCC

There have been a handful of genomic studies aimed at determining somatic genetic architecture of heritable synchronous ccRCCs relative to sporadic tumors. Frequent loss of 3p was observed in a comparison of copy-number profiles across ccRCCs from 90 total heritable cases, sporadic cases, and cell lines (Beroukhim et al., 2009). Copy-number profiles among the three groups were heterogeneous and variants were highly fragmented within the cell lines compared to the patient samples. Generally, more SCNAs were observed in in sporadic tumors than in heritable tumors. When comparing synchronous tumors, breakpoints at 3p were inconsistent, which supports the theory that each synchronous tumor was independent in origin.

Recently, I participated in the analysis of whole-genome shotgun sequencing of 40 synchronous ccRCCs from six VHL syndrome patients (Fei et al., 2016). All tumors were

determined to be independent primary tumors with no clonal relationship. Over 90% of all sSNVs were tumor specific (Figure 2A) and most of the sSNVs shared between two tumors from the same patient were determined to be kidney specific (Figure 2B). We can extend this result to VHL Syndrome renal cancer and generalize that all (or most) tumor initiate and evolve independently while under the same host environment.

Consistent with sporadic ccRCC, we observed recurrent loss of 3p in the tumors of all six patients (Figure 3A). However, patient Green presented with loss of whole chromosome 3 in ten of thirteen tumors from one patient. The loss of whole chromosome 3 in ten of thirteen tumors from one individual is extremely non-random (p-value of $2.1*10^{-8}$) and highly significant. It remains unclear why in some individuals with ccRCC, whole chromosome 3 loss occurs, but in most individuals with ccRCC loss of 3p is observed. What is clear, is that there exists a personal bias in somatic alterations in tumors, supporting convergence towards specific somatic genotypes.

In this dissertation, I further examine genomic features of VHL Syndrome ccRCCs, including reanalysis of previously studied cases, additional cases from previously studied patients, and new cases. In two studies, I aim to illustrate how the germline and somatic environments can constrain the evolutionary paths of tumors. The first study asks the question: if somatic driver events shape tumor grade divergence and convergence. The second study is concerned with the observation of non-random loss of chromosome 3 and germline pressures on SCNAs. Together, these studies should demonstrate the wide potential of VHL Syndrome ccRCCs as biological system for probing tumor evolution.

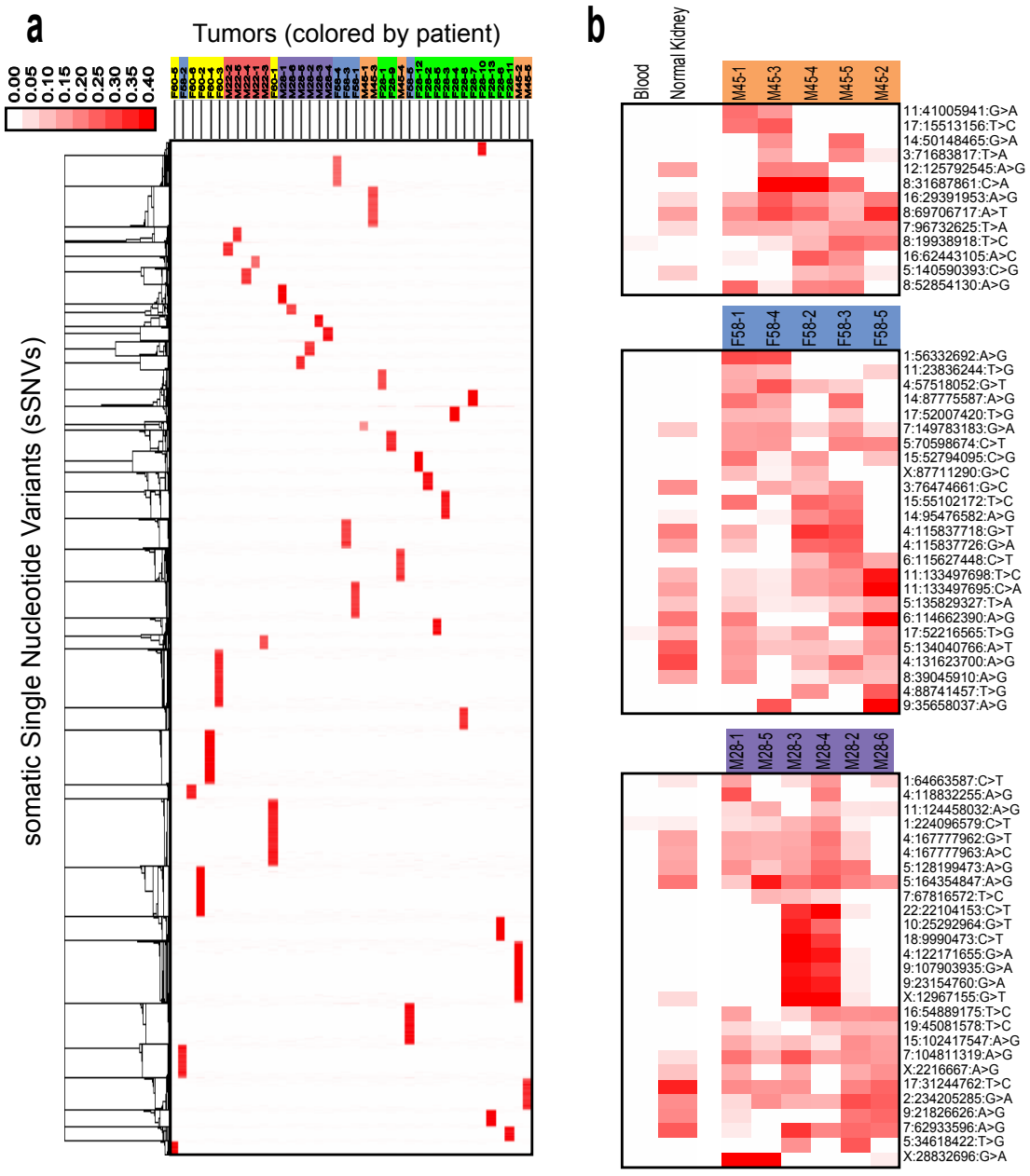


Figure 2: Somatic SNVs reveal no clonal relationship between tumors.[1] **A**) A heatmap of all private sSNVs (variants unique to a single tumor). 90.2% of the 100,677 sSNVs called were private. Each column represents one tumor and tumors from the same patient are similarly colored. Rows represent variants and the variant allele frequency (VAF) is represented by the shading (highest VAFs are in red and lowest VAFs are in white). Hierarchical clustering was performed across both columns and rows. **B)** Heatmaps of all sSNVs shared between two or more tumors from the same patient. The three patients with both kidney and blood normal samples are shown. Most shared variants were present in the kidney normal.

[1] "Figure 2: The somatic SNVs revealed that each tumour was independent from the other tumours." by Fei et al. is licensed under CC by 4.0.

a
somatic Copy Number Variants (sCNVs)
M22-1
M22-2
M22-3
M22-4
M22-5
M45-1
M45-2
M45-3
M45-4
M45-5
F60-1
F60-2
F60-3
F60-4
F60-5
F60-6
F28-1
F28-2
F28-3
F28-4
F28-5
F28-6
F28-7
F28-8
F28-9
F28-10
F28-11
F28-12
F28-13
F58-1
F58-2
F58-3
F58-4
F58-5
M28-1
M28-2
M28-3
M28-4
M28-5
M28-6
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
b
somatic Single Nucleotide Variants (sSNVs)
13:98578836:C>T
14:73834002:T>C
2:102821999:G>A
X:97320672:C>T
17:5651221:A>G
13:64134511:A>G
13:64134510:G>A
5:115663272:T>A
9:114938837:C>T
10:33358412:T>C
2:20436196:C>T
2:20436193:T>C
15:24619317:C>T
21:16620288:T>C
15:31104021:A>G
4:70046987:A>G
1:149697579:C>T
8:28831333:T>G
4:131405866:T>C
15:28688189:T>C
22:21644452:T>C
4:79439596:A>G
5:2160589:C>A
11:58519757:T>C
6:32717136:G>C
6:32717137:T>C
8:61103498:T>A
15:94745594:T>A
7:32186899:A>G
17:34855515:T>C
4:48537609:C>T
5:66623586:G>T
4:79544686:T>G
F28-2
F28-12
F28-1
F28-7
F28-6
F28-4
F28-5
F28-9
F28-10
F28-3
F28-11
F28-13
F28-8

Figure 3: Somatic copy number alteration profiles corroborate tumor independence and expose patient-specific variation.[2] A) A heatmap of copy-ratios depicting somatic copy number alterations (SCNAs) across all 40 tumors. Each row represents a different tumor with patient color codes. Blue represents a loss event, while red represents a gain. Shading of SCNAs is due to the log2 of the copy-ratio and can be influenced by the tumor purity and clonality of the SCNA. **B)** A heatmap of all sSNVs shared among tumors from patient Green. Rows represent each sSNV and the shading designates the VAF of that sSNV. Each column is a different tumor. Columns shaded darker green have loss of whole chromosome 3, while lighter green columns have loss of 3p. Hierarchical clustering was performed across both rows and columns.

[2] "Figure 4: Copy number variants confirmed that the tumours are independent but showed within-patient patterns.' by Fei et al. is licensed under CC by 4.0.

Chapter 2. High grade VHL Disease clear-cell renal cell carcinomas are associated with the presence of clonal driver mutations

Asia D. Mitchell, Suzanne S. Fei, Christopher J. Ricketts, Christopher Boniface, Cathy D. Vocke, W. Marston Linehan, Paul T. Spellman

All data presented in this chapter were analyzed by Asia Mitchell. The experimental design was led by Asia Mitchell with assistance from Paul Spellman. Suzanne Fei and Christopher Ricketts consulted on the initial hypothesis and early-stages of the experimental design. Previous data obtained and published by Suzanne Fei provided the foundation for this study. Christopher Boniface prepared the DNA libraries for sequencing. Cathy Vocke and Marston Linehan provided all patient materials, including DNA and clinical data. All figures and text were composed by Asia Mitchell.

2.1 Abstract:

VHL Syndrome is a rare genetic disorder with a dominant mode of inheritance, which predisposes carriers to benign and malignant neoplasms in a variety of tissues, including renal cell carcinomas. Individuals with VHL Syndrome, who develop renal cancer, commonly present with multiple, malignant clear-cell renal carcinomas (ccRCC). These are synchronous tumors, meaning they initiate, mature and evolve, at the same time and within the same germline environment. We previously published the whole genome sequences of 40 ccRCCs from six VHL Syndrome patients. While these genomes were sequenced deeply enough to call most clonal mutations, they lacked sufficient depth to definitively identify subclonal single nucleotide variants in known cancer driver genes. To determine if there were missed driver mutations in these tumors we performed exome resequencing on five high-grade and five low-grade synchronous ccRCC tumors from three VHL syndrome patients obtaining 131X median coverage. Resequencing provided sufficient power to call variants across all known cancer driver genes. We applied methods to assess clonality within these tumor samples to improve detection of true drivers of tumor evolution with respect to Fuhrman grade. We show that the high-grade tumors were elevated for nonsynonymous driver gene variants. In one patient we observed linear evolution at *SETD2* in a high-grade tumor. In tumors from the same individual, we observe clonal loss-of-function *BAP1* mutations in two high-grade tumors and a clonal loss-of-function *PBRM1* mutation in one low-grade tumor. Together our observations suggest that clonal somatic mutations in known cancer genes are determinants of Fuhrman grade in VHL Syndrome ccRCCs.

2.2 Background

Clear-cell renal cell carcinoma (ccRCC) is the most common subtype of renal cancers, and accounts for three-quarters of the nearly 50,000 new annual renal cancer cases. Prognosis is largely associated with biological aggressiveness of the tumor. The most widely used method for classifying renal tumors is the Fuhrman nuclear grading system. This method categorizes renal tumors based on the microscopic morphology of the nucleus and nucleolar prominence (Ficarra et al., 2005; Fuhrman et al., 1982). High Fuhrman grade correlates with increased ability for metastasis and lower survival (Nishikimi et al., 2011; Sukov et al., 2012) and is the most effective indicator of a tumor's metastatic potential.

There are distinct genetic features of ccRCCs that are well characterized from large-scale genomic studies (The Cancer Genome Atlas Research Network, 2013). Inactivating mutations or epigenetic modifications in the tumor suppressor gene *VHL* are observed in two-thirds of sporadic ccRCCs (The Cancer Genome Atlas Research Network, 2013). Biallelic inactivation of *VHL* is a known early driver of ccRCC (Gerlinger et al., 2012, 2014), but tissue specific biallelic inactivation of *VHL* is not sufficient to produce ccRCC in mouse models of ccRCC (Haase et al., 2001; Hou and Ji, 2018). *VHL* is located on 3p as are *BAP1*, *PRBM1* and *SETD2*, which are three of the other tumor suppressor genes commonly mutated in ccRCC. Loss of 3p is observed in over 90% of all ccRCCs. *BAP1* and *PBRM1* mutations are mutually exclusive in sporadic ccRCCs and *BAP1* loss of function is associated with higher tumor grades (Peña-Llopis et al., 2012). Additionally, loss of *BAP1* is observed before and during metastatic progression (Eckel-Passow et al., 2017) and correlates with larger tumor volume (Minardi et al., 2016).

Approximately 2% of individuals diagnosed with ccRCC have VHL syndrome, caused by inherited loss-of-function (LOF) mutations in *VHL*. VHL syndrome associated ccRCCs often

manifest as multiple tumors at an early age. In VHL Syndrome carriers who are screened for ccRCCs, the majority of these tumors are detected at Fuhrman grade 2 and are 3 cm or less in the largest dimension. However, if monitoring is delayed or a tumor is particularly aggressive, tumors may advance to higher Fuhrman grades and larger volumes.

We previously published the whole-genome sequences (WGS) of 40 synchronous ccRCCs from six VHL syndrome patients (Fei et al., 2016). Over 90% of all somatic mutations called were tumor specific. Further, half of the mutations that were shared between at least two tumors from the same individual were identified as kidney-specific. These data show that synchronous tumors in VHL syndrome are not clonally related and are independent primary tumors.

Sporadic ccRCCs have been subjected to multi-regional analyses, where distinct parts of the same tumor have been analyzed independently with the aim of understanding evolutionary trajectories of tumors (Gerlinger et al., 2012, 2014; Mitchell et al., 2018; Turajlic et al., 2018b, 2018a). These analyses have determined that sporadic ccRCCs can have regional clonal architecture and some variants that are nearly clonal in one region are not observed in other regions. These multiregional studies also show that the majority of variants predicted to be clonal in one region are observed (and clonal) in other regions assessed.

VHL Syndrome provides one of the few human systems to study the evolution of multiple tumors in the same genetic background. It has been hypothesized that evolution is dependent on prior steps that constrain future evolutionary routes and lead to divergent genotypic and phenotypic features that can have significant clinical value (Swanton, 2015). Here, we provide a comparison of somatic driver events in Fuhrman grade 2 and 3 synchronous ccRCC tumors and discuss how these mutations impact the evolution of the tumor.

2.3 Methods

Tumor samples and DNA extraction

Patient phenotypes and other clinical data were obtained from the National Cancer Institute Clinical Research Information System or patient charts. This study was approved by the Institutional Review Board of the National Cancer Institute (IRB study 16626). All patients provided written informed consent. Sample procurement details and DNA sequencing methods were described in Fei et al. 2016.

Statistical Analysis of Phenotypes

Clinical phenotypes and mutation counts based on WGS were tabulated for all 40 tumor samples. Features included: tumor grade, age at tumor resection, ellipsoid tumor volume, total somatic single nucleotide variant (sSNV) count from WGS, and tumor purity estimate. To determine relationships between features and tumor grade, we performed multivariate logistic regression. Tumor grade was binarized, where low-grade tumors (grade 2) were 0 and high-grade tumors (grade 3) were 1. Features containing continuous or non-zero data was normalized by median centering. The logistic regression model was fit across normalized features using Scipy Stats module in Python.

Exome Sequencing

A subset of 10 tumors were selected for exome sequencing and Aligent SureselectXT All Exon V5 (cat 5190-6208) was used to perform the exome DNA library capture. We sequenced these exomes using Illumina HiSeq at the Oregon Health & Science University Sequencing Core with a median coverage of 75X. All sequencing data files and analyses were performed on the OHSU Exalab compute cluster.

Exome Sequence Alignment

Paired-end reads were aligned to the hg19 human reference genome using bwa-mem (Li and Durbin, 2009). Output SAM files were converted to BAM files, sorted, and indexed using Samtools v0.1.17 (Li et al., 2009). MarkDuplicates, part of Picard Tools v1.51 (McKenna et al., 2010), was used to mark duplicate reads generated during the PCR amplification stage. Fine-tuning of the alignment was performed using GATK Best Practices workflow. For each sample, local positions to target for realignment were called using RealignerTargetCreator and then realigned using IndelRealigner. Quality scores were then recalibrated using BaseRecalibrator and PrintReads, which bins reads based on the original quality score, the dinucleotide, and the position in the read. A second round of realignment and base recalibration was applied to each tumor-normal pair, to improve alignment around germline variants.

Post-alignment Processing

Whole-exome sequencing (WES) and WGS BAMS were further processed to extract only reads overlapping to Agilent exome probes. BAM files were processed using BEDtools (Quinlan and Hall, 2010) and a BED file containing hg19 genomic positions for all exons. For each sample, we merged both the WES-exon-only BAM and the WGS-exon-only BAM then performed another round of local realignment and base recalibration.

SNV calling and filtering

Somatic single-nucleotide variants were called for each tumor-normal pair using MuTect2 (Cibulskis et al., 2013; Poplin et al., 2018). To reduce false positives, we performed two filtering steps: sample-level and dataset-level. Sample-level filtering considered each sample independently. Called sSNVs were discarded if they had fewer than 14 reads in the tumor, fewer than 8 reads in the normal, power ≥ 0.8 to detect mutation with allele fraction ≥ 0.3 at site, present

in DbSNP (but not present in COSMIC), and did not pass each of the 14 integrated quality filters in MuTect2. Data-set level filtering considered all samples collectively. sSNVs were also discarded if they were suspected to be a single-nucleotide polymorphism (SNP) due to presence in normal alignments of other patients. They were also discarded if the region contained other and different variants across two or more samples from different patients. The neighboring alignment of all variants was manually reviewed in IGV and variants residing in poorly aligned regions were discarded.

SNV Annotation

Oncotator was used to annotate variants as synonymous or nonsynonymous, and to annotate genes [Ramos et al., 2015]. Synonymous variants were filtered out and not included in further analysis. The candidate driver list of 260 genes was curated from three publications: The Cancer Genome Atlas (TCGA) ccRCC (The Cancer Genome Atlas Research Network, 2013), TCGA Pan Can – 12 tumor types (Kandoth et al., 2013), TCGA 21 tumor types (Lawrence et al., 2014). All genes listed as significant drivers were included and reduced to a unique list of genes. Nonsynonymous variants were assessed for functional impact by using three approaches:

(1) Missense variants were assessed using five bioinformatics tools: MutationAssessor (Reva et al., 2011), PolyPhen2 (Adzhubei et al., 2010), CADD (Kircher et al., 2014), SIFT (Ng and Henikoff, 2003), and iCAGES (Dong et al., 2016). Each tool uses different methods and evidence to make their predictions of functional impact. MutationAssessor considers evolutionary conservation of the locus across protein homologs and has been successfully applied to several cancer studies (including TCGA). PolyPhen2 considers multiple sequence-based and structural-based features of the variant allele and at the locus. CADD and iCAGES incorporate results of other

variant assessment tools as features for training machine learning classifiers. CADD also utilizes data from functional genomic studies as well as conservation data. iCAGES specifically assesses cancer variant and incorporates data from cancer mutation databases such as COSMIC and TCGA. Finally, SIFT considers evolutionary conservation at the locus and the changes to physio-chemical properties of amino acids. Variants were classified as "loss-of-function" using the following thresholds: a functional impact score ≥ 1.9 from MutationAssessor; a phred score ≥ 15 from CADD; a prediction of "possibly damaging" or "probably damaging" by PolyPhen2; a prediction of "damaging" by SIFT; and a prediction of "damaging" by iCAGEs. Variants determined to be damaging by three or more of the aforementioned tools were marked as candidate loss-of-function variants.

(2) Splice site variants were assessed using NNSPLICE 0.9 (Reese et al., 1997), a tool that uses an artificial neural network to predict structural changes leading disruption of the splice site. This tool gives a metric from zero to one that represents the strength of the splice site (values closest to one indicate strong splicing). Splice site metrics from NNSPLICE 0.9 were previously computed by Piva et al. for all known mutations in *BAP1*, *PBRM1*, and *SETD2*. We used these metrics to assess any splice site variants in these three genes (Piva et al., 2015).

(3) All variants were compared against variant curation databases ClinVar (Landrum et al., 2014) and COSMIC (Forbes et al., 2017), as well as peer-reviewed literature. If a variant was found in ClinVar, the clinical significance, assertion method, and supporting evidence were all considered. If the clinical significance was listed as "pathogenic" or "likely pathogenic" then the variant in question was marked as a candidate pathogenic variant. If the variant was found in COSMIC or in published

literature, the cancer type and mutation frequency were considered. If the variant was previously observed in a renal cancer, then it was marked as a candidate pathogenic variant.

CNV Calling and Annotation

WGS datasets were used to determine segmented copy number variant sites base on read depth using BIC-seq v1.1.2 (Xi et al., 2010) as described in Fei et. al 2016. Genotypes were previously called using Unified Genotyper (Poplin et al., 2018). To identify copy-neutral loss of heterozygosity, genotype VCFs were used with the scatter plot function from CNVKit (Talevich et al., 2016) to visualize and detect sites of allelic imbalance overlapping regions lacking read depth-based copy number variation.

Tumor purity estimation

Tumor purity was estimated using filtered synonymous and nonsynonymous variants derived from WGS, segmented copy number data, and the ABSOLUTE (Carter et al., 2012) method as described in Fei et al. 2016.

Cancer cell fraction calculation

The cancer cell fraction (CCF) is a ratio of the average number of mutations per cell and the expected number of mutations per cell and was approximated using PyClone (Roth et al., 2014). PyClone requires filtered SNVs, minor and major integer copy number, and tumor purity as input. Prior genotypes weights were derived based on the assumption that mutations at (i) diploid loci were almost always heterozygous, (ii) haploid loci were effectively homozygous, and (iii) triploid loci were a mixture of majority AAB and minority ABB, where A is the mutant allele. The highest weight was always applied to heterozygous genotypes. The binomial model was run using 10,000 iterations, an error rate of 0.001, and burn-in of 1,000.

2.4 Results

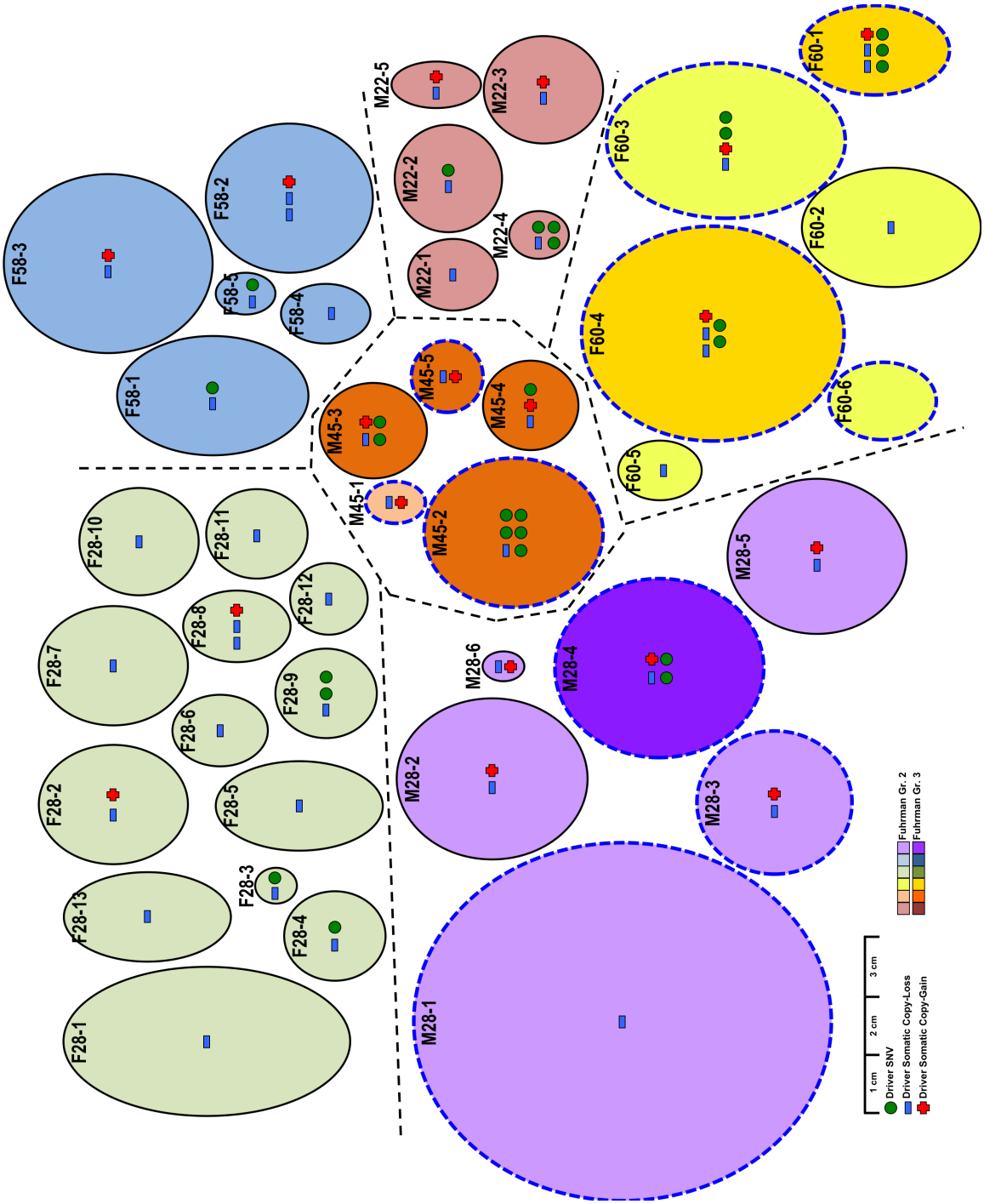
Few candidate drivers discovered in WGS

We identified non-synonymous mutations in coding sequences from our previously published WGS of 40 VHL syndrome ccRCCs from six individuals (Fei et al., 2016). We filtered SNVs as described in the methods and identified variants residing within known cancer genes from three sources: TCGA ccRCC working group (n=17) (The Cancer Genome Atlas Research Network, 2013), TCGA PanCan working group (n=114) (Kandoth et al., 2013), and a secondary TCGA PanCan assessment across 21 tumor types (n=260) (Lawrence et al., 2014). Few coding variants were found in known cancer genes and only 14 out of 40 tumors (35%) had somatic mutations in any of these genes (Figure 1A). Median coverage across these genes was 39X with tumor purities ranging from 0.25 to 0.75. This sequencing depth only provided sufficient power to detect somatic SNVs (sSNVs) at 82% of loci across known cancer genes (Table 1).

In all but one case, the inherited null allele of *VHL* was revealed by either the partial or complete loss of chromosome 3 encoding the wild-type allele, or copy-neutral loss of heterozygosity (LOH) (Figure 1A-B). Full somatic copy-number profiles for each tumor are available in chapter 1 figure 3. Only 5 out of 40 tumors (12%) contained mutations in driver genes identified as significantly mutated by TCGA ccRCC analysis. These five tumors had mutations in *MTOR* (one mutation), *PBRM1* (one mutation), *BAP1* (two mutations in two tumors), and *SETD2* (two mutations in one tumor).

Tumor ID	Power (drivers)	Power (ccRCC)
M22-1	0.88	0.82
M22-2	0.88	0.80
M22-3	0.93	0.90
M22-4	0.85	0.76
M22-5	0.88	0.85
M45-1	0.90	0.89
M45-2	0.88	0.82
M45-3	0.89	0.84
M45-4	0.78	0.74
M45-5	0.89	0.87
F60-1	0.89	0.82
F60-2	0.85	0.80
F60-3	0.91	0.80
F60-4	0.88	0.84
F60-5	0.67	0.61
F60-6	0.76	0.74
F28-1	0.93	0.92
F28-2	0.91	0.89
F28-3	0.88	0.80
F28-4	0.69	0.61
F28-5	0.85	0.80
F28-6	0.89	0.82
F28-7	0.78	0.72
F28-8	0.69	0.67
F28-9	0.92	0.89
F28-10	0.80	0.74
F28-11	0.84	0.67
F28-12	0.80	0.67
F28-13	0.93	0.90
F58-1	0.88	0.80
F58-2	0.74	0.55
F58-3	0.88	0.82
F58-4	0.88	0.84
F58-5	0.89	0.76
M28-1	0.87	0.76
M28-2	0.89	0.85
M28-3	0.88	0.84
M28-4	0.87	0.80
M28-5	0.87	0.80
M28-6	0.89	0.87

Table 1: Median power to detect clonal mutations across all driver gene positions and across ccRCC driver genes (ccRCC) in 40 tumors with WGS


A
F28-1
F28-13
F28-4
F28-3
F28-5
F28-2
F28-6
F28-7
F28-8
F28-9
F28-12
F28-11
F28-10
F58-3
F58-1
F58-4
F58-5
F58-2
M22-5
M22-3
M22-2
M22-4
M22-1
M45-5
M45-4
M45-3
M45-1
M45-2
M28-6
M28-4
M28-2
M28-1
M28-3
M28-5
F60-5
F60-4
F60-3
F60-6
F60-2
F60-1
1 cm
2 cm
3 cm
Driver SNV
Driver Somatic Copy-Loss
Driver Somatic Copy-Gain
Fuhrman Gr. 2
Fuhrman Gr. 3

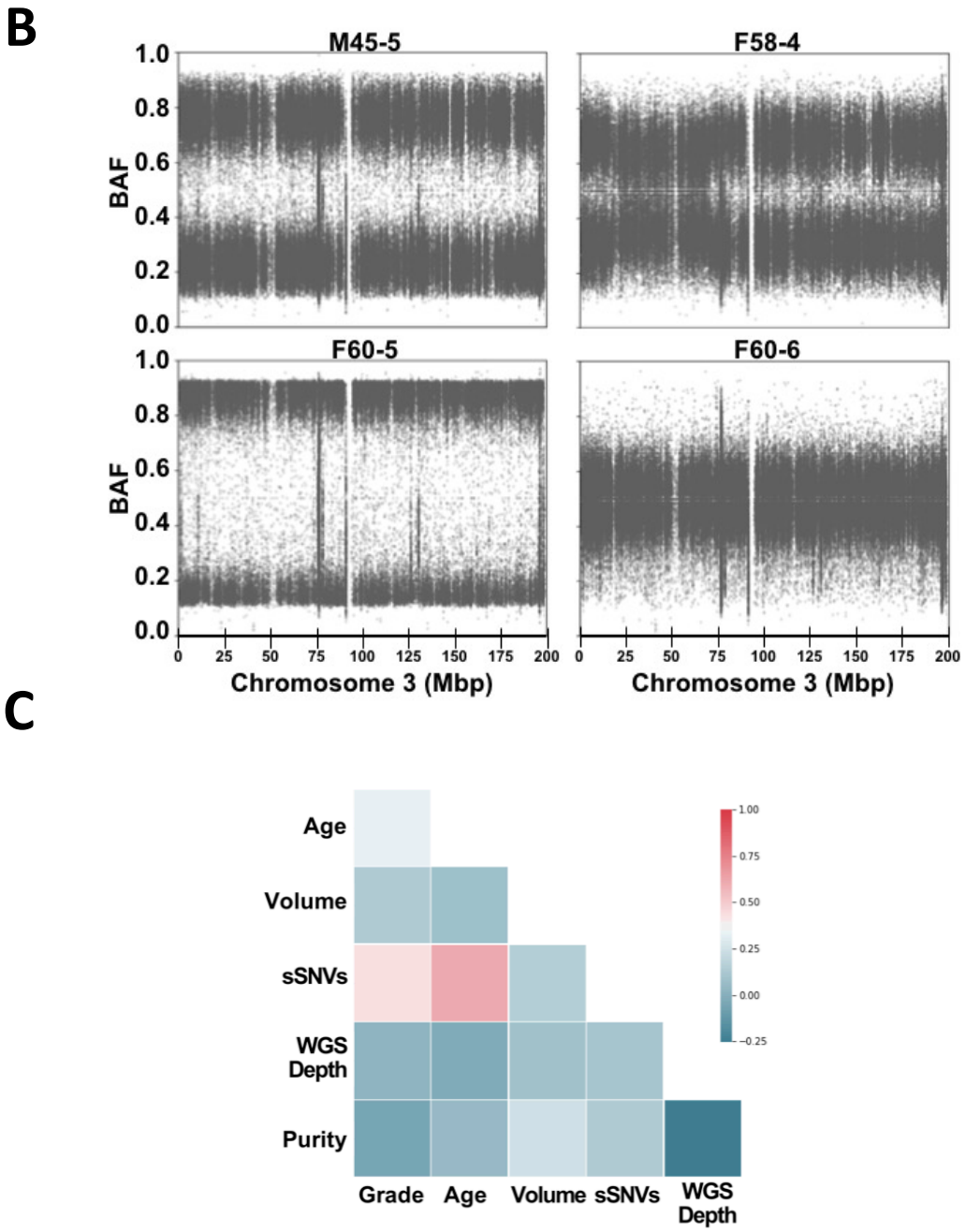


Figure 1: Candidate somatic driver mutation events called from WGS in 40 VHL Syndrome ccRCC tumors. A) Each oval represents an independent tumor drawn to scale using the two largest dimensions measured after surgical resection. Tumors from the same patient are the same color, and tumors of the same Fuhrman grade are similarly shaded. Somatic mutation events are annotated within each oval: green circle represent candidate driver mutations, red plus signs represent a copy-gain of 5q, and blue minus signs represent a copy-loss of 3p and/or 14q. Tumor ovals with a blue dotted outline were further profiled using WES. **B)** B-allele frequency (BAF) plots for chromosome 3 for all fours tumors lacking copy-number LOH from Fei et al. 2016. **C)** Spearman rank correlation matrix of various features of across the tumors. Correlated features are represented in red.

We grouped tumors by phenotypic features to identify any relationships between phenotypes and to recurrent or driver genomic events (Figure 1C). We confirmed the previously reported positive correlation between patient age and total number of sSNVs (Fei et al., 2016). No obvious pattern could be observed between the size of the tumors and the acquisition of cancer driver sSNVs. Additionally, there was no correlation with tumor size or grade. We did identify a small positive correlation between tumor grade and total number of sSNVs (Figure 1C). Upon performing multivariate logistic regression analysis, we determined that the total number of candidate driver variants is positively influenced by the total number of sSNVs.

We observed nonsynonymous mutations in known driver cancer genes in 6 out of 7 (86%) of the Fuhrman grade 3 tumors, compared to 8 out of 33 (24%) of the Fuhrman grade 2 tumors. Additionally, 6 out of 7 (86%) of the Fuhrman grade 3 tumors had gain of chromosome 5q in comparison to 11 out of 33 (33%) of the Fuhrman grade 2 tumors. These results suggest that increasing Fuhrman grade is associated with an accumulation of cancer associated genomic alterations in these VHL Syndrome ccRCCs.

Merged WGS and WES

To improve our power to detect variants across all loci within known cancer driver genes, we performed whole exome sequencing (WES) on a subset of tumors from our cohort. We selected five Fuhrman grade 2 and five Fuhrman grade 3 tumors from the three patients with both high and low grade tumors and obtained a median WES coverage of at least 70X. The WES and WGS genomes were independently aligned (methods). For each tumor-normal pair, the exome portion of the WGS alignment was extracted and merged with the WES alignment (Figure 2). Local realignment around known indels and base recalibration were both performed on the merged tumor-normal pairs prior to calling variants using MuTect2.

As expected, deeper WES uncovered additional mutations in cancer driver genes in both grade 2 and grade 3 VHL ccRCC tumors (Figure 3A). The median power to call clonal variants at candidate gene locations improved by ≥ 9% in all ten tumors and ranged from 96% - 100% (Table 2). The addition of WES revealed four additional candidate driver gene mutations with variant allele frequency (VAF) between 10% and 15% (Figure 3A) that passed our variant filters.

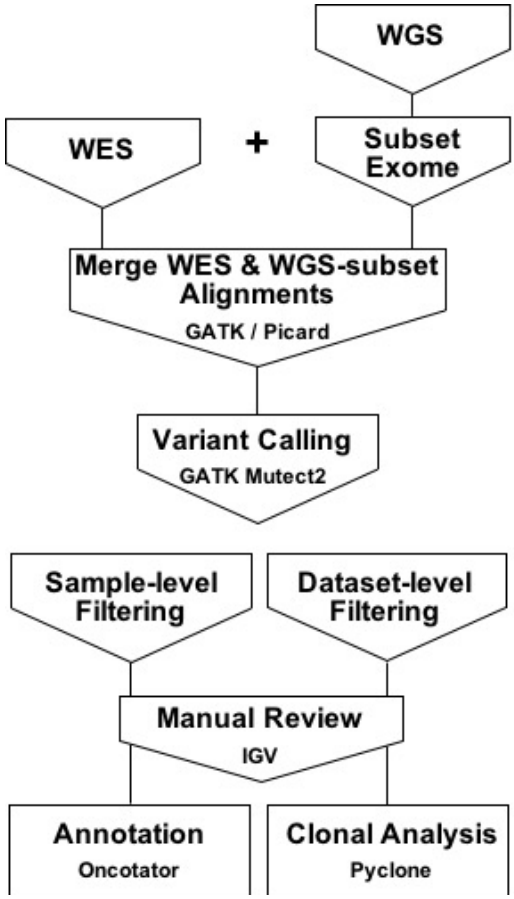


Figure 2: Overview of the bioinformatic workflow design to obtain merged genome and high confidence variant calls.

ID	WGS			Merged (WGS + WES)		
	Median Depth	% >=50X	Median Power	Median Depth	% >=50X	Median Power
M45-B	52	54%	N/A	114	96%	N/A
M45-1	45	24%	90%	124	96%	99%
M45-2	42	20%	88%	133	97%	99%
M45-5	43	19%	89%	134	96%	99%
F60-B	43	18%	N/A	150	99%	N/A
F60-1	43	19%	89%	193	99%	100%
F60-3	46	33%	91%	134	99%	100%
F60-4	43	22%	88%	131	97%	100%
F60-6	34	1%	76%	128	97%	100%
M28-B	39	8%	N/A	134	96%	N/A
M28-1	38	7%	87%	111	93%	99%
M28-3	44	28%	88%	123	96%	99%
M28-4	40	15%	87%	107	93%	96%

Table 2: Median power and depth to detect clonal mutations across candidate driver gene positions in WGS and Merged genome

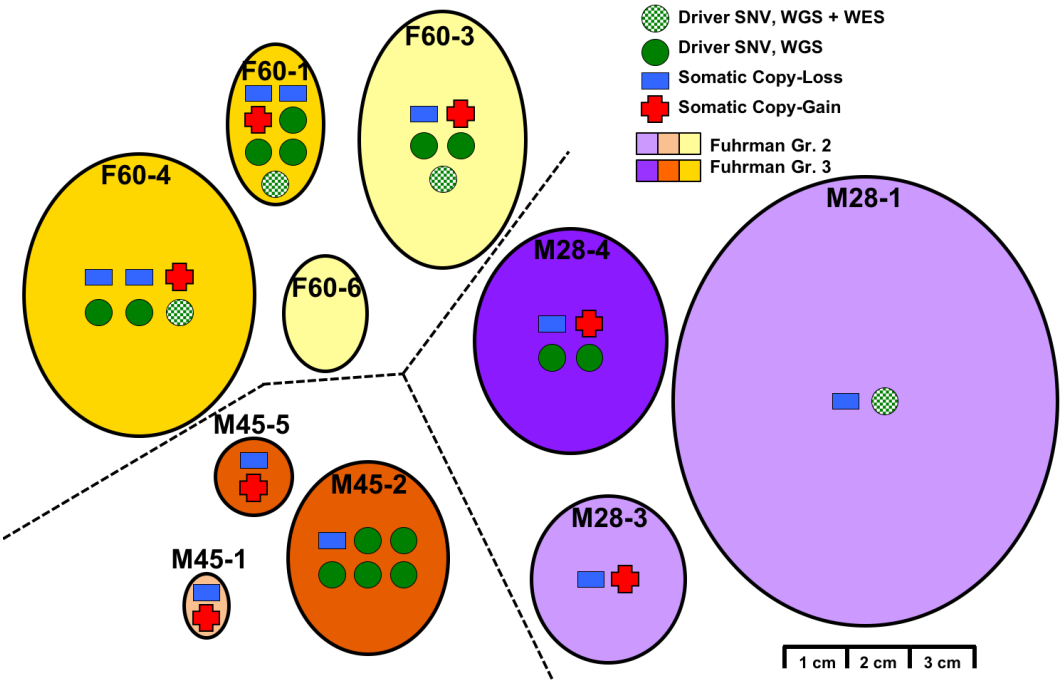


Figure 3: Candidate driver events identified in VHL Syndrome ccRCCs merged genomes. Overview of candidate driver mutations and CNVs across 10 tumors. Data are represented as described in figure 1. Candidate driver sSNVs identified in the merged genome are represented by hashed interior green circles.

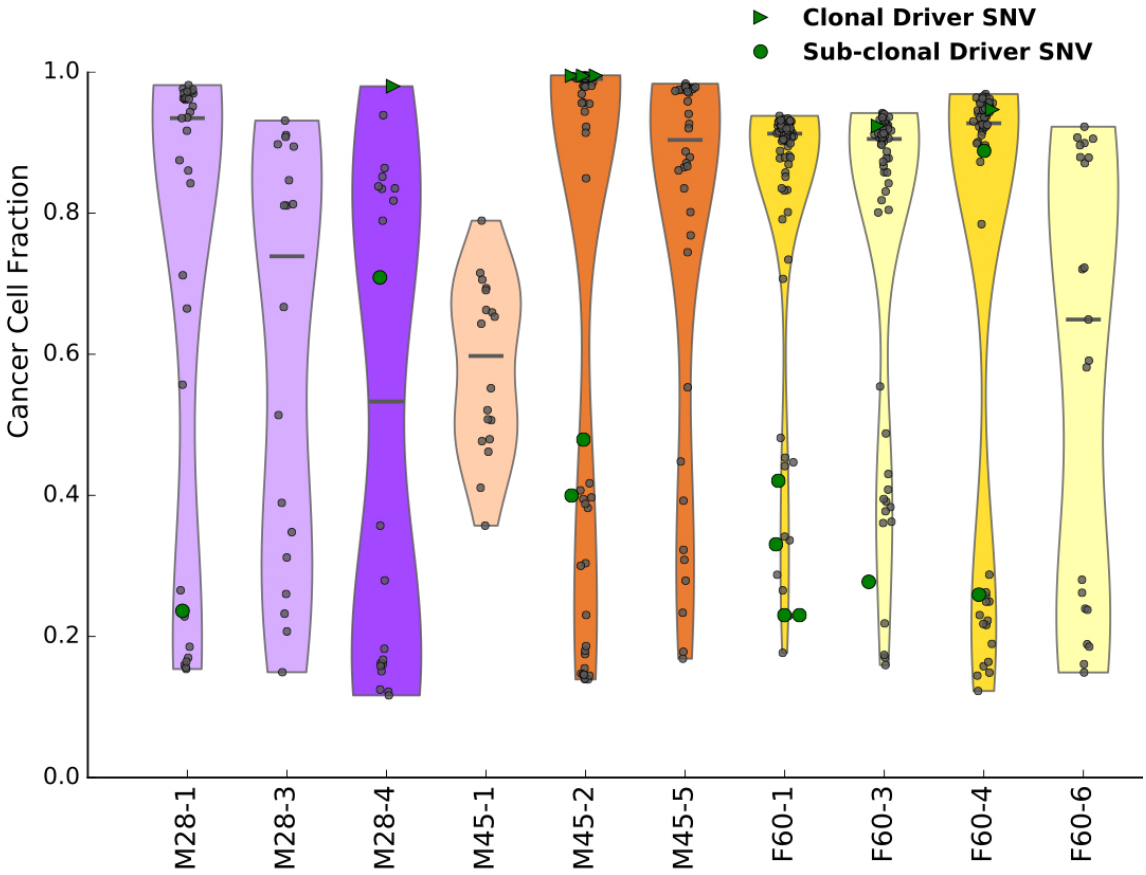


Figure 4: Distribution of cancer cell fraction (CCF) for all nonsynonymous variants in each tumor. Coloring of the distribution density represents is patient specific and shading represents the tumor grade (darker shades are high grade, while lighter shades are low grade tumors). Green triangles represent clonal candidate driver variants, while green circles represent subclonal candidate driver variants. Dark-grey points denote clonal passenger variants, CCF ≥ 0.9, and light-grey points denote subclonal passenger variants.

ID	Gr.	Purity	Total WES Variants	% SNVs	% Indels	% Nonsynonymous	% Clonal
F60-1	3	0.64	622	72.3%	27.7%	28.8%	78.9%
F60-3	2	0.55	429	82.8%	17.2%	25.4%	77.0%
F60-4	3	0.74	435	80.7%	19.3%	26.7%	71.4%
F60-6	2	0.67	303	75.2%	24.8%	22.8%	33.3%
M28-1	2	0.83	268	83.2%	16.8%	28.0%	39.4%
M28-3	2	0.71	234	80.3%	19.7%	18.8%	22.2%
M28-4	3	0.65	225	80.0%	20.0%	25.8%	9.1%
M45-1	2	0.27	254	82.7%	17.3%	26.0%	16.7%
M45-2	3	0.82	477	80.1%	19.9%	26.8%	58.3%
M45-5	3	0.52	320	77.8%	22.2%	24.4%	65.8%

Table 3: Variant type frequencies

Functional Impact & Clonal Architecture

To understand the role of each candidate driver mutation, we assessed the clonality and the functional impact of each variant. Functional impact was determined by taking the consensus output of five variant assessment tools as outlined in the methods. To approximate the clonal architecture of these tumors, we utilized PyClone to calculate the cancer cell fraction (CCF) for all nonsynonymous variants (Figure 4). The CCF is a ratio of the average number of mutations per cell and the expected number of mutations per cell. Given that our study used bulk tumor samples, our CCF values represent an average across the region sequenced.

Four of the five high-grade tumors carry candidate driver variants and in three of these high-grade tumors, all candidate driver variants detected were clonal. One low grade tumor contained clonal candidate driver variants, while three out of five low-grade tumors did not contain any candidate drivers. CCF distributions of nonsynonymous variants (Figure 4) do not show trends by patient, grade, or size. No trends were observed when considering the size of mutation (SNV vs indel) or the rate of nonsynonymous variants (Table 3). It should be noted that the trunk size (the fraction of clonal mutations) tended to be higher in the high-grade tumors (Table 3). The trunk size of F60-3, a low-grade tumor, was more similar to the trunk sizes of the two high-grade tumors, F60-1 and F60-4, than to that of the low-grade tumor F60-6.

Four of the ten exome sequenced tumors contained TCGA ccRCC significantly mutated genes. Both high grade tumors from patient F60 contained KIRC drivers. Notably, F60-1 contained a single nucleotide deletion in exon 13 of *BAP1* determined to cause loss-of-function and F60-3 contained a 23 bp insertion also in *BAP1*. The F60-3 *BAP1* insertion occurred four bp upstream of the most frequent causative allele of BAP1 tumor predisposition syndrome. Additionally, COSMIC reported four renal cancers reporting mutations at this allele. Therefore, we determined both BAP1 mutations to be pathogenic. One low-grade tumor F60-3 contained a splice-site mutation

in *PBRM1*, near exon 16, determined to disrupt normal splicing and cause loss-of-function of *PBRM1* transcripts. M45-2 contained two *SETD2* mutations (Table 4). One mutation was determined to be a clonal missense mutation and predicted to be a loss-of-function mutation. The second *SETD2* mutation was determined to be a subclonal protein truncating mutation and also predicted to result in loss-of-function. We observed loss-of-heterozygosity at the *SETD2* locus, so both *SETD2* mutations are likely on the same allele. In multiregional sequencing studies of sporadic ccRCC (Gerlinger et al., 2012, 2014) multiple independent *SETD2* mutations have been observed in several ccRCC tumors and provide evidence for parallel evolution.

Remarkably one tumor, F60-6, lacked any identifiable copy number variant on 3p, 5q or 14q and did not present any candidate driver variants. However, the total sSNV count, 1430, [Fei et al 2016] of this tumor was consistent with another low-grade tumor (F60-5) from the same patient with similar tumor purity and dimensions. One high-grade tumor M45-5 lacked any candidate driver variant. However, this tumor did contain copy-neutral LOH of whole chromosome 3 (Figure 1B) and gain of 5q (Figure 1A).

ID	Gr.	Gene	Chr.	Pos.	Nuc. Change	AA Change	LOF	Depth	VAF	CCF
M28-1	2	*NBPF1*	1	16891347	C>T	R1044K	Neutral	50	0.10	0.24
M28-4	3	*ACVR2B*	3	38495828	G>C	W5C	Neutral	16	0.63	0.91
M28-4	3	*TP53BP1*	15	43767876	G>T	C319*	X	46	0.21	0.70
M45-2	3	*NTN4*	12	96180874	A>T	V143E	X	140	0.43	0.97
M45-2	3	***SETD2***	3	47059198	G>T	P2488H	X	45	0.71	0.97
M45-2	3	***SETD2***	3	47161898	C>A	E1410*	X	105	0.25	0.45
M45-2	3	*SF3B1*	2	198268446	C>A	A528S	X	164	0.42	0.97
M45-2	3	*TPX2*	20	30386234	G>T	R671L	X	75	0.20	0.47
F60-1	3	*XIRP2*	2	168104667-8	GA>TC	M2080I, K2081Q	X	350	0.07	0.23
F60-1	3	*NOTCH1*	9	139400027	G>A	P1441S	Neutral	154	0.10	0.34
F60-1	3	*SMC1A*	X	53440031	G>A	L225F	X	223	0.27	0.43
F60-1	3	***BAP1***	3	52437750	delC	G470fs	X	211	0.70	0.99
F60-3	2	*ALPK2*	18	56204766	C>G	V885L	Neutral	118	0.31	0.99
F60-3	2	*CDKN2C*	1	51439577	G>C	G48R	X	84	0.12	0.45
F60-3	2	***PBRM1***	3	52643328	C>A	Splice Site	X	34	0.35	0.98
F60-4	3	*ANK3*	10	61958142	C>T	A549T	Neutral	87	0.28	0.91
F60-4	3	*APC*	5	112173923	T>A	L878M	Neutral	199	0.37	0.95
F60-4	3	***BAP1***	3	52436620	Tins24	*685fs	X	61	0.50	0.90
F60-4	3	*CARD11*	7	2962289	G>A	H750Y	Neutral	97	0.11	0.30

Gr. = Fuhrman Nuclear Grade, ccRCC driver genes are in bold, Chr. = Chromosome, Pos. = Position, Nuc. = Nucleotide, AA = Amino Acid, LOF = Loss-of-Function, VAF = Variant Allele Frequency, CCF = Cancer Cell Fraction

Table 4: List of candidate driver gene mutations across 10 tumors

2.5 Discussion

We have identified that ccRCC tumors associated with VHL Disease have few driver events, and in many tumors, the germline *VHL* mutation is the sole observable driver event. If high tumor grade represents a more evolved tumor, then high-grade tumors should have more clonal driver events than low grade tumors. Notably, low grade tumors appear to be driven by loss of *VHL* and 3p and do not have other driver events that appear to be early. Conversely, grade 3 tumors are driven by *VHL*, loss of 3p, and additional early mutational events.

A large number of the original 40 tumor genomes seem to have gained no additional mutations in cancer associated genes other than the germline *VHL* mutations and yet still developed into large tumors. The presence of known cancer-associated chromosomal alterations other than chromosome 3p loss is more common in these tumors than mutations, yet the largest tumor assessed (M28-1) demonstrated no additional known driver mutations or chromosomal alterations. Upon deeper sequencing, a few additional candidate driver mutations were identified within these 10 tumors, including one candidate driver mutation in M28-1.

We still do not know the minimum number or combination of mutations required for tumorigenesis but, our observations provide further evidence that this number can be small, on the order of one to ten driver mutations per tumor (Martincorena et al., 2017). For example, other than germline VHL, no known cancer driver variant or recurrent copy-number variant was observed in F60-6, a small low-grade tumor. Despite the peculiar lack of identifiable somatic variants supporting evolution towards tumorigenesis, the somatic mutation count and pathological features provide evidence for F60-6 being a tumor and not a cyst. Although, ccRCCs can often have cystic components and can be difficult to distinguish from grade 1 ccRCCs (Moch et al., 2016). The total sSNV count of F60-6 was very similar to that of tumors of a similar tumor

purity, grade, and age of diagnosis. We have observed cysts to have whole-genome sSNVs counts below 10 (data not shown), which is two-orders of magnitudes less than the whole-genome sSNV count observed in ccRCC. Perhaps the biopsy taken from this tumor did not include the tumor region containing driver variants. We speculate that complete inactivation of VHL in this tumor is due to methylation of the wild-type allele, which is observed in nearly 10% of sporadic ccRCCs (The Cancer Genome Atlas Research Network, 2013). There may be other epigenetic modifications that drive tumorigenesis for F60-6.

Structure of Tumor Evolution

Tumor evolution is also concerned with observing the degree of clonal selection. Competing models of evolution in sporadic and heritable ccRCCs exist. A previous study of four ccRCC tumors from an individual with VHL Syndrome (Fisher et al., 2014) identified few coding mutations and few subclonal mutations. The observed mutations occurred in a step-wise pattern which suggested a linear model of tumor evolution for heritable ccRCCs. However, other studies of sporadic ccRCC tumors (Gerlinger et al., 2012, 2014) observed a branched model of evolution. These studies identified several-fold more coding mutations, a higher degree of intra-tumor heterogeneity, and the presence of multiple subclones. Heritable and sporadic ccRCC are assumed to be similar diseases (Beroukhim et al., 2009) yet exhibit two different evolutionary trajectories.

With bulk sequencing, we lack the empirical data to fully determine the model of evolution. However, our study did identify novel results that provide some insight into evolutionary characteristics of heritable ccRCCs. Consider the two *SETD2* mutations identified in M45-2. One mutation, missense, was clonal while the truncation mutation was subclonal. M45-2 exhibits LOH at the *SETD2* locus, thus both mutations occur on the same *SETD2* allele. This suggests a linear relationship exists between these two *SETD2* mutations, where the missense

mutation occurred before the truncating mutation. We hypothesize that the missense mutation may have reduced the functionality of *SETD2* products and the truncating mutation caused a full loss-of function; Otherwise, there would be no evolutionary advantage of acquiring the second coding mutation within the same allele.

Multiple *SETD2* mutations have been previously identified in sporadic ccRCCs. However, in these studies, the *SETD2* mutations were determined to be subclonal and to exist on different branches (Gerlinger et al., 2012, 2014). Thus, in sporadic ccRCCs, multiple *SETD2* mutations represent a branched relationship and, with our observation, further suggests that the evolutionary characteristics of heritable and sporadic ccRCCs differ. High depth multiregional or single-cell sequencing performed on a subset of these tumors would be necessary to better determine the evolutionary architecture of *SETD2* mutations in heritable ccRCCs.

Model of Tumor Grade Evolution

We hypothesize the following model for grade determination in tumor evolution: early driver mutations are responsible for early determinants of tumor grade. Recent developments in mouse models of VHL Syndrome ccRCCs support our observations in human tumors. In mouse models, complete knockout of *VHL* with conditional knockout of *BAP1* led to high grade tumors, while conditional knockout of *PBRM1* led primarily to low grade tumors (Hou and Ji, 2018). *BAP1* mutations occur in 10-15% of sporadic ccRCC while 50% of sporadic ccRCC have *PBRM1* mutations (The Cancer Genome Atlas Research Network, 2013).

Rarely *BAP1* and *PBRM1* mutations are found in the same tumor. However, multiregional studies have only observed *BAP1* and *PBRM1* mutations in separate regions of the same tumor. Gene expression profiling of *BAP1*-only mutant and *PBRM1*-only mutant tumors depict different

profiles and little pathway overlap, suggesting that these tumors may represent different ccRCC subtypes (Joseph et al., 2016).

These mouse models put context around an intriguing result for patient F60. We observed two high grade tumors with *BAP1* mutations and one low grade tumor with a *PBRM1* mutation. All three variants were predicted to be both pathogenic and clonal within the bulk tumor sample, suggesting that these mutations were indeed early tumor drivers. Since these tumors arose from the same host environment, it argues against the germline or host environment contributing to differing tumor grades in human ccRCCs.

Chapter 3. Patient-Specific Recurrent Loss of Whole Chromosome 3 in von-Hippel Lindau Syndrome Clear-cell Renal Cell Carcinomas

Authors

Asia D. Mitchell, Christopher Boniface, Suzanne S. Fei, Cathy D. Vocke, W. Marston Linehan, Paul T. Spellman

All data presented in this chapter were analyzed by Asia Mitchell. The experimental design was led by Asia Mitchell with assistance from Paul Spellman. Previous data obtained and published by Suzanne Fei provided the foundation for this study. Christopher Boniface prepared the DNA libraries for sequencing. Cathy Vocke and Marston Linehan provided all patient materials, including DNA and clinical data. All figures and text were composed by Asia Mitchell.

3.1. Abstract

Somatic copy number alterations (SCNAs) affect a larger fraction of the genome than any other somatic genetic alteration in cancer and are thus of significant interest to carcinogenesis. SCNAs are often the site of known cancer driver genes and overlap with significantly mutated genes. In clear-cell renal cell carcinoma, four of the five most frequently mutated tumor suppressor genes are located along the p-arm of chromosome 3. The loss of 3p is observed in 90% of clear cell renal cell carcinomas (ccRCC). Of the ccRCCs that lose 3p, 9% also lose the q-arm of chromosome 3 and thus have whole loss of chromosome 3. It is currently unknown if loss of whole chromosome 3 is a significant genomic alteration in ccRCCs. We performed low-pass whole genome sequencing to assess arm-level copy-number across 143 synchronous and metachronous clear-cell renal cell carcinomas from 24 individuals with VHL Syndrome. Our profile of at least three primary tumors arising from the same environment provided sufficient power to identify patient specific constrains on arm-level somatic copy number alterations, including loss of whole chromosome 3. We observed recurrent loss of whole chromosome 3 in two unique patients. This supports our hypothesis that an undefined combination of genotype and environment can apply selective pressure for loss of whole chromosome 3.

3.2. Background

Somatic copy number alterations (SCNAs) affect a larger fraction of the genome than any other somatic genetic alteration (Beroukhim et al., 2010). SCNAs fall into four classes: focal events involving one or a few genes, arm-level events involving an entire chromosome arm, aneuploidy events involving whole chromosomes, and copy-neutral loss of heterozygosity. SCNAs are frequent across nearly all tumor types and are present in over 85% of the solid tumors profiled by TCGA (Zack et al., 2013). On average, tumor genomes contain 40 SCNAs of any length (Beroukhim et al., 2010; Zack et al., 2013), but discovery of SCNAs of all lengths is dependent on the experimental approach..

SCNAs are sometimes the site of known cancer driver genes and overlap with significantly mutated genes (Zack et al., 2013). Focal amplified regions are predicted to harbor oncogenes, while focal loss regions are predicted to harbor tumor suppressors. Focal alterations are often tissue specific, such as amplification of 17q23 in 18% of breast cancers corresponding to *ERBB2* (Zack et al., 2013) and loss of 10q23.31 in glioblastomas corresponding to *PTEN* (Brennan et al., 2013).

Less understood yet affecting more of tumor genome than any other alteration, are aneuploidy and arm-level SCNAs. The most frequently recurrent arm-level SCNAs occur two orders of magnitude more often than that of focal alterations. Some arm-level alterations exhibit near uniform patterns of alteration in certain tumor types, such as loss of 3p observed in over 91% of clear-cell renal cell carcinomas (ccRCCs) (The Cancer Genome Atlas Research Network, 2013). It is understood that 3p loss is common in ccRCC because four of the most frequently mutated ccRCC tumor suppressor genes (*VHL*, *PBRM1*, *BAP1*, and *SETD2*) are located within a narrow locus on 3p. Intratumor heterogeneity (Gerlinger et al., 2012, 2014) and mutation timing

studies (Gerstung et al., 2017; Mitchell et al., 2018) have identified loss of 3p as a clonal event, occurring early ccRCC progression.

Of the ccRCCs that lose 3p, 10% also lose the q-arm of chromosome 3 and thus have whole loss of chromosome 3 (The Cancer Genome Atlas Research Network, 2013). Other frequent arm-level SCNAs in ccRCC include gain of 5q and loss of 14q (Gunawan et al., 2001; Klatte et al., 2009; Mitsumori et al., 2002). Gain of 5q has been identified as concurrent with early loss of 3p (Mitchell et al., 2018). Loss of 14q is associated with the loss of *HIFA* (Shen et al., 2011), a key component of the hypoxia-inducible transcription factor complex and a target of pVHL.

A few groups, including our own, have studied copy-number variation in heritable ccRCCs from individuals with VHL Syndrome (Beroukhim et al., 2010; Fei et al., 2016; Fisher et al., 2014; Shuib et al., 2011). Individuals with VHL Syndrome have a germline loss-of-function mutation in VHL, resulting in essentially 100% lifetime risk of developing multiple ccRCCs (Prowse et al., 1997). The multiple ccRCCs can be synchronous (tumors originating in the same tissue at the same time), or metachronous (tumors originating in the same tissue across different time periods). We have previously published results from the whole genome sequencing (WGS) of 40 synchronous ccRCCs from six VHL Syndrome patients (Fei et al., 2016). Notably, we observed loss of whole chromosome 3 in ten of thirteen tumors within one patient, a statistically unlikely event if the process of chromosome loss is random.

Here, we profile over 150 synchronous and metachronous ccRCCs from 25 individuals with VHL Syndrome in order to investigate patient-specific constraints on somatic copy-number variation.

3.3. Methods

Samples and DNA extraction

Patient phenotypes and other clinical data were obtained from the National Cancer Institute Clinical Research Information System or patient charts. This study was approved by the Institutional Review Board of the National Cancer Institute (IRB study 16626). All patients provided written informed consent.

Tumor DNA was extracted from frozen normal or tumor tissue with Maxwell 16 Tissue DNA purification kits (Promega) using the "tissue" program. Blood DNA was extracted from EDTA-anticoagulated peripheral blood samples using Maxwell 16 Blood DNA purification kits (Promega) with the "buffy coat" program. DNA concentrations were determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). Twenty normal, nontumor, DNA samples were obtained from the Knight Cancer Institute BioArchive at Oregon Health and Science University. DNA was extracted from buffy coat or whole blood.

Whole Genome Library Preparation and Sequencing

Beginning with at least 20 ng of genomic DNA for all tumor and normal samples, transposase (tagmentation) reactions were performed using 2 uL of 5 ng/uL DNA, 7 uL of nuclease-free H_2O, 10 uL buffer, and 1 uL of Tn5 enzyme. Tagmentation enzymes provided by Adey lab. Transposase reactions were then incubated at 55°C for 5 minutes before adding 10 uL of 5M guanidine thiocyanate (Sigma G9277). Following tagmentation, samples were cleaned up using Ampure XP beads (Beckman-Colter A63881) to retain DNA fragments larger than 100 bps. Bead purification was performed at a ratio of 1.3 parts sample to 1 part Ampure XP beads, incubated at room temperature for 5 minutes, washed once with 70% ethanol, and finally eluted with 20 ul of nuclease-free H_2O.

Library amplification and indexing was performed using KAPA HiFi Hotstart 2X master mix (KAPA KK2602) by combining 25 uL master mix with 20 uL of eluted DNA (from previous reaction), 4.5 uL of nuclease-free H_2O, 0.5 uL of i7 indexed primer, and 0.5 uL of i5 indexed primer (primers from IDT). PCR was run under the following conditions: 72°C for 5 minutes, 98°C for 30 seconds, then 8 cycles of 98°C for 10 seconds, 62°C for 20 seconds and 72°C for 40 seconds, and finally 72°C for 10 seconds. PCR reactions were purified using the Ampure XP beads at a ratio of 1 part beads to 1 part sample. Samples were incubated at room temperature for 5 minutes, washed once with 70% ethanol, and eluted with 25 ul of nuclease-free H_2O.

Bead purified DNA libraries were assessed for quality using Agilent's High Sensitivity DNA kit (Agilent 5067-4626) with the BioAnalyzer 2100. DNA quantification was performed by qPCR using KAPA Library Quantification kit (KAPA KK4824). Single ended 50 bp sequencing was performed using Illumina HiSeq at the Oregon Health & Science University Sequencing Core. In order to maximize resolution of copy-number segments, targeted sequencing depth was calculated based on the equations below:

Equation 1

$$Reads\ per\ Sample = \frac{Genome\ Size}{Read\ Length} * Coverage$$

Equation 2

$$\#\ Windows = \frac{Reads\ per\ Sample}{Target\ Reads\ per\ Window}$$

Equation 3

$$CNV\ Resolution = \frac{Genome\ Size}{\#\ Windows}$$

Equation 4

$$\#\ Samples\ per\ Lane = \frac{Reads\ per\ Lane}{Reads\ per\ Sample}$$

For equation 1, genome size is 3.2 billion bp and read length is 50 bp (single ended reads). For equation 2, target reads per window was 300 and this value is based on a publication using LP-WGS for calling fetal CNAs from maternal blood samples [Dong et al., 2016}. This study was referenced due to the similarity in challenges between maternal-fetal DNA contamination and normal-tumor DNA contamination. I used equations 1 – 4 to derive the values in table 1 and chose a target coverage of 0.05X, or approximately 3.3 million reads per sample.

WGS Coverage	Reads per Sample	# Windows	CNV Resolution (bp)	# Samples per Lane
2	132,000,000	440,000	7,500	2
1	66,000,000	220,000	15,000	4
0.5	33,000,000	110,000	30,000	7
0.25	16,500,000	55,000	60,000	14
0.1	6,600,000	22,000	150,000	36
0.05	3,300,000	11,000	300,000	71
0.015	990,000	3,300	1,000,000	237
0.01	660,000	2,200	1,500,000	356
0.005	330,000	1,100	3,000,000	712
0.001	66,000	220	15,000,000	3,561

Table 1: Target coverage calculations for LP-WGS method of detecting SCNAs

Whole Genome Sequence Alignment

Single-end reads were aligned to the hg19 human reference genome using bwa-aln and bwa-samse (Li and Durbin, 2009). Output SAM files were converted to BAM files, sorted, and indexed using Samtools v0.1.17 (Li et al., 2009). MarkDuplicates, part of Picard Tools v1.51 (McKenna et al., 2010), was used to mark duplicate reads generated during the PCR amplification stage. Final BAMs were reindexed using the BuildIndex tool from GATK4. Alignment sequencing quality was assessed by viewing the distribution of MAPQ scores for mapped reads the total number of uniquely mapped reads across all samples. Samples with library sizes below 600,000 were removed from further analysis.

Somatic Copy Number Alteration (SCNA) Calls

Copy number ratios (log_2 scaled) were determined using CNVKit (Talevich et al., 2016) followed circular binary segmentation (CBS) to determine copy number segments. First, a flat reference was computed for the hg19 reference fasta file. This process determines GC-content and RepeatMasker information needed for bias correction. Next, in "wgs" mode, a pooled reference was built from the BAM files of all the non-tumor normal samples and the average target bin size was set to 1 Mbp. While still in "wgs" mode, copy-number ratios were computed for all tumor samples using the pooled reference. Finally, tumor copy-number ratio tables were segmented using the CBS algorithm. The final output file contains arm-level copy-number segments and the log_2 tumor to normal copy-number ratio.

Tumor Purity Calculation

To estimate tumor purity, I relied on the following assumptions: (1) Loss of 3p is an early event in all ccRCCs, therefore it's clonal and present in nearly all tumor cells. (2) The largest contiguous segment on 3p to represents the clonal 3p loss event. This segment will have the highest number of probes. (3) If a tumor does not have loss of 3p, we cannot use the method below to estimate purity. Copy number ratios are log2 scaled. Copy loss is a log2 ratio less than -0.10 (copy ratio of 0.933). A total of 143 of 158 tumor passed the above constraints. The equation (Eqn. 5) used to estimate purity is the percent error. Here, p is purity and CNR is the log_2 copy-number ratio of the largest contiguous segment on 3p in tumors with 3p loss.

Equation 5

$$p = 1 - 2 \times \left(2^{CNR} - {}^{1}/_{2}\right)$$

To test the accuracy of this tumor purity estimation method, we compared tumor purity estimates obtained from the ABSOLUTE method (Carter et al., 2012) for 10 samples previously profiled with 30X WGS (as described in Fei et al. 2016). I calculated the difference between the two methods then used these values to calculate the standard error (SE) of our method, as outlined in equation 6, where σ is the standard deviation of the difference between the two purity estimation methods and n is the number of samples compared.

Equation 6

$$SE = \frac{\sigma_{(p_{ABSOLUTE} - p_{3p})}}{\sqrt{n}}$$

Clonal Tumor Comparisons.

To compare complete SCNA profiles of tumors, copy number segments were divided into smaller segments. The smaller segments where chromosomal g-bands from hg19. This created 862 segments for each tumor. Chromosomal band segments and their $\log_2$ copy-number ratio were stored into a dataframe in Python. The columns of the data frame represented each tumor and the rows of the dataframe represented each chromosomal band. Bands lacking copy number for all tumors from an individual were removed from the matrix.

The complete matrix was used to calculate rho, the Spearman rank correlation coefficient. First, I computed then plotting rank values of a subset of tumor pairs to determine if our data fit the assumptions of the Spearman rank correlation. I identified a monotonic relationship between each pair assessed and thus determined that the Spearman's rho was an appropriate statistical test of correlation for our dataset.

To perform the test, first rho values were calculated for each pair of unrelated tumors, totaling unique comparisons. Next, we used the rho values for unrelated tumor pairs to create a

distribution of rho value for independent tumors. I used this distribution to determine clonal tumors to have a rho value ≥ 99.9th percentile. Finally, I computed rho values for each pair of related tumors and compared these values to the 99.9th percentile threshold of the rho value distribution for independent tumors. If the rho value was ≥ 99.9th percentile, then the tumor pairs were determined to be clonally related.

Patient-Specific Recurrent SCNA Analysis.

The identification of patient-specific recurrent SCNAs is based on the binomial distribution, Equation 7, where the probability of success is equal to the rate of the event in sporadic ccRCC cases from TCGA (The Cancer Genome Atlas Research Network, 2013). Here, p is the binomial probability, x is the number of tumors with an SCNA, F is the rate of that SCNA in sporadic ccRCC, and n is the total number of tumors from a patient or case. An SCNA would be significantly recurrently if the binomial probability is < 0.05. The binomial distribution relies on the assumption that all comparisons are independent, therefore, tumors determined to be clonally related were removed from the SCNA analysis.

Equation 7

$$p(x; n, F) = \sum_{n}^{x} {}_nC_x \times F^x \times (1 - F)^{n-x}$$

3.4. Results

Sample population

To better understand the SCNA landscape of synchronous tumors, I profiled SCNAs from additional cases of VHL Syndrome ccRCC. Statistical modeling indicated that at least three tumors per cases were necessary to identify patients with recurrent loss of whole chromosome 3 (Figure 1). This per patient tumor number was based on the binomial distribution where the probability of success is equal to 0.09, the rate of whole chromosome 3 loss in sporadic ccRCCs (The Cancer Genome Atlas Research Network, 2013). In patients with three or four tumors profiled, loss of whole chromosome 3 would be recurrent if observed in at least two tumors (Figure 1a-b). In patients with five tumors sequenced, the SCNA would be recurrent if observed in at least three tumors (Figure 1c).

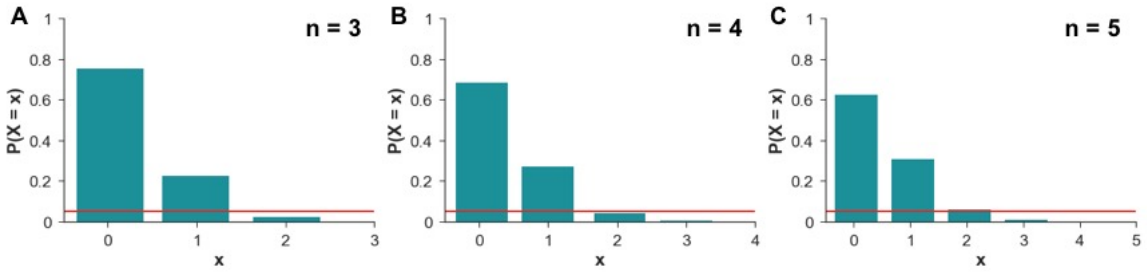


Figure 1: Sample size calculations. Significance calculations based on the binomial distribution, where n is the number of tumors per patient, x is the number of tumors with a specific SCNA, P(X = x) is the probability of the SCNA, and the red horizontal line is the significance threshold (p-value < 0.05). **A)** n = 3 tumors, **B)** n = 4 tumors, **C)** n = 5 tumors.

We identified 22 VHL syndrome patients with at least three synchronous or metachronous ccRCCs that were not part of our previous study (Table 2). Table 2 provides the naming schema for the surgery IDs - the patient's sex (M or F) followed by the patient's age at the time of tumor resection. Color codes and surgery IDs have been retained for tumors studied from patients also in Fei et al. 2016 (denoted in bold) and were added to the surgery ID for the one new

patient with metachronous tumors (Brown). Duplicate case IDs were followed by a letter to create unique case IDs (e.g. M27A and M27B). The median number of tumors per patient was four (Table 2). We also identified three additional cases from two previously characterized individuals (patient Green F28 and patient Red M22) (Fei et al., 2016). This amounted to 143 synchronous and metachronous ccRCC tumors from 24 individuals with VHL Syndrome.

Patient Color ID	Surgery ID	Number of Tumors	Resection Site	VHL Mutation	Relative
Brown	F57	1	R	CD 117kb	Green
Brown	F58	2	R	CD 117kb	Green
Brown	F59	1	R	CD 117kb	Green
Green	F20	6	L	CD 117kb	Brown
Green	**F28**	13	R	CD 117kb	Brown
Green	F29	5	L	CD 117kb	Brown
Orange	**M45**	3	R	fsR176	
Purple	**M28**	3	R	N78S	
Red	**M22**	5	R	P86R	
Red	M34	11	R	P86R	
Yellow	**F60**	3	R	W117C	
	F24	5	L	P86S	
	F49	6	R	L158P	M60
	F53	4	R	PD exon 3, 19kb	
	F59	6	L	W88R	
	F63	4	R	PD exon 2-3, 19.5 kb	
	F66	4	L	PD exon 2-3, 7.6kb	
	F73	4	L	R167W	
	M26	3	L	G104W, fs*28	
	M27A	4	R	delF76	
	M27B	16	L	PD exon 3	
	M32	4	L	delF76	
	M34	8	R	L118P	
	M37A	4	R	PD exon 2	
	M37B	4	R	splice acc exon 2	
	M39	7	L	PD exon 1, 5,646 bp	
	M42	6	R	PD exon 1, 2,700 bp	
	M44	6	R (3), L (3)	L135*	
	M47	4	L	PD exon 1, 1,089 bp	
	M48	4	L	R161*	
	M60	5	R	L158P	F49
	M84	4	R	fs codon 178	

PD = partial deletion, CD = complete deletion, splice acc = splice site acceptor, fs = frame shift, * = stop gain. Cases from Fei et al 2016 are in **BOLD**

Table 2: Patient and sample information

In addition to the new VHL Syndrome ccRCC cases, we resequenced nine tumors across three patients from our previous study (Fei et al., 2016) serve as a positive control and to determine the accuracy of our methods. Twenty diploid non-tumor DNA samples from whole blood were used as a pool of normals.

Our previous work identified patient-specific recurrent whole chromosome 3 loss in patient Green, F28, (Fei et al., 2016). In order to assess if this event was due to this individual's germline environment, we needed to study tumors from different time points. For patient Green, we acquired six tumors resected eight years prior from the left kidney (surgery ID is F20) and acquired five tumors resected one year later from the left kidney (surgery ID is F29). We also obtained four tumors from three surgeries (each surgery separated by one year) from the mother of this patient (surgery IDs are F57, F58, and F59), herein called patient Brown. We were interested in profiling tumors from related individuals to determine if recurrent SCNAs might be due to the inherited VHL mutation or other germline chromosome 3 variants in cis with the VHL mutation. For both the patient Green and Brown, the germline VHL mutation was a 117KB deletion that resulted in a complete deletion of VHL, a rare mutation in VHL Syndrome.

Lastly, we included tumors from a later resection for patient Red. The SCNA profiles for M22 were noisy in our previous study (Fei et al., 2016), but whole loss of chromosome 3 was apparent in two out of the five tumors profiled. We resequenced all five of these tumors and eleven additional tumors (surgery ID is M34) from a resection occurring twelve years later to determine the copy number status of chromosome 3 in patient Red's ccRCCs.

Low-pass whole genome sequencing

We performed low-pass WGS (LP-WGS) on our sample cohort to generate high quality arm-level copy-number profiles. We called somatic copy-number alterations (SCNAs) using CNVKit and a panel of normals, as described in the methods and outlined in Figure 2. CNVKit performs a genome-wide GC content correction and repeat-masked fraction correction before computing tumor-normal CNRs. Copy number segments were determined from the tumor-normal copy-number ratios using CBS. To reduce batch effects during library preparation and sequencing, we randomized sample order across tumor and normal samples, as well as by tumors from the same patient.

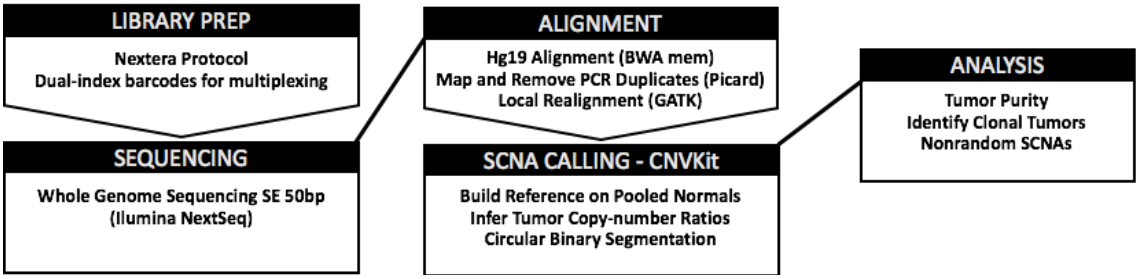


Figure 2: The Bioinformatic workflow for calling somatic copy-number alterations (SCNAs).

We removed two normal samples and one tumor sample from further analysis due to insufficient read depth. Figure 3 shows the distribution of total uniquely mapped reads following alignment to hg19 and removal of PCR duplicates for all remaining tumor samples. We achieved a range of 642 thousand to 14.34 million uniquely mapped reads and a median of 4.39 million uniquely mapped reads. With this depth, we achieved an approximate median genome-wide coverage of 0.05X.

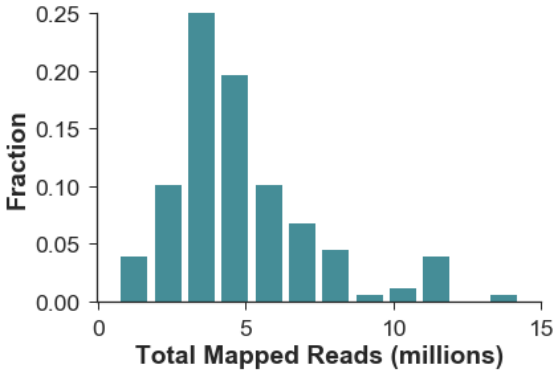


Figure 3: Histogram of the total uniquely mapped reads across all tumor samples. Uniquely mapped reads exclude PCR duplicates and unmapped reads.

To assess the accuracy of our arm-level SCNA calling approach, we compared arm-level SCNA calls from the LP-WGS method to those from 30X WGS across the nine resquenced tumor samples. These tumor samples had a tumor purity range of 25% to 75%. We were interested in determining the combined effect of tumor purity and total uniquely mapped reads on our ability to call loss of 3p in these tumors. For each tumor, we down-sampled the sequence library across a range of 300 thousand reads to the total number of reads then called SCNAs as outlined in figure 2. Figure 4 shows the $\log_2$ copy-ratios and 95% confidence across a variety of library sizes (total mapped reads). For high purity tumors, 65% to 75%, $\log_2$ copy-ratios from the 30X WGS fell within the 95% confidence interval of LP-WGS approach across nearly all library sizes. For the least pure tumors, 25% and 46%, the LP-WGS approach slightly underestimated the $\log_2$ copy-ratio. However, at library sizes of at least 1 million reads, the difference between the two approaches was able to distinguish a copy-loss event (middle plot in figure 4B) from no SCNA (top plot in figure 4B).

Tumor Purity

As previously mentioned, 3p copy-number can be difficult to distinguish in tumors of low purity and at low read depth. ccRCC can be highly cystic, resulting in low tumor cellularity. To account for contamination of non-tumor diploid cells and the likelihood of cystic structures within

the bulk tumor sample, we estimated the tumor purity for all tumors containing 3p alterations. Our tumor purity calculation relies on the assumption that loss of 3p is one of the earliest genomic alterations and is present in all tumor cells. Figure 5A provides a distribution of the tumor purities. The median purity was 56.94%, and the range of tumor purity was 17.27% to 90.77%. The purity range and distribution are on par with purity estimates for sporadic ccRCC (The Cancer Genome Atlas Research Network, 2013). We used median values from Figure 4 to remove any samples where tumor purity and library size may be too low to distinguish copy-loss of 3p from non-altered 3p. No samples fell below our thresholds for low tumor purity (<30%) and small library size (< one million reads); Thus, purity and read depth did not limit our ability to call arm-level SCNAs (Figure 5B).

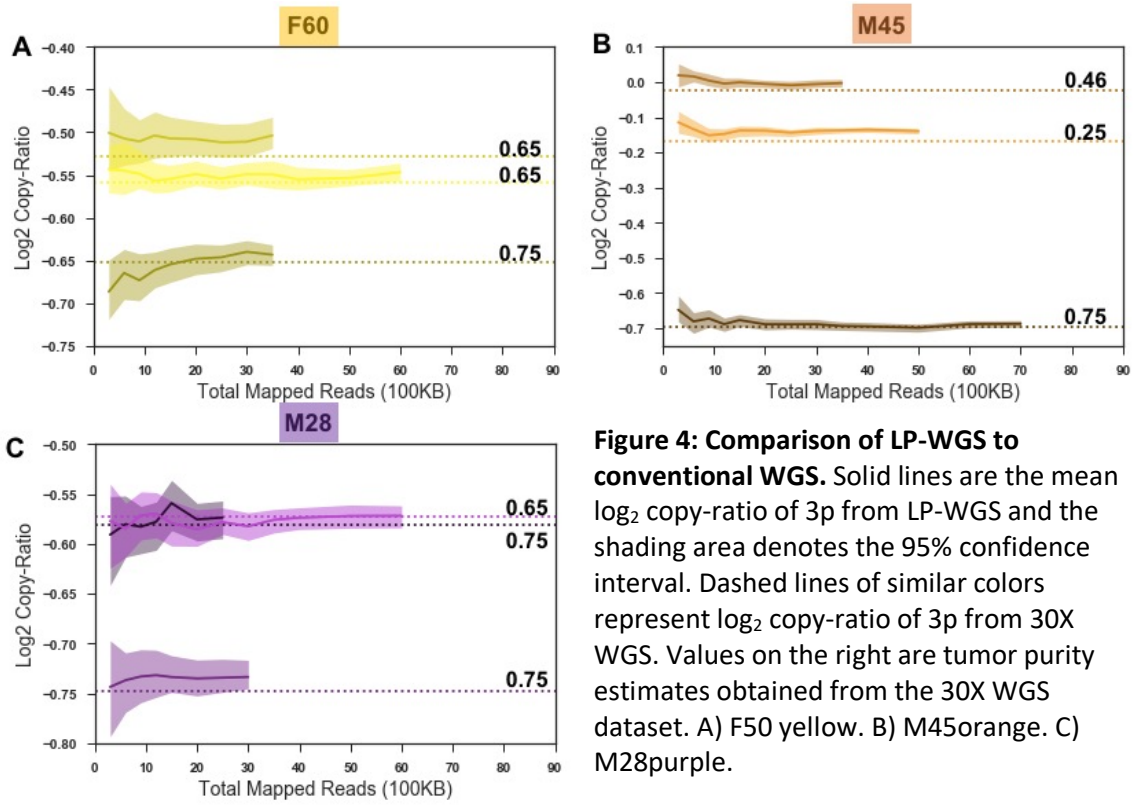


Figure 4: Comparison of LP-WGS to conventional WGS. Solid lines are the mean $\log_2$ copy-ratio of 3p from LP-WGS and the shading area denotes the 95% confidence interval. Dashed lines of similar colors represent $\log_2$ copy-ratio of 3p from 30X WGS. Values on the right are tumor purity estimates obtained from the 30X WGS dataset. A) F50 yellow. B) M45orange. C) M28purple.

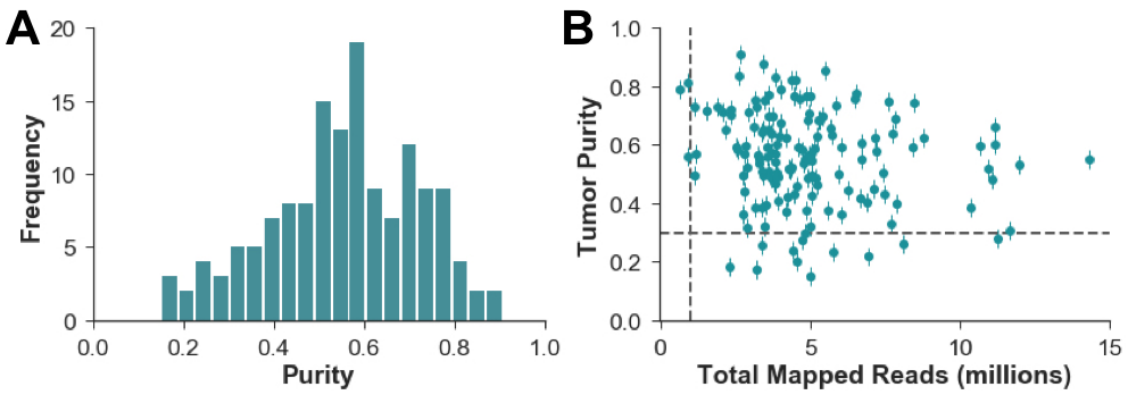


Figure 5: Assessment of tumor purities. A) Histogram of estimated tumor purities. **B)** Estimated tumor purity plotted vs total number of uniquely mapped reads. The horizontal dashed line at tumor purity of 0.3, represents the threshold for low tumor purity. The vertical dashed line at one million total mapped reads represents the threshold for small library size. Error bars are based on the calculations provided in Table 3.

To test the accuracy of this tumor purity estimation method, we compared tumor purity estimates obtained from the ABSOLUTE method (Carter et al., 2012) for ten tumors previously profiled with WGS (Fei et al., 2016). We calculated the difference between the two methods then used these values to calculate the standard error (SE) of our method, as outlined in the methods and shown in Table 3. The standard deviation for the difference is 1.6%. The purity estimate described here is +/- 3.2% (95% confidence) of the purity estimate from the ABSOLUTE method.

Tumor (Patient)	Purity (Absolute) 30X+ WGS	Purity (3p loss) Low-pass WGS	Difference
M28-1 (purple)	81.00%	83.80%	2.80%
M28-2 (purple)	68.00%	71.00%	3.00%
M28-3 (purple)	61.00%	68.20%	7.20%
F60-1 (yellow)	23.00%	20.30%	2.70%
F60-2 (yellow)	82.00%	76.00%	6.00%
F60-3 (yellow)	61.00%	64.20%	3.20%
M45-2 (orange)	62.00%	65.90%	3.90%
M45-3 (orange)	73.00%	75.50%	2.50%

Table 3: Comparison of two tumor purity estimation methods

Identifying Clonal Relationships

Previous genomic analyses of synchronous and metachronous VHL Syndrome ccRCCs (Beroukhim et al., 2009; Fei et al., 2016; Fisher et al., 2014) have not observed clonally related tumors from the same patient. However, it is possible that tumors from the same patient could be clonally related. We assumed that clonally related tumors would have SCNA profiles that were more similar than any two non-related tumors. To identify clonal tumors, we applied the Spearman Rank correlation test in a pairwise fashion to unrelated tumors in our cohort as outlined in Figure 6A and further detailed in the methods. The distribution of rho values across all unrelated tumor pairs are shown Figure 6B. We then applied the same Spearman Rank correlation test to each pair of related tumors. Significantly clonally related tumors were determined to have a rho value greater than the 99.9th percentile (rho < 0.802) of unrelated tumor pair rho values. This threshold is represented by the red line in figure 7B.

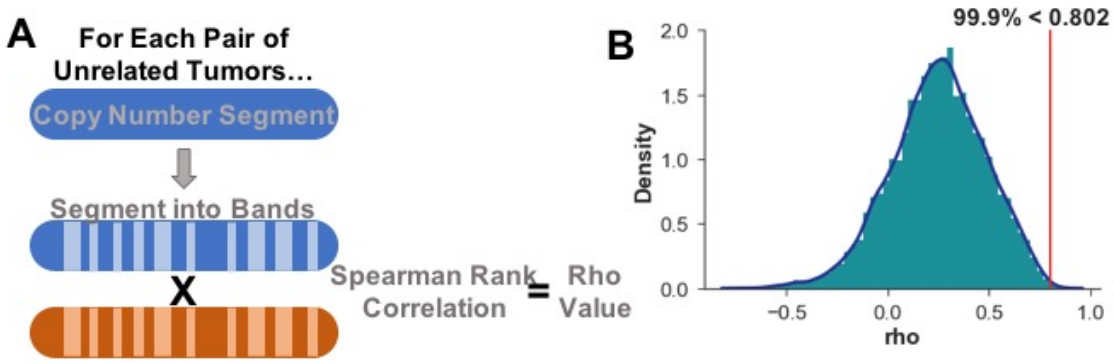


Figure 6: Spearman rank correlation determines clonal relationships. A) Diagram of the method for determining clonal relationships between related tumors. Copy number segments for each tumor are divided into chromosome band segments, then rho values of the Spearman Rank Correlation are computed for each pair of unrelated tumors. **B)** The density distribution of the rho coefficient from the Spearman Rank correlation test for all comparisons of unrelated tumor pairs. The blue outline represents the kernel density estimate. The red line represents the 99.9th percentile threshold for distinguishing a clonal relationship between tumor pairs.

We identified significant clonal relationships within two patients (Figure 7 and 8). In one patient, patient Brown, all tumors studied were determined to be clonally related (Figure 7A-C). For this patient, we used the rho values from the spearman rank correlation (Figure 7B) to reconstruct the phylogenetic tree in order to better visualize the clonal hierarchy (Figure 7C). Notably, the phylogenetic tree is consistent with the age of resection. The tumor resected at the youngest age (57 years, F57) was less correlated with all other tumors, while a tumor resected a year later (58 years, F58-1) had the highest correlation with all other tumors from patient brown. These observations suggested F57 it to be closest to the most recent common ancestor (MRCA), and F58-1 to be centrally located within the phylogenetic tree. Tumor F57 from patient Brown, was the only tumor to acquire loss of chromosome 4 and lacked loss of chromosome 13. Two tumors were resected at age 58 and differ by one SCNA. The loss of whole chromosome 10 occurs in F58-2, while F58-1, and all other tumors from patient brown, only loose 10q.

For patient M39, five of its seven tumors demonstrated clonal relationships (Figure 8A-C). We used the rho values from the spearman rank correlation to reconstruct a phylogenetic tree of clonal relationships (Figure 8B-C). It appears that the MRCA had several loss events including 3p 8p and 11l, as well as several gain events such as 3p and whole chromosome 20. Except for loss of 3p, tumors M39-6 and M39-7 do not share any of these early SCNAs. Tumors M39-1, M39-3, M39-4, and M39-5 share a gain of 5q, which is frequently observed with loss of 3p. M39-7 also has gain of 5q but lacks gain of 3q observed in the previously mentioned clonal tumors.

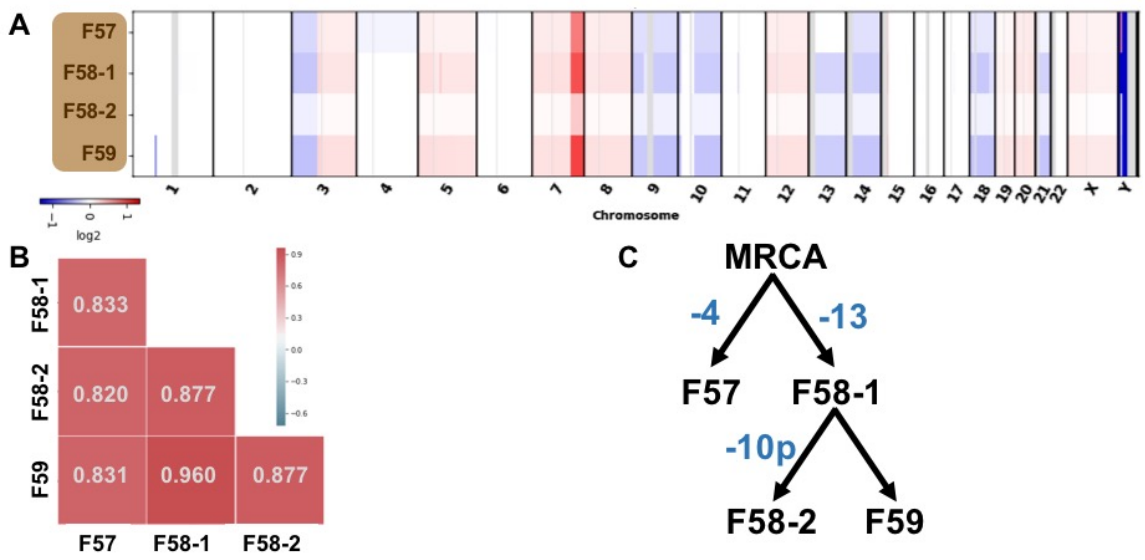


Figure 7. Clonal relationships in patient Brown. A) Genome wide SCNA profile for all five tumors across three resections from patient Brown. Different resections are denoted along the vertical axis by tumor ID, where 57, 58, and 59 represent the patient's age at the time of resection. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue. **B)** Correlation matrix for tumors in 5A based on Spearman rank correlation. The rho correlation coefficient is provided with the matrix. Strong positive correlations are dark red, while strong negative correlations are dark blue. **C)** Phylogenetic tree of evolutionary relationship for tumors in 7A-B built using deductive reasoning informed by correlation coefficients in 7B. SCNA events connecting branches are given in blue, where a minus (-) represents a somatic copy loss event.

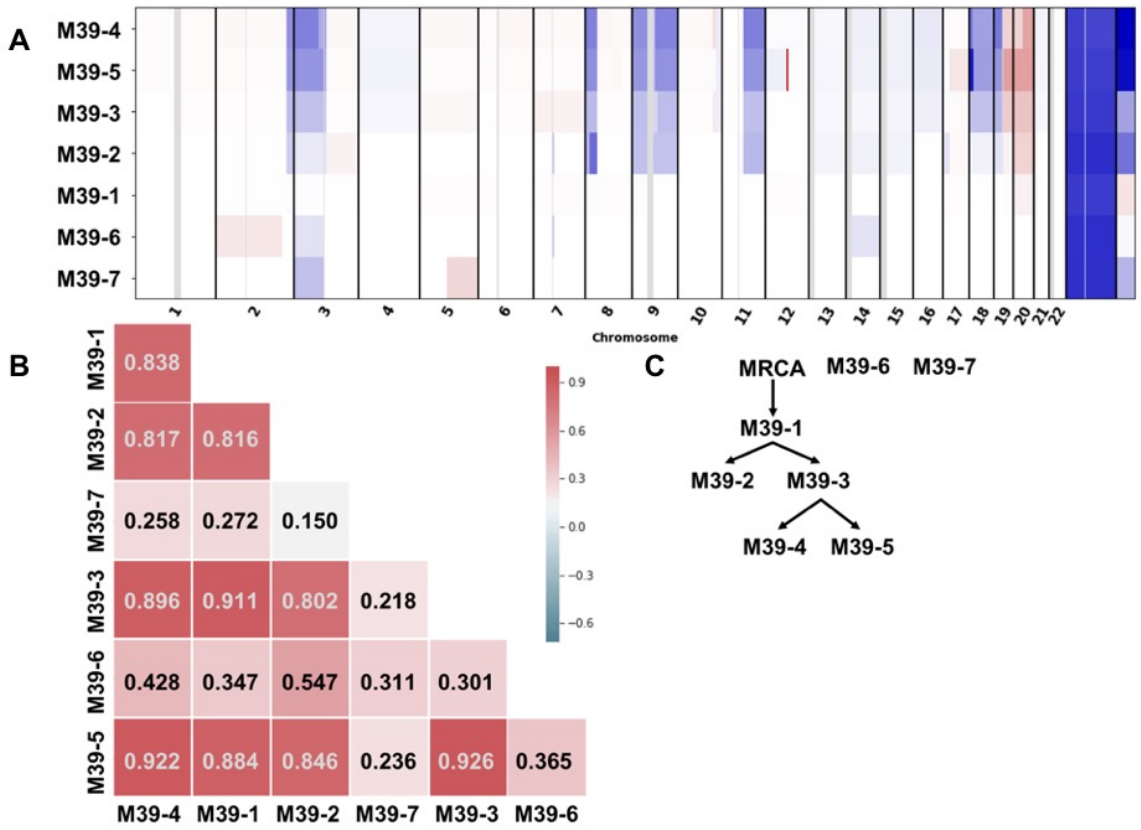


Figure 8. Clonal relationships in M39. A) Genome wide SCNA profile for all seven tumors across one resection from patient M39. Different tumors are denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue. **B)** Correlation matrix for tumors in 5A based on Spearman rank correlation. The rho correlation coefficient is provided with the matrix. Strong positive correlations are dark red, while strong negative correlations are dark blue. Clonal relationships have grey rho values. Independent relationships have black rho values. **C)** Phylogenetic tree of evolutionary relationship for tumors in 8A-B built using deductive reasoning informed by correlation coefficients in 8B.

SCNA Profiles

The genome-wide arm-level SCNA profiles for all tumors of remaining patients are provided in Figures 9 – 11 and in appendix B. Each figure contains the SCNA profile of ccRCC from a single patient. The SCNA profiles from 30X WGS from our previous publication (Fei et al., 2016) are provided for reference for the thirteen tumors for M22, patient Red, (Figure 9) and from F28, patient Green, (Figure 10). Loss of 3p and gain of 5q were the only two arm-level alterations observed in every patient. Collectively, 91% of tumors had loss of 3p and 57% of tumors had gain of 5q. These frequencies are consistent with sporadic ccRCC. Four tumors from four different patients (M45-3, M34-5, F24-6, and F73-4) had no alteration on chromosome 3 yet had other alterations. We presume these have copy-neutral LOH of chromosome 3 because we previously identified copy-neutral LOH of whole chromosome 3 in three ccRCCs discussed in chapter 2.

Patient-Specific Recurrent SCNAs

We were interested in identifying cases of recurrent events at arm-level and whole chromosome SCNAs observed in sporadic ccRCCs. Table 4 list frequencies of these SCNAs in sporadic ccRCC (from TCGA et al 2013) and the minimum number of cases per individual needed to identify a recurrent SCNA. For example, we are powered to detect patient-specific recurrent SCNAs occurring at a frequency $\leq$ 13% if at least 3 tumors per patient were available. Patients were removed from recurrent SCNA analysis if they lacked enough tumors to identify a statistically significant event. The binomial distribution assumes events are independent, thus clonally related tumors were not considered for recurrent SCNA analysis. Tests for recurrent events were only performed under two conditions: (1) patient met the minimum sample size for each SCNA listed in table 4; and (2) the SCNA was observed at least 3 times if the total number of tumors was

greater than 3. These conditions reduced the number of statistical tests performed and the need for multiple test correction.

Four patients (M27B, M34, F63, and M47) had a total of eight tumors lacking any identifiable SCNA. Although, we are unable to confirm that these samples contained tumor DNA, we still considered these tumors within the patient-specific recurrent SCNA analysis. Our previous study of 40 VHL Syndrome ccRCC identified one tumor containing no identifiable SCNA, which supports our rationale to consider these tumors.

We identified three new cases of patient-specific recurrent SCNAs, all involving loss of whole chromosome 3. These recurrent SCNAs are summarized in table 5. No other SCNA in table 4 met the two previously mentioned selection conditions necessary to perform the test for recurrence; and thus only three statistical tests for recurrent events were performed. M27B (figure 11) and both resections from patient Green (figure 9) were determined to exhibit patterns of patient-specific recurrent loss of whole chromosome 3. Whole loss of chromosome 3 was thought to be observed in the previous WGS of M22 (patient Red), but the copy-number calls were noisy. After resequencing with LP-WGS, whole loss of chromosome 3 was observed in two (M22-1 and M22-2) out of five tumors. However, this observation was not statistically significant and is likely random. Only one tumor out of eleven from the later surgery of patient Red lost whole chromosome 3.

SCNA	Frequency	Minimum Sample Size
+16	0.06	3
+17	0.06	3
-22	0.06	3
+1q	0.08	3
-3	0.09	3
+11	0.09	3
+18	0.09	3
+21	0.11	3
-21	0.11	3
+22	0.11	3
+19	0.12	3
-13	0.13	3
+2	0.16	4
10q	0.16	4
-1p	0.17	4
+3q	0.17	4
-18	0.18	4
-6	0.19	4
+16q	0.2	4
+20	0.23	4
-9	0.26	5
+12	0.26	5
-8p	0.28	5
+7	0.37	6
14q	0.41	6

Table 4: Frequencies of arm and chromosome- level somatic copy number alterations (SCNAs) in sporadic ccRCCs from TCGA

Surgery ID	SCNA	Tumors with SCNA	Total Tumors	p-value
M27B	-3	5	16	0.011
F20	-3	3	6	0.011
F29	-3	3	5	$6*10^{-3}$

Table 5: List of all patient-specific recurrent SCNAs

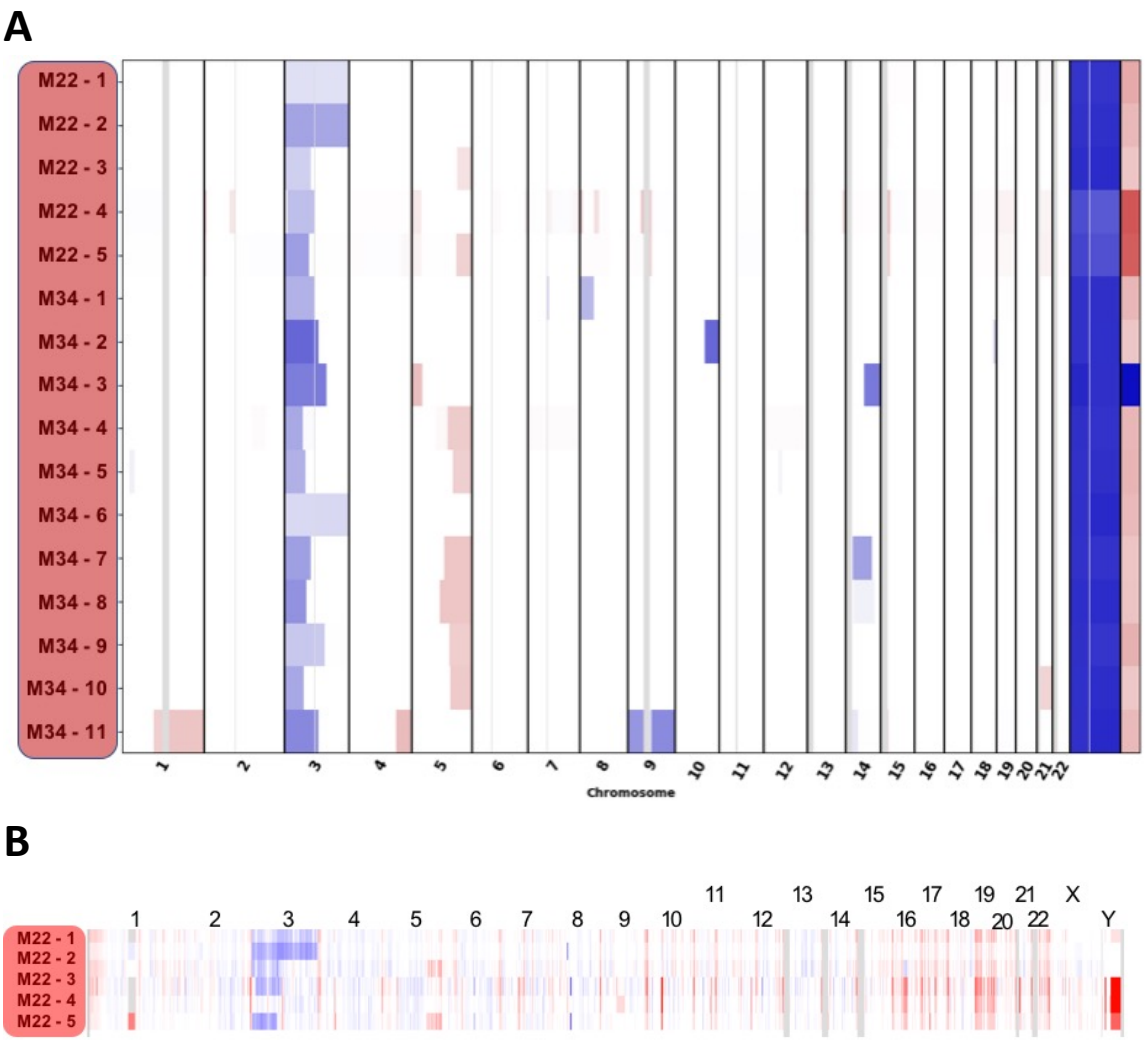


Figure 9. SCNA profiles and loss of chromosome 3 in patient Red. A) Genome wide SCNA profile from LP-WGS for 16 tumors across two resections. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue. **B)** Genome wide SCNA profile for M22 derived from 30X WGS from (Fei et al., 2016).

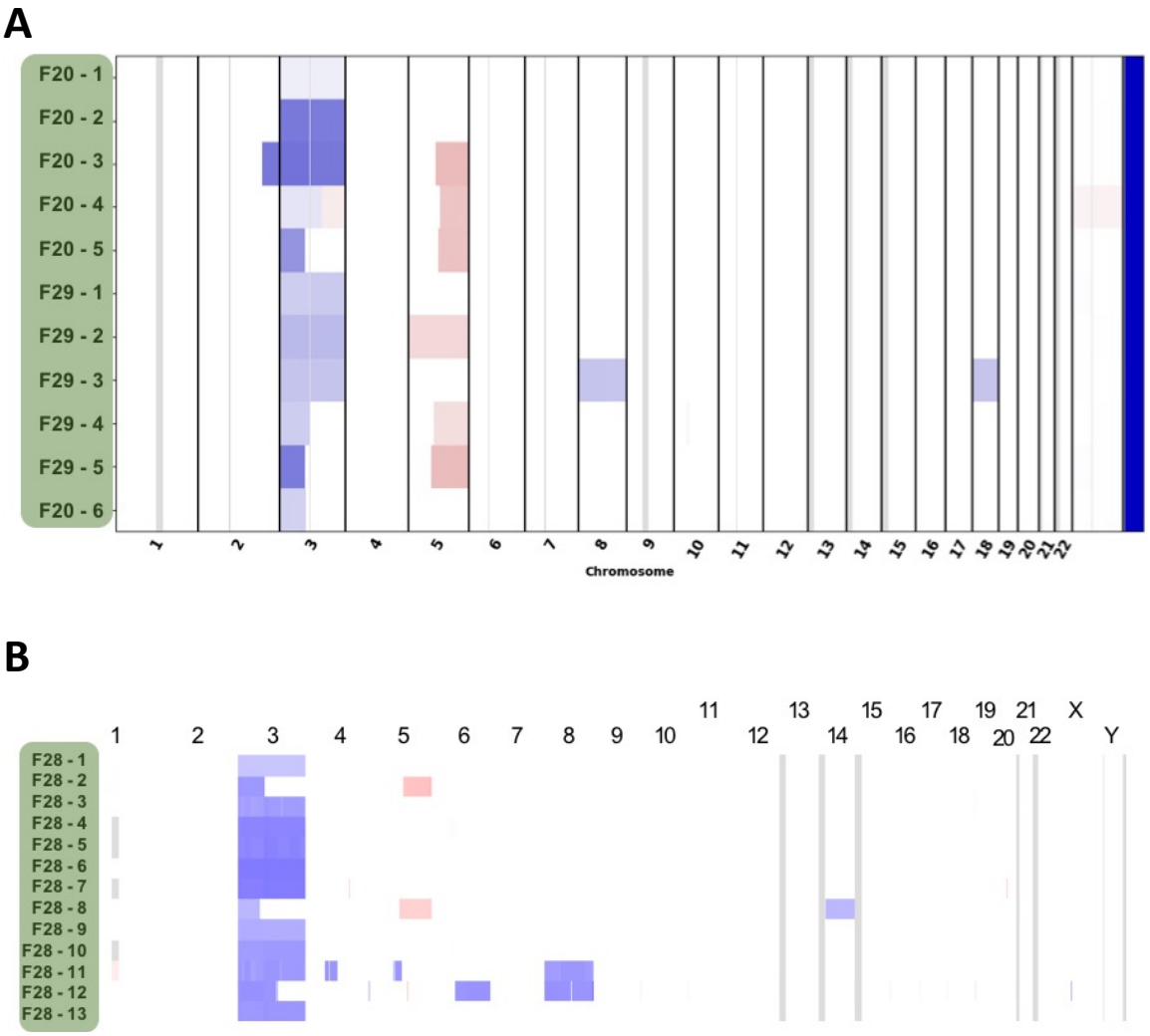


Figure 10. SCNA profiles and recurrent loss of chromosome 3 in patient Green. A) Genome wide SCNA profile from LP-WGSfor 11 tumors across two resections. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue. **B)** Genome wide SCNA profile for F28 derived from 30X WGS from (Fei et al., 2016).

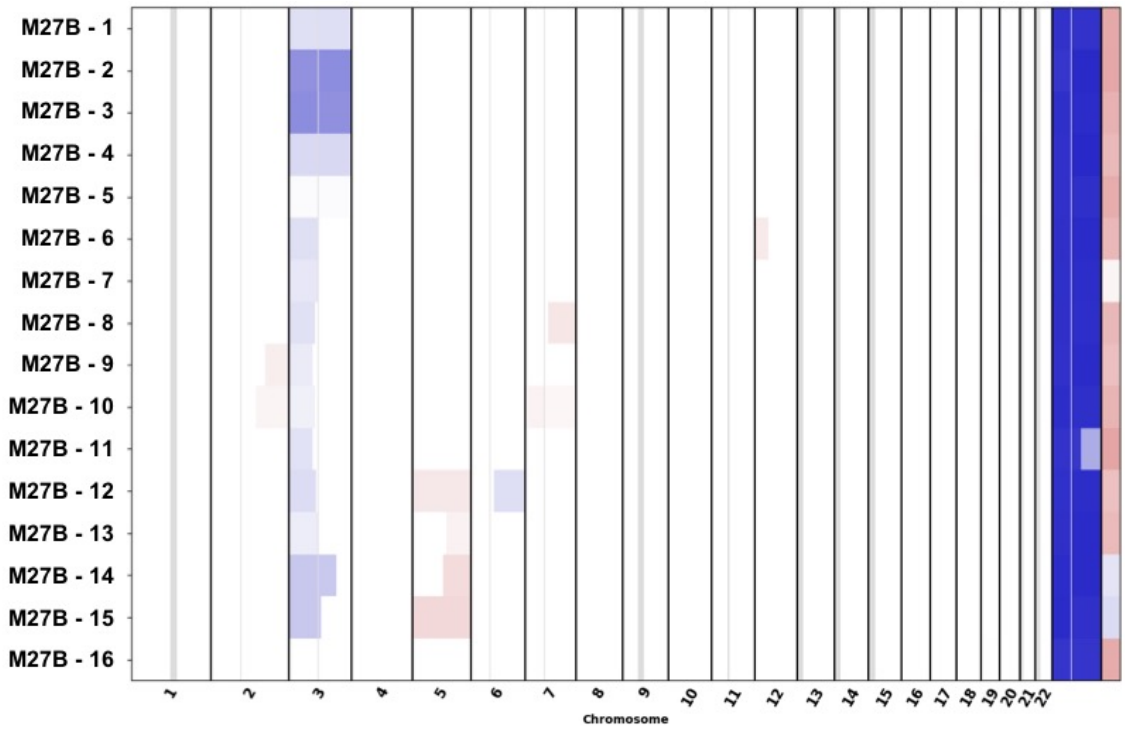


Figure 11. SCNA profile and recurrent loss of chromosome 3 in M27B. Genome wide SCNA profile from LP-WGS for 16 tumors across one resections. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue. M27B – 16 did not contain any arm-level SCNA.

3.5. Discussion

Loss of 3p represents an essential mechanism for driving tumorigenesis of ccRCC due to the loss-of-heterozygosity of tumor suppressor genes such as *VHL*, *PBRM1*, *SETD2*, and *BAP1*. Nearly every tumor we studied, with the exception of tumors without any observable SCNAs, contained loss of 3p. We understand loss of 3p to be an early event in the evolution of sporadic renal tumors from studies by Gerlinger et al., and we can also assume loss of 3p to be early in heritable ccRCC (Gerlinger et al., 2012, 2014). A mutation timing study by the TRACERx renal consortium, calculated loss of 3p to occur five to twenty years prior to *VHL* inactivation and three to five decades prior cancer diagnosis (Mitchell et al., 2018). This result suggests that for sporadic ccRCC, in the "two-hit" model of tumor initiation, loss of 3p is the first hit and inactivation of VHL is the rate limiting step. However, the tumor initiation model is reversed in heritable ccRCC – *VHL* inactivation is the first hit followed by loss of 3p as the rate limiting step. Therefore, the order of these two tumor initiating alterations must not be essential to tumorigenesis.

Mechanisms for loss of 3p in ccRCC

Copy-number alterations are often a product of punctuated evolutionary processes. Punctuated evolution results in a large increase in tumor heterogeneity in a short period of time due to the acquisition of multiple genetic alterations at once. Chromothripsis, or chromosome shattering, can cause punctuated evolution and chromoplexy, complex structural rearrangements, can be observed following chromothripsis. Chromoplexy has been observed as initiating events in several tumor types (Baca et al., 2013; Gao et al., 2016) including ccRCC (Mitchell et al., 2018). In ccRCC, recurrent breakpoint fusions were detected between 3p and 5q in tumors with 3p loss and 5q gain. In the rearrangement, a "derived chromosome" is formed between 3p and 5q. This alteration is common in sporadic ccRCC, occurring in up to 60% of tumors (The Cancer Genome Atlas Research Network, 2013). Other structural abnormalities between 3p

and other chromosomal arms in sporadic ccRCCs from TCGA as well as the observation of subclonal gain of 5q in many ccRCC both suggest there to be many mechanisms for loss of 3p in ccRCC.

TRACERx also applied their method to reanalyze the 40 heritable ccRCC whole genomes from our previous work, to determine if there was evidence for chromothripsis at 3p-5q breakpoint fusions. Similar to sporadic ccRCC, clustered rearrangements were detected at 3p and 5q breakpoints suggesting translocation of 3p to 5q , t(3:5), caused by chromothripsis (Mitchell et al., 2018). Breakpoint rearrangement alterations differed between tumors from the same patient, suggesting there to be no patient specific constraint in chromothripsis.

Here, we observed ccRCCs with concurrent 3p loss and 5q gain in nearly every patient, with this alteration occurring in well over half of all tumors. Several patients had concurrent 3p loss and 5q gain in all of their tumors. Our clonality analysis ruled out similarities in 3p breakpoint locations. These two observations lead us to believe that the host environment could constrain the mechanism for 3p loss. We did not have sufficient whole genome coverage to call genotypes or somatic mutations. However, with additional sequencing, we could determine evidence of chromothripsis driving t(3:5).

Mechanisms for increased whole chromosome 3 loss in ccRCC

While we understand why some somatic genetic alterations are beneficial to a tumor, we do not understand how one's constitutive genome can select for or against certain copy number events. We have identified a propensity for loss of whole chromosome 3 in two individuals. The observation of preferential whole chromosome 3 loss in all surgeries from patient Green is a strong indicator that additional surgeries from M27B should be studied. Interestingly both patient Green and M27B had a large number of tumors resected at each surgery. Their tumor counts were the largest across all patients profiled. Given the nature of how and when surgical resections of

ccRCCs are performed in individuals with VHL Syndrome, it may be a leap to make inferences from the total number of tumors resected. Nonetheless, it would be beneficial to this study to review clinical surveillance data to determine if patients who exhibit recurrent loss of chromosome 3 also exhibit a higher number of renal tumors or a higher total tumor burden at a single time point.

Numerous studies support that each type of SCNA must be governed by different mechanisms. Aneuploidy can result from genome wide doubling (Sansregret and Swanton, 2017). We previously discussed that arm-level alterations can be driven by chromothripsis. However, we do not know why some individuals have a propensity for loss of whole chromosome 3 or loss of 3p in renal tumors. It has been demonstrated in a few studies that aneuploidy is positively selected for in cancer and that aneuploidy promotes tumorigenesis via loss of heterozygosity of tumor suppressor genes (Sheltzer et al., 2017; Taylor et al., 2018). In one study, aneuploidy events involving single chromosome gains (single chromosome trisomy) suppressed tumorigenesis in vivo but later drove the cells to acquire structural abnormalities to compensate for the copy-gain and increase fitness (Sheltzer et al., 2017). The Darwinian principles governing tumor evolution lead us to believe that loss of whole chromosome 3 is a stronger tumor initiating alteration than loss of 3p in these individuals. Therefore, we have hypothesized several mechanisms for preferential whole chromosome loss in ccRCC.

Perhaps there exist inherited genetic events that apply selective pressure for or against whole chromosome 3 loss. For example, individuals with 3p loss could be enriched for deleterious heterozygous germline mutations along the q-arm of chromosome in *cis* to the germline VHL mutation. In these cases, loss of the whole chromosome 3 containing the wildtype VHL allele would unmask the deleterious effects of these other germline mutations on 3q and the deleterious effect would reduce fitness of the tumor cell or be cell-lethal. We could test this hypothesis through cytogenetics experiments on normal renal cells from patient Green and M27B

in comparison to other patients. If this hypothesis is true, whole chromosome 3 loss would be present in some normal renal cells from patient Green and M27B, but not present in normal renal cells of other patients.

Alternatively, there could exist low penetrance polymorphisms on chromosome 3 may modify chromosome 3 loss in patients with recurrent loss of whole chromosome 3. These polymorphisms may confer a small risk of increase tumor fitness alone but the risk is amplified when copy-number is altered. For example, Ewart-Toland *et al.* identified a polymorphism in *STK15*, a gene that controls chromosome segregation during mitosis. The polymorphism was low-penetrance, but increased tumor susceptibility due *STK15* copy number amplification (Ewart-Toland et al., 2003). This could be assessed by performing an association study limited to SNPs residing on chromosome 3. As with any association study, the effect size is likely small and large samples sizes would be necessary to achieve the needed power to perform this study.

We also considered if perhaps the germline VHL mutation modulated chromosome 3 loss. Patient Green had a 25 Kbp deletion encompassing the entire VHL gene. Complete deletion of VHL is less common in VHL Syndrome cases presenting with ccRCC because large deletion events can affect the function and expression of neighboring genes. This then affects the phenotypic presentation of VHL Syndrome. For example, one member of a family with a 50 Kbp deletion at VHL presented with bilateral ductal breast cancer (Krzystolik et al., 2014). It was later discovered that the large deletion affected the *FANCD2* gene, an important member of DNA repair processes. We hoped to compare SCNA profiles from patient Green and their mother, patient Brown. However, all tumors profiled from patient Brown were determined to be clonal. After identifying whole chromosome 3 in a new patient, M27B, with a different VHL mutation, it is unlikely that recurrent loss of chromosome 3 is due to the germline VHL mutation. Further, M27B's VHL

mutation, a partial deletion of exon 3, is common in VHL Syndrome patients presenting with ccRCC.

The identification of recurrent loss of whole chromosome 3 expands our understanding of patient specific pressures on tumor evolution. It is important to study the collection of somatic genetic alterations in the context of the underlying germline background so we can identify mechanisms that constrain tumor evolution.

Chapter 4 – Summary and Conclusion

Tumorigenesis as an evolutionary process has been substantiated by many studies since the 1950's. Early tumorigenesis studies relied on epidemiological data from non-endocrine tumors and generalized cancer to evolve in discrete stages. Despite observing the mutagenic potential of carcinogens such as x-rays as early as the late 1800's, cancer was not considered to be largely influenced by genetic factors until the 1960's. Advancements in molecular genetics provided the tools to identify and study individual cancer-causing genes. By the 1990's, we arrived at models of evolution that were tumor type specific and described the order of mutation events. The introduction of high-throughput genomic sequencing coupled with wide-spread collaborative cancer consortia have propelled tumor evolution research and contributed to thousands of peer-reviewed studies. We now understand cancer to be a dynamically adapting system, most often governed by stochastic and deterministic genetic processes.

Stochastic processes such as genetic drift introduce a degree of unpredictability. In cancer genetics, passenger mutations are a product of genetic drift. These are random mutations that do not change the fitness of the tumor. On the other hand, driver mutations do affect the fitness of a tumor and do not occur randomly. The acquisition of driver mutations is due to positive selection, a deterministic process. There is strong evidence that Darwinian positive selection is a major feature of tumor evolution, such as the enrichment of inactivating frameshift mutations in tumor suppressor genes (Yang et al., 2010) or the enrichment of certain mutation signatures in a particular tumor type (Alexandrov et al., 2018).

Tumor Evolution is Deterministic

The deterministic nature of tumor evolution suggests that the order and type of events are significant. While we do not know the minimum number of alterations or steps required to

initiate then drive tumor development, we know that there must be a few initiating alterations and multiple stages in tumor evolution. Clear-cell renal cell carcinoma (ccRCC) provides a good example of this. The loss of 3p and inactivating *VHL* mutation are requirements for tumor initiation (Mitchell et al., 2018). Later stages of tumor evolution usually include successions of clonal expansion and clonal selection, which may be driven by nongenetic factors. This multistage process results in increasing genotypic and phenotypic heterogeneity. It is the intratumor heterogeneity that creates massive challenges in the delivery and identification of targeted therapeutics. Additionally, if tumor evolution is completely controlled by Darwinian evolution, then cancer cells will invariably evolve to resist treatment until they outgrow their host environment, which is evident by the metastasis of late-stage tumors. Therefore, it is of interest to cancer medicine to understand the evolutionary dynamics and architecture of tumors, and to determine the contribution of Darwinian processes.

Tumor evolution is dependent on the context of the host environment, which consists of the germline background at a given time within a specific tissue type. However, the influences of the host environment on the evolution of a tumor are complex and make it difficult to derive reproducible conclusions from tumor evolution studies. But, if we observe the evolution of multiple tumors arising within the same host environment, what do we learn? In rare cases, individuals may present with multiple primary tumors. Tumors are synchronous if they are present in the same tissue at the same time, while metachronous tumors represent multiple primary tumors separated by at least six months. Both scenarios largely control for space and/or time, reducing complexity when studying the stochastic nature of tumor evolution. Can we utilize these constraints to predict prognostic features of tumors, or the types of therapies that might be effective? These are questions that drive further consideration of synchronous tumors in tumor evolution studies.

VHL Syndrome as a Biological Model for Probing Tumor Evolution

Synchronous renal tumors of the clear cell subtype arise in individuals with VHL Syndrome. This biological system is underutilized in cancer genetics studies. Perhaps rare diseases are overlooked in studies traditionally relying on big data either from large sample sizes or high-resolution data collection to reach sufficient power. However, the data shown within this dissertation recommends researchers in cancer genetics to consider if better systems exist for studying tumor evolution. The term "biological system" is used in reference to VHL Syndrome ccRCC to reflect that it a system of natural, spontaneous existence.

All individuals with VHL Syndrome have a germline mutation in the tumor suppressor gene, *VHL*. There are a variety of phenotypes including neoplasms from a variety tissues of origin. Patients presenting with renal tumors, the only malignant phenotype in VHL Syndrome, have loss-of-function (LOF) *VHL* mutations usually involving a deletion within the first three exons. Additionally, two-thirds of sporadic ccRCCs have mutated or epigenetically silenced *VHL*, causing *VHL* to be the most frequently mutated gene observed in this tumor type. Under normal conditions, pVHL functions as a negative regulator of angiogenesis through its interaction with hypoxia inducible factors (HIFs). Angiogenesis is a hallmark of cancer. Inactivation of *VHL* disrupts the regulation of HIFs under hypoxic conditions and increases transcription of downstream targets that activate angiogenesis.

Like other tumor suppressors, an inactivating *VHL* mutation is unmasked by loss of the wild-type allele. Loss-of-heterozygosity (LOH) at *VHL* nearly always involves loss of 3p and is observed in over 90% of sporadic ccRCCs. The near requirement of VHL inactivation in both heritable and sporadic ccRCCs provide further interest of studying the evolution of ccRCCs and suggest similar evolution models between the two. The early age of onset for ccRCC in VHL Syndrome patients is consistent with a single mutation model (Maher et al., 1992). The model

suggests the rate-limiting step for tumorigenesis in heritable ccRCC to be LOH at 3p. A simple model of ccRCC tumorigenesis would include an inactivating *VHL* mutation, then loss of the remaining wild-type *VHL* allele via LOH at 3p. However, mouse models have demonstrated that biallelic inactivation of *VHL* is not sufficient to drive ccRCCs but does induce multicystic renal disease (Rankin et al., 2006). Due to lack of synteny between humans and mice, it is not possible to replicate 3p LOH in a VHL Syndrome mouse model. Recent VHL Syndrome mouse models have successfully driven ccRCC under conditional knock-outs of other known tumor suppressor genes such as *TP53*, *RB1*, *BAP1*, and *PBRM1* (Gu et al., 2017; Harlander et al., 2017; Hou and Ji, 2018).

We have learned from several ccRCC genetics studies that four of the five most frequently mutated genes are located on 3p (The Cancer Genome Atlas Research Network, 2013). This includes *VHL*. The three remaining genes are all chromatin remodelers: *BAP1* a methyltransferase (mutated in 8-11% ccRCCs), *SETD2* a deubiquitinase (mutated in 3-12% of ccRCCs), and *PBRM1* a member of the SWI/SNF complex (mutated in 33-41% of ccRCCs). Both *SETD2* and *BAP1* directly modify histones; while *PBRM1* is responsible for altering nucleosome to DNA interactions. *SETD2* mutant ccRCCs are associated with increase hypermutation (Ricketts et al., 2018) and hypermutated tumors correlated with decreased survival. *BAP1* mutated ccRCCs also correlate with decreased survival and worst prognosis. Interestingly, *BAP1* and *PBRM1* mutations are mutually exclusive and ccRCC with these mutations display different gene expression profiles (Joseph et al., 2016). The consistencies in genomic location, functions in chromatin remodeling, mutation frequencies, and clinical outcomes as well as the variance in concurrent mutations and gene expression must not be coincidental. Each overlapping and distinctive observation provides insight into the constraints of ccRCC evolution.

Observation of Constrained Evolution

We have provided supporting evidence that the genomic analysis of synchronous and metachronous ccRCCs across multiple patients with VHL Syndrome provide an effective system for understanding tumor evolution. The VHL biological system allows for the study of multiple independent primary tumors under the same host environment. We consider the host environment to include the genotypic and phenotypic features of the environment. There are many models and trajectories of tumor evolution, and any single tumor can exhibit multiple evolutionary trajectories during its lifespan. I argue that tumor evolution models are not mutally exclusive but are subjective to the features you consider within the system that you are exploring. In controlling the host environment and in some cases varying time, I identified divergent, parallel, and constrained evolutionary trajectories as well as positive selection within our two studies.

In chapter two, somatic driver events were analyzed across a selection of high and low-grade tumors. Patient F60 demonstrated both parallel (Figure 1A) and divergent (Figure 1B) evolutionary trajectories in relation to Fuhrman grade. Parallel trajectory towards high-grade ccRCC was observed between F60-1 and F60-4. These two high-grade tumors possessed clonal inactivating *BAP1* mutations, each resulting in a frameshift and both under LOH. *BAP1* mutant ccRCC have previously been associated with high Fuhrman grade and increased metastatic rate (Minardi et al., 2016). Both tumors acquired different pathogenic mutations in the same gene, *BAP1*, early in tumor development and both evolved into high-grade tumors. In comparison, F60-3 acquired an early pathogenic mutation in *PBRM1*, with LOH, and instead evolved into a low-grade tumor. Therefore, there is also evidence of divergent evolution in F60 at clonal inactivating *BAP1* and *PBRM1* mutations in regards to tumor grade.

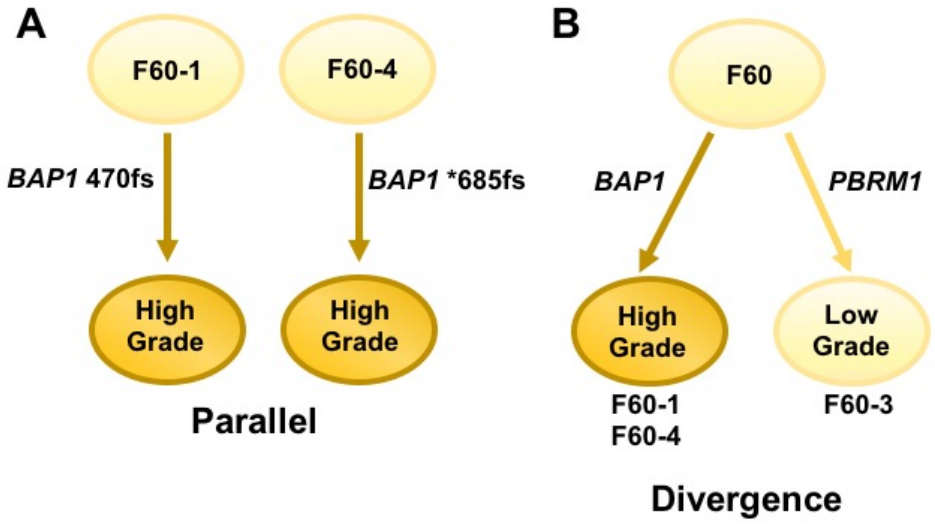


Figure 1: Model of clonal *BAP1* and *PBRM1* inactivating mutations as determinants of tumor grade in F60. A) Different clonal inactivating frameshift mutations in *BAP1* in tumors F60-1 and F60-4 lead to parallel evolution towards high Fuhrman grade (dark shading). **B)** Clonal inactivating mutations in tumor suppressor genes, *BAP1* and *PBRM1*, lead to divergence upon tumor grade. High Fuhrman grade is represented by dark shading, while low Fuhrman grade is represented by light shading.

An example of positive selection was observed in M45-2 (Figure 2), a tumor with two different *SETD2* mutations, one clonal and one subclonal, on the same allele. While both *SETD2* mutations were predicted to be inactivating, it is seemingly impossible to acquire two mutations in the same gene by chance. One theory is that the clonal *SETD2* mutation, P2488H, resulted in a partial inactivation or a reduction in function of the *SETD2* gene products, and the subsequent *SETD2* mutation, E1410*, resulted in a completed inactivation. The subclonal *SETD2* mutation must present a fitness advantage in order for it to persist. This theory is supported by Darwinian positive selection as a prominent evolutionary feature of tumors. If this theory were true and tumor M45-2 continued to evolve, then we would observe clonal out-growth of cells containing the E1410* in *SETD2* over time.

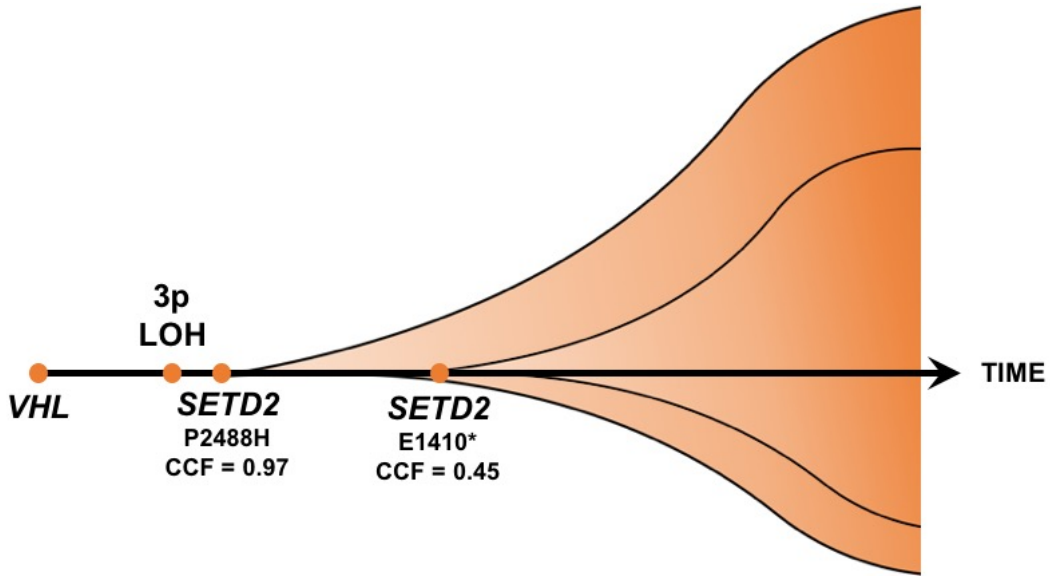


Figure 2: Linear evolution at *SETD2* in M45-2. Solid orange circles represent mutational events during the lifetime of tumor M45-2. The germline mutation in *VHL* is the first driver event followed by the loss of heterozygosity (LOH) at 3p. The next driver mutation is P2488H in *SETD2*, represented by a cancer cell fraction (CCF) of 97%. Here, clonal expansion is initiated. Both *SETD2* and *VHL* are located on 3p and thus are in LOH. A second mutation, E1410*, occurs on the same allele of *SETD2* and is present in approximately 45% of the cancer cells.

The observation of two mutations in *SETD2* in tumor M45-2 also support that the evolutionary architecture around *SETD2* mutations can be linear in VHL Syndrome (heritable) ccRCCs. Figure 2 models the linear evolutionary architecture with respect to SETD2 in M45-2. This observation is in contrast to the branched evolutionary architecture around SETD2 mutations in several sporadic ccRCCs (Gerlinger et al., 2014). Linear evolution has been previously observed in two VHL Syndrome ccRCCs (Fisher et al., 2014). The observations of linear evolution in VHL Syndrome ccRCC does not rule out the presence of branched architecture in all heritable ccRCCs. Similarly, this does not suggest that heritable and sporadic ccRCCs are largely different diseases and are not comparable. We have not studied enough VHL Syndrome ccRCCs with the appropriate genomic methods to be able to both fully characterize and generalize the evolutionary architecture of these tumors. Experimental approaches that include genomic analysis of multiple spatial-temporal biopsies (e.g. multi-regional or single-cell

approaches) from several independent tumors across several patients would be necessary to potentially observe branching evolutionary architecture.

It may also be that in order to observe branched evolution architecture, there must be an opportunity for competing clones to emerge. Perhaps a higher number of cell divisions, and time, are required for clonal competition. In general, sporadic ccRCCs evolve across double the time span than early heritable ccRCCs. With increased cell divisions, there is more chance that a selective sweep is incomplete providing an opportunity for divergence. We have also seen in models of colorectal cancer that linear and branched evolution are not mutually exclusive (Suzuki et al., 2017). If the same can be said for renal cancers, then I propose that linear evolution is a predominant feature in less aged tumors (such as most heritable ccRCCs) and branched evolution is a feature of tumors of greater age (such as most sporadic ccRCCs).

Constraints on tumor evolution were profiled in chapter three. The identification of two unique individuals, M27B and patient Green, with recurrent somatic loss of whole chromosome 3 demonstrate that the host environment can constrain the evolution of the tumor's somatic genome (Figure 3). I am unsure of the fitness advantage of whole chromosome 3 loss in ccRCC or in these two patients, but there must be one. However, this observation encourages additional genomic studies of larger sample sizes of ccRCCs and detailed molecular biology experiments on tumor and normal samples from these two individuals. I am also unsure if there exists a patient-specific bias towards whole chromosome 3 loss, or other somatic copy-number alterations, in sporadic ccRCC. I also identified that synchronous and metachronous ccRCCs can be clonally related and complete independence should not always be assumed when studying synchronous or metachronous of any tumor type. Experimental protocols should always contain an approach for determining evolutionary independence.

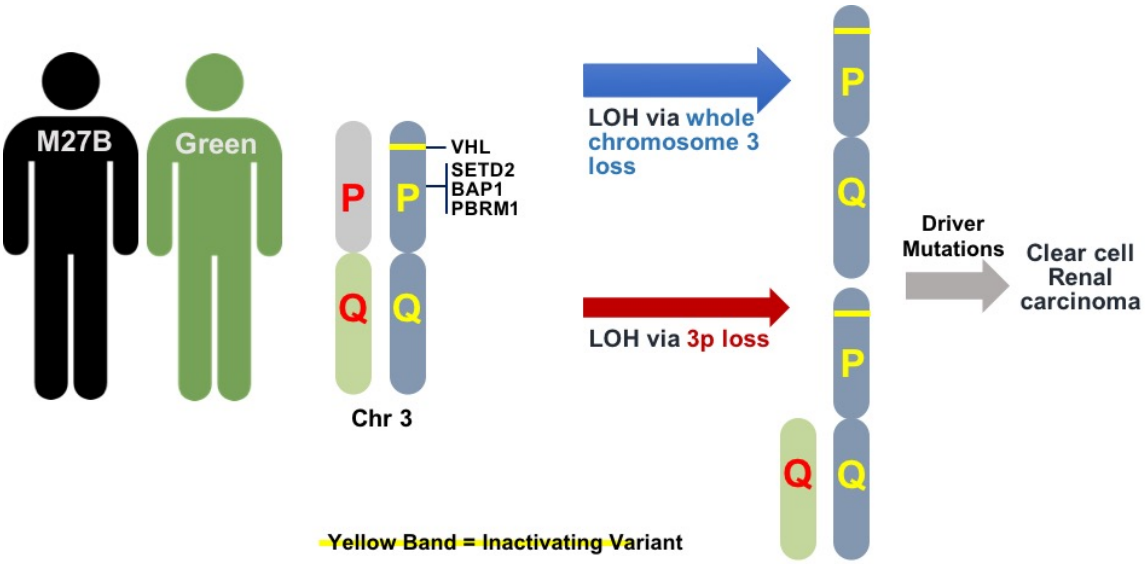


Figure 3: Overview for patient-specific recurrent loss of whole chromosome 3. Two patients, M27B and patient Green, exhibited significant recurrent loss of whole chromosome 3. Almost ubiquitously, LOH of *VHL* is achieved through loss of 3p and rarely through loss of whole chromosome 3. In M27B and patient Green, there is an increased rate for whole chromosome 3 compared to what we observe in sporadic ccRCC, represented by the wide width of the blue arrow. Alternatively, there is a decreased rate for loss of 3p, as represented by the narrow width of the red arrow.

Recommendations for future tumor evolution studies

As previously mentioned the assessment of intratumor heterogeneity (ITH) can provide prognostic value. Yet, there has only been one study of heritable ccRCCs aimed at quantifying ITH in VHL Syndrome ccRCCs (Fisher et al., 2014). Only two tumors from one patient were studied and no intratumor heterogeneity was observed in these tumors. Both tumors were low-grade and likely less evolved. Perhaps these tumors had not undergone clonal divergence or the experimental approach was not powered to detect it. Nonetheless, this result is surprising and worth questioning because sporadic ccRCCs exhibit significant ITH at both the genotypic and phenotypic level (Gerlinger et al., 2014; Turajlic et al., 2018). Studying ITH in relation to prognostic phenotypes such as tumor grade would be insightful. In order to assess genotypic

ITH, multiple tumor regions or single tumor cells must be studied. Both approaches require considerable foresight in sample collection and processing methods.

The design of this study lacked the sample resources to perform functional assays such as gene expression profiling, protein signaling dynamics, and cellular assays. These assays have been performed on sporadic ccRCCs in large-scale. Gene expression and epigenetic modification analysis of tumors analyzed in chapter 2 could have explained Fuhrman grade in tumors lacking inactivating driver mutations. Studying protein signaling dynamics across synchronous tumors might reveal additional modes of convergent, divergent, and parallel evolution. Both functional assays could be used to determine a mechanism for preferential loss of whole chromosome 3. In order to test this, normal tissue or cells would be required from patients with nonrandom whole chromosome 3 loss and those without.

Future recommendations are largely dependent on incorporating different sample collection methods. Collection of sufficient synchronous ccRCC is already challenging due to the low national and international incidence of VHL Syndrome. To mitigate this, increased collaboration among research biobanks is needed to acquire and process a variety of synchronous ccRCC samples for tumor evolution studies.

Models are inherently data driven. As we acquire more genomic data at higher resolution across many temporal and spatial contexts, we improve our power to detect reproducible patterns. A multistage model of cancer is still well supported by our data but each stage may be driven by different mutational processes and thus tumor evolution would be better described by integrating models. The order and context of the evolution processes must be pivotal to tumorigenesis. We need better systems to study tumor evolution. It is of the advantage of the greater cancer genetics research community to consider how to maximize the use of models already in existence.

Appendix A

Suzanne Fei, **Asia D Mitchell**, Michael Heskett, Cathy Vocke, Christopher J Ricketts, Myron Peto, Nicholas Wang, Kemal Sonmez, W. Marston Linehan, and Paul Spellman. "Patient-specific factors influence somatic variation patterns in von Hippel-Lindau disease renal tumors." Nature Communications 2016; 7: n. pag. Web.

Contributions: In collaboration with Suzi Fei, a former post-doc in the Spellman Lab, I assisted in the analysis of variants from the whole-genome sequencing of 40 clear-cell renal cell carcinomas (ccRCCs) from six individuals with VHL Syndrome. I also assisted in manuscript writing and formatting for publication. This is the original dataset providing the foundation for my dissertation research.

Approximately 10% of somatic single nucleotide variants (sSNVs) were shared between at least two tumors from the same patient. All sSNVs were called using whole-blood as the normal genome. However, we also had normal kidney tissue for three patients and called sSNVs using the kidney normal as well. Upon comparing shared sSNVs called from the blood normal to those called using the kidney normal, half of the shared sSNVs were determined to be kidney-specific sSNVs. For the remaining 5% of shared sSNVs, we wanted to determine if they fit any of three scenarios: (1) germline contamination or kidney-specific mosaic sSNV, (2) clonal sSNV, or (3) sequencing artifact. Using variant allele and total read counts for each sSNVs, I developed a Python script to parse this data table and bin each variant into the most likely of one of the three scenarios.

Germline and kidney-specific mosaic sSNVs would be present in multiple tumors from a single patient but not found in any other patients. These sSNVs would vary in variant allele frequency (VAF) and some tumors may have fewer variant reads or a lower VAF than our

minimum cutoff for calling a sSNV. To account for this, I included all sSNVs with a total read depth of at least 15 and at least one variant read observed with a MAPQ score $\geq$ 30.

Sequencing artifacts are a bit easier to identify in larger sequencing studies. To categorize shared sSNVs as erroneous, we assumed they could be present in multiple tumors across different patients. Additionally, we considered that if a particular locus was multi-allelic at low VAFs across multiple tumors, it was likely erroneous. Given the large volume of sSNVs to assess, I followed up with a manual review of alignment quality in IGV for a few candidate erroneous sSNVs.

To identify variants associated with clonal relationships, we compared groups of sSNVS that were shared across tumors from the same patient. If two tumors shared a clonal relationship, we would expect there to be at least 100 shared sSNVs across the whole-genome. Differences in the VAFs of each sSNVs between the tumor would demonstration clonal ordering. A tumor that disseminated from another tumor would have a collection of shared sSNVs at a lower VAF than the tumor that preceded it. If two tumors share a most recent common ancestor (MRCA), then both tumors could share similar VAFs across the shared sSNVs.

We were able to categorize > 91% of all shared sSNVs using the methods described above. 82% of shared sSNVs (2,018 variants) were categorized as sequencing artifacts or errors. The remaining 9% of shared sSNVs (226 variants) were categorized as kidney-specific mosaic sSNVs. There were not large enough collections of shared sSNVs to suggest clonal relationships exist between any of these tumors. Interestingly, chromosome 22 was over-represented (relative to size and average mutation rate in ccRCC) and chromosome 18 was under-represented for candidate kidney-specific variants. Chromosome 11 was also over-represented for candidate kidney-specific variants. This observation is interesting because in sporadic ccRCCs chromosome 11 is the least frequently altered in respect to copy-number alterations.

Appendix B – Somatic Copy Number Alteration Profiles

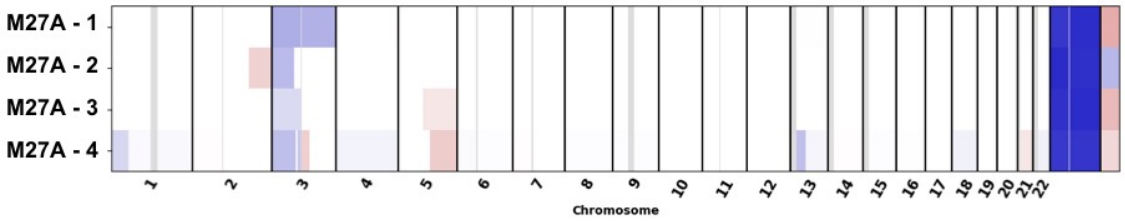


Figure 1. **SCNA profile for M27A.** Genome wide SCNA profile for all four tumors across one resection from patient M27A. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

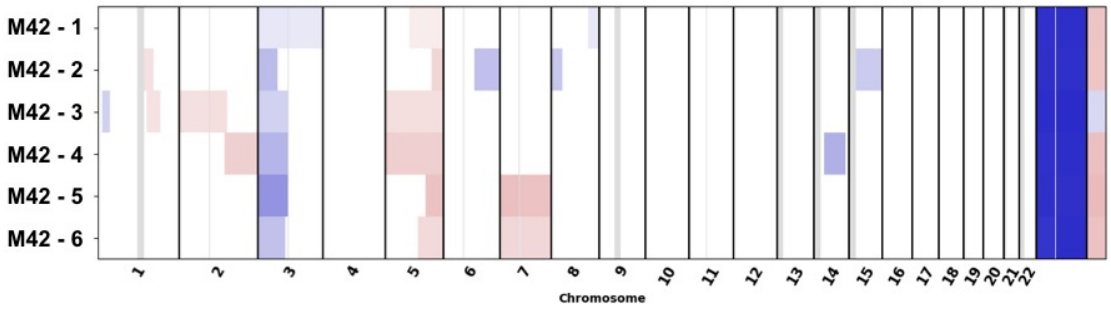


Figure 2. SCNA profile for M42. Genome wide SCNA profile for all six tumors across one resection from patient M42. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

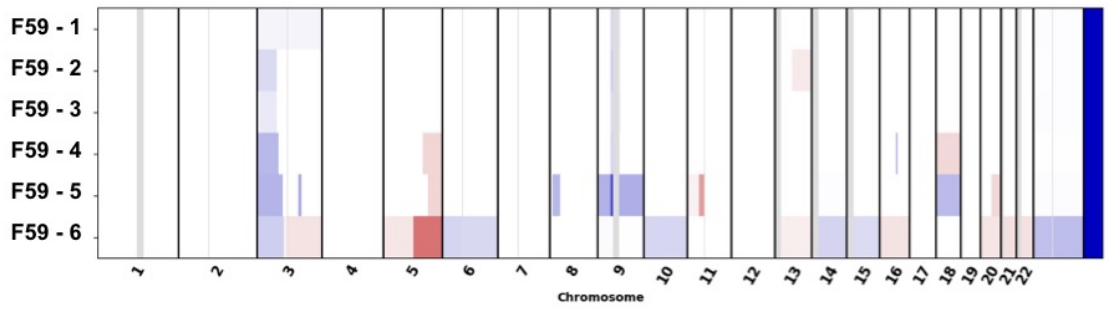


Figure 3. SCNA profile for F59. Genome wide SCNA profile for all six tumors across one resection from patient F59. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

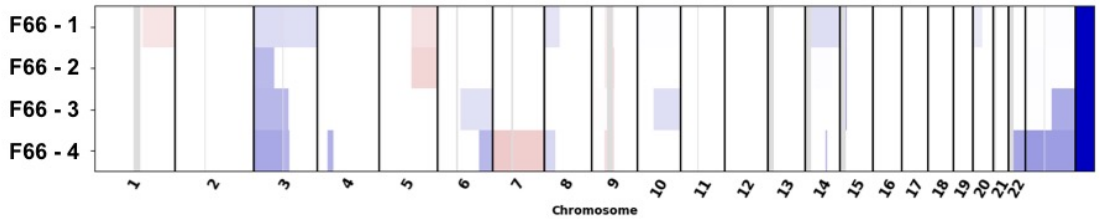


Figure 4. SCNA profile for F66. Genome wide SCNA profile for all four tumors across one resection from patient F66. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

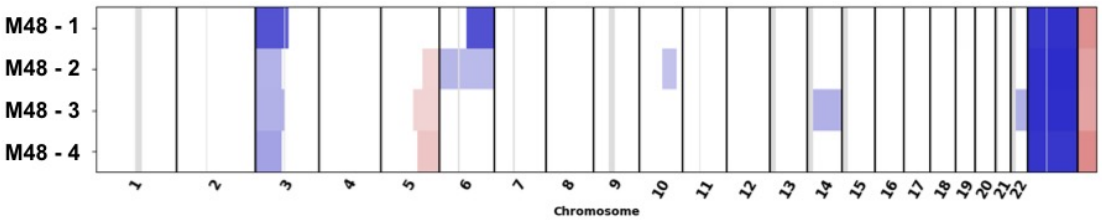


Figure 5. SCNA profile for M48. Genome wide SCNA profile for all four tumors across one resection from patient M48. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

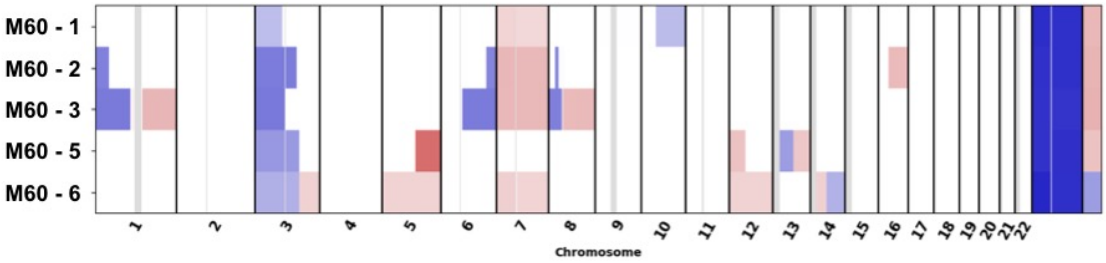


Figure 6. SCNA profile for M60. Genome wide SCNA profile for all six tumors across one resection from patient M60. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

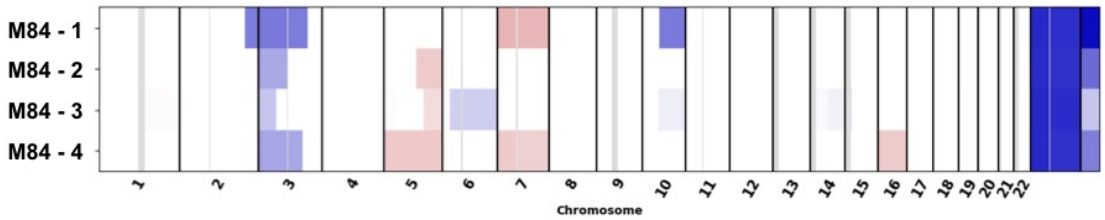


Figure 7. SCNA profile for M84. Genome wide SCNA profile for all four tumors across one resection from patient M84. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

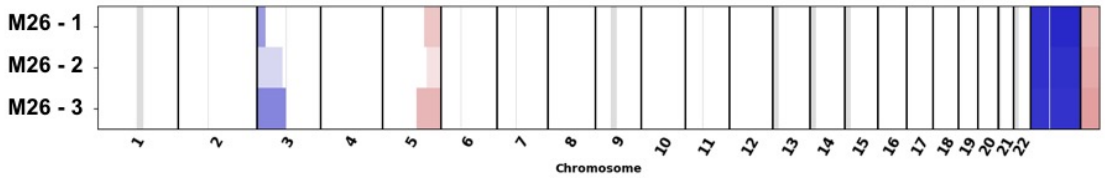


Figure 8. SCNA profile for M26. Genome wide SCNA profile for all three tumors across one resection from patient M26. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

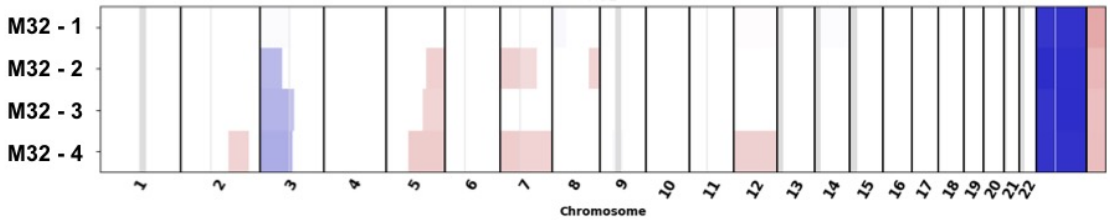


Figure 9. SCNA profile for M32. Genome wide SCNA profile for all four tumors across one resection from patient M32. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

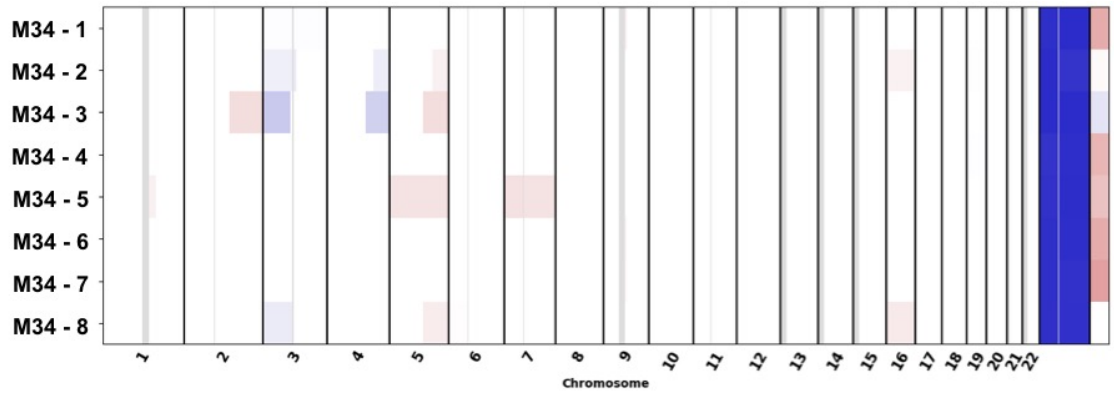


Figure 10. SCNA profile for M34. Genome wide SCNA profile for all eight tumors across one resection from patient M34. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue. M34-4, M34-6, and M34-7 did not contain any arm-level SCNA.

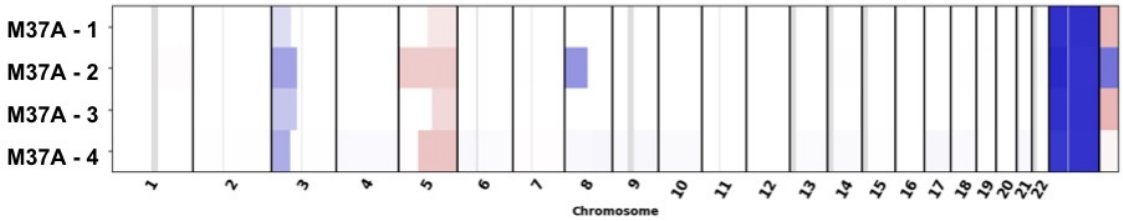


Figure 11. **SCNA profile for M37A.** Genome wide SCNA profile for all four tumors across one resection from patient M37A. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

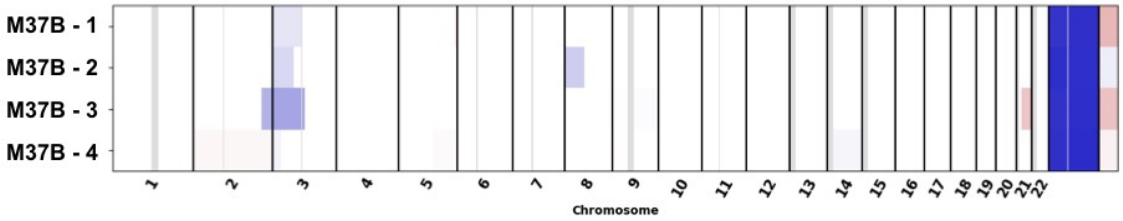


Figure 12. SCNA profile for M27B. Genome wide SCNA profile for all four tumors across one resection from patient M27B. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

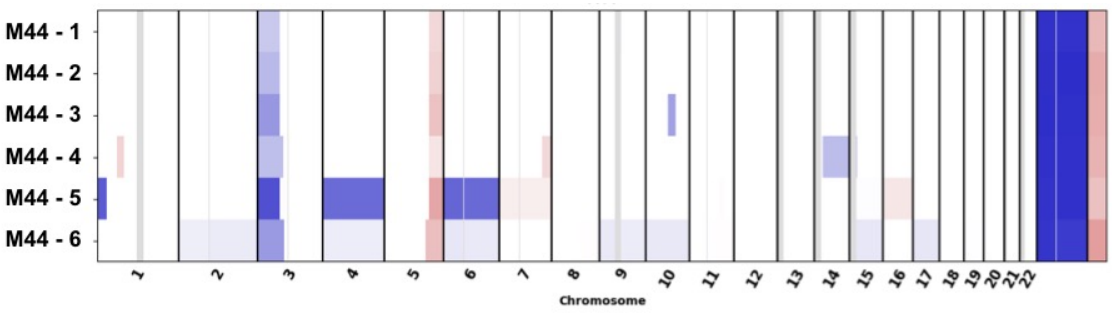


Figure 13. SCNA profile for M44. Genome wide SCNA profile for all six tumors across one resection from patient M44. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

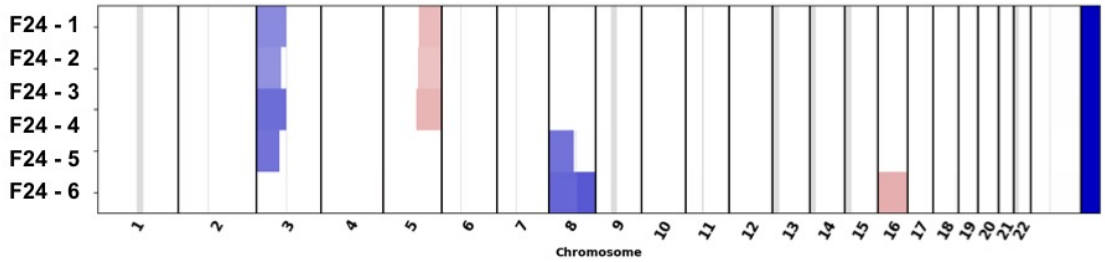


Figure 14. **SCNA profile for F24.** Genome wide SCNA profile for all six tumors across one resection from patient F24. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

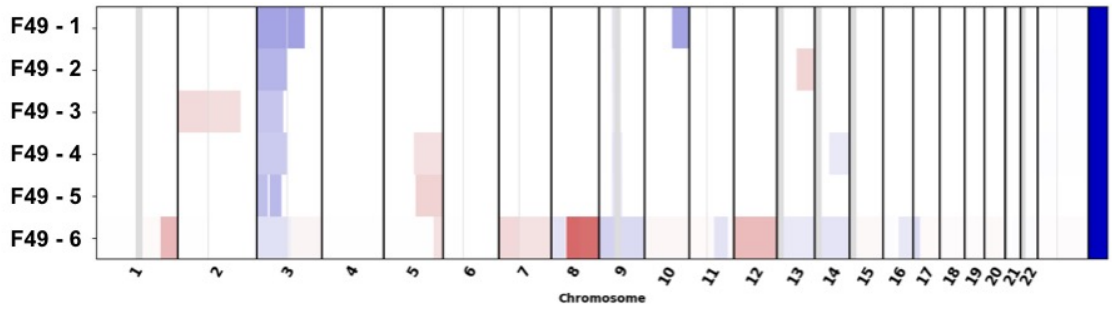


Figure 15. SCNA profile for F49. Genome wide SCNA profile for all six tumors across one resection from patient F49. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

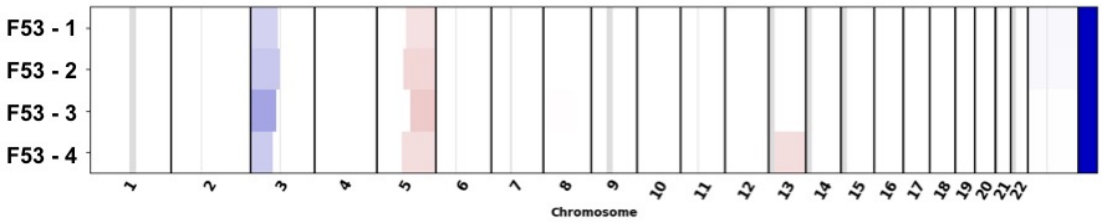


Figure 16. **SCNA profile for F53.** Genome wide SCNA profile for all four tumors across one resection from patient F53. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

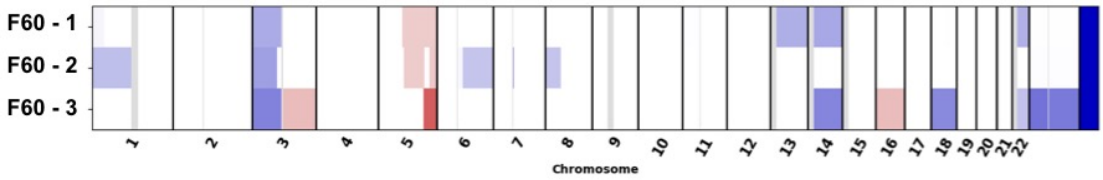


Figure 17. **SCNA profile for F60.** Genome wide SCNA profile for all three tumors across one resection from patient F60. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

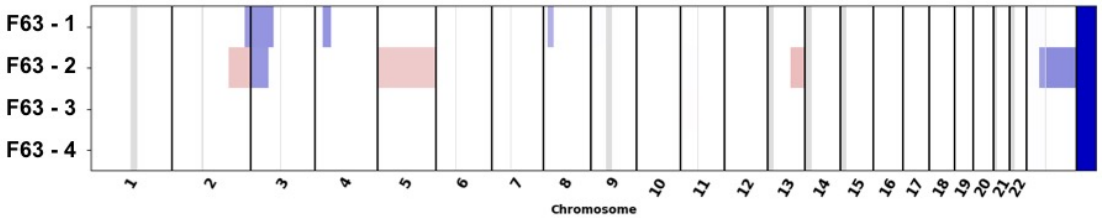


Figure 18. SCNA profile for F63. Genome wide SCNA profile for all four tumors across one resection from patient F63. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue. F63-3 and F63-4 did not contain any arm-level SCNA.

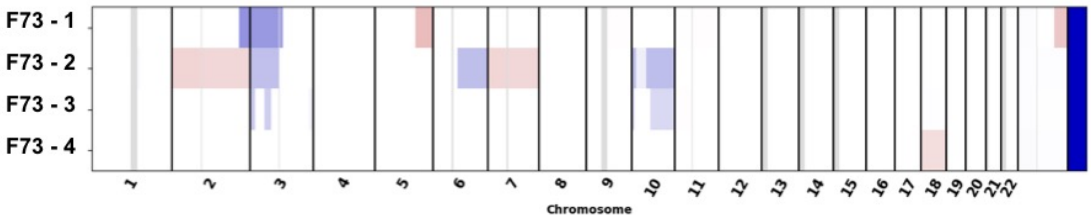


Figure 19. SCNA profile for F73. Genome wide SCNA profile for all four tumors across one resection from patient F73. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue.

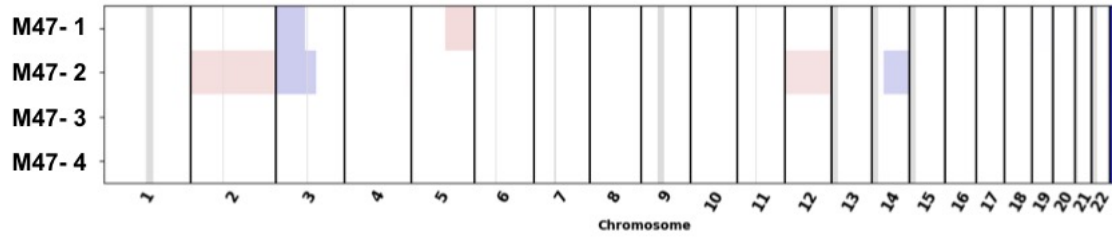


Figure 20. SCNA profile for M47. Genome wide SCNA profile for all four tumors across one resection from patient M47. Each tumor is denoted along the vertical axis. Chromosome numbers are given along the horizontal axis. The heatmap provides $\log_2$ copy-ratios, where gains are in red and losses are in blue. M47-3 and M47-4 did not contain any arm-level SCNA.

References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249.

Alexandrov, L., Kim, J., Haradhvala, N.J., Huang, M.N., Ng, A.W.T., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E., Lopez-Bigas, N., et al. (2018). The Repertoire of Mutational Signatures in Human Cancer. BioRxiv 322859.

Alexandrov, L.B., and Stratton, M.R. (2014). Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. Current Opinion in Genetics & Development 24, 52–60.

Alexandrov, L.B., Jones, P.H., Wedge, D.C., Sale, J.E., Campbell, P.J., Nik-Zainal, S., and Stratton, M.R. (2015). Clock-like mutational processes in human somatic cells. Nature Genetics 47, 1402–1407.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al. (2013). Signatures of mutational processes in human cancer. Nature 500, 415.

"American Cancer Society | Cancer Facts & Statistics." (Aug. 28, 2018). American Cancer Society | Cancer Facts & Statistics, cancerstatisticscenter.cancer.org/.

Amos, C.I., Dennis, J., Wang, Z., Byun, J., Schumacher, F.R., Gayther, S.A., Casey, G., Hunter, D.J., Sellers, T.A., Gruber, S.B., et al. (2017). The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. Cancer Epidemiol Biomarkers Prev 26, 126–135.

Anderson, K., Lutz, C., van Delft, F.W., Bateman, C.M., Guo, Y., Colman, S.M., Kempski, H., Moorman, A.V., Titley, I., Swansbury, J., et al. (2011). Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature 469, 356–361.

Armitage, P., and Doll, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. British Journal of Cancer 8, 1.

Armitage, P., and Doll, R. (1957). A Two-stage Theory of Carcinogenesis in Relation to the Age Distribution of Human Cancer. British Journal of Cancer 11, 161–169.

Baca, S.C., Prandi, D., Lawrence, M.S., Mosquera, J.M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T.Y., Ghandi, M., et al. (2013). Punctuated Evolution of Prostate Cancer Genomes. Cell 153, 666–677.

Beerenwinkel, N., Antal, T., Dingli, D., Traulsen, A., Kinzler, K.W., Velculescu, V.E., Vogelstein, B., and Nowak, M.A. (2007). Genetic progression and the waiting time to cancer. PLoS Computational Biology 3, e225.

Benita, Y., Kikuchi, H., Smith, A.D., Zhang, M.Q., Chung, D.C., and Xavier, R.J. (2009). An integrative genomics approach identifies Hypoxia Inducible Factor-1 (HIF-1)-target genes that form the core response to hypoxia. Nucleic Acids Research 37, 4587–4602.

Beroukhim, R., Brunet, J.-P., Di Napoli, A., Mertz, K.D., Seeley, A., Pires, M.M., Linhart, D., Worrell, R.A., Moch, H., Rubin, M.A., et al. (2009). Patterns of Gene Expression and Copy-Number Alterations in von-Hippel Lindau Disease-Associated and Sporadic Clear Cell Carcinoma of the Kidney. Cancer Res 69, 4674.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature 463, 899–905.

Bertheau, P., Lehmann-Che, J., Varna, M., Dumay, A., Poirot, B., Porcher, R., Turpin, E., Plassa, L.-F., de Roquancourt, A., Bourstyn, E., et al. (2013). p53 in breast cancer subtypes and new insights into response to chemotherapy. The Breast 22, S27–S29.

Bielski, C.M., Zehir, A., Penson, A.V., Donoghue, M.T.A., Chatila, W., Armenia, J., Chang, M.T., Schram, A.M., Jonsson, P., Bandlamudi, C., et al. (2018). Genome doubling shapes the evolution and prognosis of advanced cancers. Nature Genetics 50, 1189.

Blackford, A., Parmigiani, G., Kensler, T.W., Wolfgang, C., Jones, S., Zhang, X., Parsons, D.W., Lin, J.C.-H., Leary, R.J., Eshleman, J.R., et al. (2009). Genetic Mutations Associated With Cigarette Smoking in Pancreatic Cancer. Cancer Research 69, 3681–3688.

Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., Karchin, R., Kinzler, K.W., Vogelstein, B., and Nowak, M.A. (2010). Accumulation of driver and passenger mutations during tumor progression. Proceedings of the National Academy of Sciences 107, 18545–18550.

Brandt, A., Bermejo, J.L., Sundquist, J., and Hemminki, K. (2008). Age of onset in familial cancer. Ann Oncol 19, 2084–2088.

Brennan, C.W., Verhaak, R.G.W., McKenna, A., Campos, B., Noushmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al. (2013). The Somatic Genomic Landscape of Glioblastoma. Cell 155, 462–477.

Brinkmann, B., Klintschar, M., Neuhuber, F., Hühne, J., and Rolf, B. (1998). Mutation Rate in Human Microsatellites: Influence of the Structure and Length of the Tandem Repeat. The American Journal of Human Genetics 62, 1408–1415.

Cancer Genome Atlas Research Network, Brat, D.J., Verhaak, R.G.W., Aldape, K.D., Yung, W.K.A., Salama, S.R., Cooper, L.A.D., Rheinbay, E., Miller, C.R., Vitucci, M., et al. (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. N. Engl. J. Med. 372, 2481–2498.

Carter, H., Marty, R., Hofree, M., Gross, A.M., Jensen, J., Fisch, K.M., Wu, X., DeBoever, C., Van Nostrand, E.L., Song, Y., et al. (2017). Interaction Landscape of Inherited Polymorphisms with Somatic Events in Cancer. Cancer Discovery 7, 410–423.

Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. 30, 413–421.

Chen, W., Hill, H., Christie, A., Kim, M.S., Holloman, E., Pavia-Jimenez, A., Homayoun, F., Ma, Y., Patel, N., Yell, P., et al. (2016). Targeting renal cell carcinoma with a HIF-2 antagonist. Nature 539, 112–117.

Cho, H., Du, X., Rizzi, J.P., Liberzon, E., Chakraborty, A.A., Gao, W., Carvo, I., Signoretti, S., Bruick, R.K., Josey, J.A., et al. (2016). On-target efficacy of a HIF-2α antagonist in preclinical kidney cancer models. Nature 539, 107–111.

Chojniak, R., and Younes, R.N. (2003). Pulmonary Metastases Tumor Doubling Time: Assessment by Computed Tomography. American Journal of Clinical Oncology 26.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature Biotechnology 31, 213–219.

Crabtree, M., Sieber, O.M., Lipton, L., Hodgson, S.V., Lamlum, H., Thomas, H.J.W., Neale, K., Phillips, R.K.S., Heinimann, K., and Tomlinson, I.P.M. (2003). Refining the relation between /`first hits/' and /`second hits/' at the APC locus: the /`loose fit/' model and evidence for differences in somatic mutation spectra among patients. Oncogene 22, 4257–4265.

D'Antonio, M., Tamayo, P., Mesirov, J.P., and Frazer, K.A. (2016). Kataegis Expression Signature in Breast Cancer Is Associated with Late Onset, Better Prognosis, and Higher HER2 Levels. Cell Reports 16, 672–683.

Davis, A., Gao, R., and Navin, N. (2017). Tumor evolution: Linear, branching, neutral or punctuated? Biochim Biophys Acta 1867, 151–161.

De, S., and Michor, F. (2011). DNA secondary structures and epigenetic determinants of cancer genome evolution. Nat. Struct. Mol. Biol. 18, 950–955.

Dong, C., Guo, Y., Yang, H., He, Z., Liu, X., and Wang, K. (2016). iCAGES: integrated CAncer GEnome Score for comprehensively prioritizing driver genes in personal cancer genomes. Genome Medicine 8, 135.

Droz, D., Zachar, D., Charbit, L., Gogusev, J., Chrétein, Y., and Iris, L. (1990). Expression of the human nephron differentiation molecules in renal cell carcinomas. Am. J. Pathol. 137, 895–905.

Duffey, B.G., Choyke, P.L., Glenn, G., Grubb, R.L., Venzon, D., Linehan, W.M., and Walther, M.M. (2004). The relationship between renal tumor size and metastases in patients with von Hippel-Lindau disease. J. Urol. 172, 63–65.

Dunford, A., Weinstock, D.M., Savova, V., Schumacher, S.E., Cleary, J.P., Yoda, A., Sullivan, T.J., Hess, J.M., Gimelbrant, A.A., Beroukhim, R., et al. (2017). Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. Nat Genet 49, 10–16.

Dworkin, A.M., Ridd, K., Bautista, D., Allain, D.C., Iwenofu, O.H., Roy, R., Bastian, B.C., and Toland, A.E. (2010). Germline Variation Controls the Architecture of Somatic Alterations in Tumors. PLoS Genetics 6, e1001136.

Eble, J.N., and Bonsib, S.M. (1998). Extensively cystic renal neoplasms: cystic nephroma, cystic partially differentiated nephroblastoma, multilocular cystic renal cell carcinoma, and cystic hamartoma of renal pelvis. Semin Diagn Pathol 15, 2–20.

Eckel-Passow, J.E., Serie, D.J., Cheville, J.C., Ho, T.H., Kapur, P., Brugarolas, J., Thompson, R.H., Leibovich, B.C., Kwon, E.D., Joseph, R.W., et al. (2017). BAP1 and PBRM1 in metastatic clear cell renal cell carcinoma: tumor heterogeneity and concordance with paired primary tumor. BMC Urology 17, 19.

Eeles, R., Al Olama, A.A., Berndt, S., Wiklund, F., Conti, D.V., Ahmed, M., Benlloch, S., Easton, D., Kraft, P., Chanock, S.J., et al. (2017). Prostate cancer meta-analysis from more than 145,000 men to identify 65 novel prostate cancer susceptibility loci. JCO 35, 1–1.

Eirew, P., Steif, A., Khattra, J., Ha, G., Yap, D., Farahani, H., Gelmon, K., Chia, S., Mar, C., Wan, A., et al. (2015). Dynamics of genomic clones in breast cancer patient xenografts at single cell resolution. Nature 518, 422–426.

Elorza, A., Soro-Arnáiz, I., Meléndez-Rodríguez, F., Rodríguez-Vaello, V., Marsboom, G., de Cárcer, G., Acosta-Iborra, B., Albacete-Albacete, L., Ordóñez, A., Serrano-Oviedo, L., et al. (2012). HIF2α acts as an mTORC1 activator through the amino acid carrier SLC7A5. Mol. Cell 48, 681–691.

Ewart-Toland, A., Briassouli, P., de Koning, J.P., Mao, J.-H., Yuan, J., Chan, F., MacCarthy-Morrogh, L., Ponder, B.A.J., Nagase, H., Burn, J., et al. (2003). Identification of Stk6/STK15 as a candidate low-penetrance tumor-susceptibility gene in mouse and human. Nat Genet 34, 403–412.

Fearon, E.R. (2011). Molecular genetics of colorectal cancer. Annu Rev Pathol 6, 479–507.

Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell 61, 759–767.

Fei, S.S., Mitchell, A.D., Heskett, M.B., Vocke, C.D., Ricketts, C.J., Peto, M., Wang, N.J., Sönmez, K., Linehan, W.M., and Spellman, P.T. (2016). Patient-specific factors influence somatic variation patterns in von Hippel–Lindau disease renal tumours. Nature Communications 7, 11588.

Ficarra, V., Guillè, F., Schips, L., Taille, A. de la, Galetti, T.P., Tostain, J., Cindolo, L., Novara, G., Zigeuner, R., Bratti, E., et al. (2005). Proposal for revision of the TNM classification system for renal cell carcinoma. Cancer 104, 2116–2123.

Fisher, R., Horswell, S., Rowan, A., Salm, M.P., de Bruin, E.C., Gulati, S., McGranahan, N., Stares, M., Gerlinger, M., Varela, I., et al. (2014). Development of synchronous VHL syndrome tumors reveals contingencies and constraints to tumor evolution. Genome Biology 15, 433.

Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L., et al. (2017). COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res 45, D777–D783.

Frew, I.J., and Moch, H. (2015). A clearer view of the molecular complexity of clear cell renal cell carcinoma. Annu Rev Pathol 10, 263–289.

Fudenberg, G., Getz, G., Meyerson, M., and Mirny, L.A. (2011). High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. Nat. Biotechnol. 29, 1109–1113.

Fuhrman, S.A., Lasky, L.C., and Limas, C. (1982). Prognostic significance of morphologic parameters in renal cell carcinoma. Am. J. Surg. Pathol. 6, 655–663.

Furukawa, H., Iwata, R., and Moriyama, N. (2001). Growth Rate of Pancreatic Adenocarcinoma: Initial Clinical Experience. Pancreas 22.

Gao, R., Davis, A., McDonald, T.O., Sei, E., Shi, X., Wang, Y., Tsai, P.-C., Casasent, A., Waters, J., Zhang, H., et al. (2016). Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. Nature Genetics 48, 1119–1130.

Garcia-Murillas, I., Schiavon, G., Weigelt, B., Ng, C., Hrebien, S., Cutts, R.J., Cheang, M., Osin, P., Nerurkar, A., Kozarewa, I., et al. (2015). Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. Science Translational Medicine 7, 302ra133-302ra133.

Gerlinger, M., Horswell, S., Larkin, J., Rowan, A.J., Salm, M.P., Varela, I., Fisher, R., McGranahan, N., Matthews, N., Santos, C.R., et al. (2014). Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat Genet 46, 225–233.

Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., et al. (2012). Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. N Engl J Med 366, 883–892.

Gerstung, M., Jolly, C., Leshchiner, I., Dentro, S.C., Gonzalez, S., Mitchell, T.J., Rubanova, Y., Anur, P., Rosebrock, D., and Yu, K. (2017). The evolutionary history of 2,658 cancers. BioRxiv 161562.

Glaire, M.A., Brown, M., Church, D.N., and Tomlinson, I. (2017). Cancer predisposition syndromes: lessons for truly precision medicine. J. Pathol. 241, 226–235.

Gojobori, T., Li, W.-H., and Graur, D. (1982). Patterns of nucleotide substitution in pseudogenes and functional genes. Journal of Molecular Evolution 18, 360–369.

Gossage, L., Eisen, T., and Maher, E.R. (2015). VHL, the story of a tumour suppressor gene. Nature Reviews Cancer 15, 55–64.

Gu, Y.-F., Cohn, S., Christie, A., McKenzie, T., Wolff, N., Do, Q.N., Madhuranthakam, A.J., Pedrosa, I., Wang, T., Dey, A., et al. (2017). Modeling Renal Cell Carcinoma in Mice: Bap1 and Pbrm1 Inactivation Drive Tumor Grade. Cancer Discov 7, 900–917.

Gunawan, B., Huber, W., Holtrup, M., von Heydebreck, A., Efferth, T., Poustka, A., Ringert, R.H., Jakse, G., and Füzesi, L. (2001). Prognostic impacts of cytogenetic findings in clear cell renal cell carcinoma: gain of 5q31-qter predicts a distinct clinical phenotype with favorable prognosis. Cancer Res. 61, 7731–7738.

Haase, V.H., Glickman, J.N., Socolovsky, M., and Jaenisch, R. (2001). Vascular tumors in livers with targeted inactivation of the von Hippel–Lindau tumor suppressor. PNAS 98, 1583–1588.

Harlander, S., Schönenberger, D., Toussaint, N.C., Prummer, M., Catalano, A., Brandt, L., Moch, H., Wild, P.J., and Frew, I.J. (2017). Combined Vhl, Trp53 and Rb1 mutation causes clear cell renal cell carcinoma in mice. Nat Med 23, 869–877.

Hienonen, T., Salovaara, R., Mecklin, J.-P., Järvinen, H., Karhu, A., and Aaltonen, L.A. (2006). Preferential amplification of AURKA 91A (Ile31) in familial colorectal cancers. International Journal of Cancer 118, 505–508.

Hoffman, M.A., Ohh, M., Yang, H., Klco, J.M., Ivan, M., and Kaelin, W.G. (2001). von Hippel-Lindau protein mutants linked to type 2C VHL disease preserve the ability to downregulate HIF. Hum. Mol. Genet. 10, 1019–1027.

Hou, W., and Ji, Z. (2018). Generation of autochthonous mouse models of clear cell renal cell carcinoma: mouse models of renal cell carcinoma. Experimental & Molecular Medicine 50, 30.

Jenkins, Z.A., van Kogelenberg, M., Morgan, T., Jeffs, A., Fukuzawa, R., Pearl, E., Thaller, C., Hing, A.V., Porteous, M.E., Garcia-Minaur, S., et al. (2009). Germline mutations in WTX cause a sclerosing skeletal dysplasia but do not predispose to tumorigenesis. Nat Genet 41, 95–100.

Jönsson, G., Naylor, T.L., Vallon-Christersson, J., Staaf, J., Huang, J., Ward, M.R., Greshock, J.D., Luts, L., Olsson, H., Rahman, N., et al. (2005). Distinct Genomic Profiles in Hereditary Breast Tumors Identified by Array-Based Comparative Genomic Hybridization. Cancer Res 65, 7612.

Joseph, R.W., Kapur, P., Serie, D.J., Parasramka, M., Ho, T.H., Cheville, J.C., Frenkel, E., Parker, A.S., and Brugarolas, J. (2016). Clear Cell Renal Cell Carcinoma Subtypes Identified by BAP1 and PBRM1 Expression. J. Urol. 195, 180–187.

Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., et al. (2013). Mutational landscape and significance across 12 major cancer types. Nature 502, 333–339.

Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., et al. (2013). Mutational landscape and significance across 12 major cancer types. Nature 502, 333.

Kim, H., Zheng, S., Amini, S.S., Virk, S.M., Mikkelsen, T., Brat, D.J., Grimsby, J., Sougnez, C., Muller, F., Hu, J., et al. (2015). Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. Genome Research 25, 316–327.

King, M.-C., Marks, J.H., Mandell, J.B., and New York Breast Cancer Study Group (2003). Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science 302, 643–646.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. 46, 310–315.

Klatte, T., Rao, P.N., de Martino, M., LaRochelle, J., Shuch, B., Zomorodian, N., Said, J., Kabbinavar, F.F., Belldegrun, A.S., and Pantuck, A.J. (2009). Cytogenetic profile predicts prognosis of patients with clear cell renal cell carcinoma. J. Clin. Oncol. 27, 746–753.

Knudson, A.G. (1971). Mutation and Cancer: Statistical Study of Retinoblastoma. Proceedings of the National Academy of Sciences of the United States of America 68, 820–823.

Kondo, K., Kim, W.Y., Lechpammer, M., and Kaelin, W.G. (2003). Inhibition of HIF2alpha is sufficient to suppress pVHL-defective tumor growth. PLoS Biol. 1, E83.

Kreso, A., O'Brien, C.A., van Galen, P., Gan, O.I., Notta, F., Brown, A.M.K., Ng, K., Ma, J., Wienholds, E., Dunant, C., et al. (2013). Variable Clonal Repopulation Dynamics Influence Chemotherapy Response in Colorectal Cancer. Science 339, 543.

Krzystolik, K., Jakubowska, A., Gronwald, J., Krawczyński, M.R., Drobek-Słowik, M., Sagan, L., Cyryłowski, L., Lubiński, W., Lubiński, J., and Cybulski, C. (2014). Large deletion causing von Hippel-Lindau disease and hereditary breast cancer syndrome. Hered Cancer Clin Pract 12, 16.

Landi, M.T., Bauer, J., Pfeiffer, R.M., Elder, D.E., Hulley, B., Minghetti, P., Calista, D., Kanetsky, P.A., Pinkel, D., and Bastian, B.C. (2006). MC1R Germline Variants Confer Risk for BRAF-Mutant Melanoma. Science 313, 521.

Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 42, D980–D985.

Lattouf, J.-B., Pautler, S.E., Reaume, M.N., Kim, R.H., Care, M., Green, J., So, A., Violette, P.D., Saliba, I., Major, P., et al. (2016). Structured assessment and followup for patients with hereditary kidney tumour syndromes. Can Urol Assoc J 10, E214–E222.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 505, 495.

Le Calvez, F., Mukeria, A., Hunt, J.D., Kelm, O., Hung, R.J., Tanière, P., Brennan, P., Boffetta, P., Zaridze, D.G., and Hainaut, P. (2005). TP53 and KRAS Mutation Load and Types in Lung Cancers in Relation to Tobacco Smoke: Distinct Patterns in Never, Former, and Current Smokers. Cancer Res 65, 5076.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Maddock, I.R., Moran, A., Maher, E.R., Teare, M.D., Norman, A., Payne, S.J., Whitehouse, R., Dodd, C., Lavin, M., Hartley, N., et al. (1996). A genetic register for von Hippel-Lindau disease. J. Med. Genet. 33, 120–127.

Maher, E.R., Bentley, E., Payne, S.J., Latif, F., Richards, F.M., Chiano, M., Hosoe, S., Yates, J.R., Linehan, M., and Barton, D.E. (1992). Presymptomatic diagnosis of von Hippel-Lindau disease with flanking DNA markers. Journal of Medical Genetics 29, 902–905.

Maldonado, J.L., Fridlyand, J., Patel, H., Jain, A.N., Busam, K., Kageshita, T., Ono, T., Albertson, D.G., Pinkel, D., and Bastian, B.C. (2003). Determinants of BRAF Mutations in Primary Melanomas. JNCI: Journal of the National Cancer Institute 95, 1878–1890.

Maley, C.C., Galipeau, P.C., Finley, J.C., Wongsurawat, V.J., Li, X., Sanchez, C.A., Paulson, T.G., Blount, P.L., Risques, R.-A., Rabinovitch, P.S., et al. (2006). Genetic clonal diversity predicts progression to esophageal adenocarcinoma. Nature Genetics 38, 468–473.

Martincorena, I., Raine, K.M., Gerstung, M., Dawson, K.J., Haase, K., Van Loo, P., Davies, H., Stratton, M.R., and Campbell, P.J. (2017). Universal Patterns of Selection in Cancer and Somatic Tissues. Cell 171, 1029-1041.e21.

Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D.C., Fullam, A., Alexandrov, L.B., Tubio, J.M., et al. (2015). High burden and pervasive positive selection of somatic mutations in normal human skin. Science 348, 880.

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

Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Lush, M.J., Maranian, M.J., Bolla, M.K., Wang, Q., Shah, M., et al. (2015). Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nature Genetics 47, 373–380.

Milholland, B., Dong, X., Zhang, L., Hao, X., Suh, Y., and Vijg, J. (2017). Differences between germline and somatic mutation rates in humans and mice. Nature Communications 8, 15183.

Minardi, D., Lucarini, G., Milanese, G., Primio, R.D., Montironi, R., and Muzzonigro, G. (2016). Loss of nuclear BAP1 protein expression is a marker of poor prognosis in patients with clear cell renal cell carcinoma. Urologic Oncology: Seminars and Original Investigations 34, 338.e11-338.e18.

Mitchell, T.J., Turajlic, S., Rowan, A., Nicol, D., Farmery, J.H.R., O'Brien, T., Martincorena, I., Tarpey, P., Angelopoulos, N., Yates, L.R., et al. (2018). Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal. Cell 173, 611-623.e17.

Mitsumori, K., Kittleson, J.M., Itoh, N., Delahunt, B., Heathcott, R.W., Stewart, J.H., McCredie, M.R.E., and Reeve, A.E. (2002). Chromosome 14q LOH in localized clear cell renal cell carcinoma. J. Pathol. 198, 110–114.

Moch, H., Cubilla, A.L., Humphrey, P.A., Reuter, V.E., and Ulbright, T.M. (2016). 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.

Montani, M., Heinimann, K., von Teichman, A., Rudolph, T., Perren, A., and Moch, H. (2010). VHL-gene Deletion in Single Renal Tubular Epithelial Cells and Renal Tubular Cysts: Further Evidence for a Cyst-dependent Progression Pathway of Clear Cell Renal Carcinoma in von Hippel-Lindau Disease: The American Journal of Surgical Pathology 34, 806–815.

Moolgavkar, S.H. (1978). The multistage theory of carcinogenesis and the age distribution of cancer in man. J. Natl. Cancer Inst. 61, 49–52.

Nagy, R., Sweet, K., and Eng, C. (2004). Highly penetrant hereditary cancer syndromes. Oncogene 23, 6445–6470.

Nerli, R., Devaraju, S., Hiremath, M.B., Guntaka, A.K., Patne, P., and Dixit, N. (2014). Tumor doubling time of renal cell carcinoma measured by CT. Indian Journal of Urology : IJU : Journal of the Urological Society of India 30, 153–157.

Neumann, H.H., and Wiestler, O. (1991). Clustering of features and genetics of von Hippel-Lindau syndrome. The Lancet 338, 258.

Ng, P.C., and Henikoff, S. (2003). SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 31, 3812–3814.

Nicholson, A.M., Olpe, C., Hoyle, A., Thorsen, A.-S., Rus, T., Colombé, M., Brunton-Sim, R., Kemp, R., Marks, K., Quirke, P., et al. (2018). Fixation and Spread of Somatic Mutations in Adult Human Colonic Epithelium. Cell Stem Cell 22, 909-918.e8.

Nishikimi, T., Tsuzuki, T., Fujita, T., Sassa, N., Fukatsu, A., Katsuno, S., Yoshino, Y., Hattori, R., and Gotoh, M. (2011). The post-operative pathological prognostic parameters of clear cell renal cell carcinoma in pT1a cases. Pathology International 61, 116–121.

Nordling, C.O. (1953). A New Theory on the Cancer-inducing Mechanism. British Journal of Cancer 7, 68–72.

Nowell, P.C. (1976). The Clonal Evolution of Tumor Cell Populations. Science 194, 23–28.

Pantuck, A.J., Seligson, D.B., Klatte, T., Yu, H., Leppert, J.T., Moore, L., O'Toole, T., Gibbons, J., Belldegrun, A.S., and Figlin, R.A. (2007). Prognostic relevance of the mTOR pathway in renal cell carcinoma: implications for molecular patient selection for targeted therapy. Cancer 109, 2257–2267.

Paraf, F., Chauveau, D., Chrétien, Y., Richard, S., Grünfeld, J.P., and Droz, D. (2000). Renal lesions in von Hippel-Lindau disease: immunohistochemical expression of nephron differentiation molecules, adhesion molecules and apoptosis proteins. Histopathology 36, 457–465.

Park, C., Qian, W., and Zhang, J. (2012). Genomic evidence for elevated mutation rates in highly expressed genes. EMBO Rep. 13, 1123–1129.

Pause, A., Lee, S., Worrell, R.A., Chen, D.Y., Burgess, W.H., Linehan, W.M., and Klausner, R.D. (1997). The von Hippel-Lindau tumor-suppressor gene product forms a stable complex with human CUL-2, a member of the Cdc53 family of proteins. Proc. Natl. Acad. Sci. U.S.A. 94, 2156–2161.

PDQ Cancer Genetics Editorial Board (2018). Genetics of Breast and Gynecologic Cancers (PDQ®): Health Professional Version. In PDQ Cancer Information Summaries, (Bethesda (MD): National Cancer Institute (US)), p.

Peña-Llopis, S., Vega-Rubín-de-Celis, S., Liao, A., Leng, N., Pavía-Jiménez, A., Wang, S., Yamasaki, T., Zhrebker, L., Sivanand, S., Spence, P., et al. (2012). BAP1 loss defines a new class of renal cell carcinoma. Nat Genet 44, 751–759.

Piva, F., Giulietti, M., Occhipinti, G., Santoni, M., Massari, F., Sotte, V., Iacovelli, R., Burattini, L., Santini, D., Montironi, R., et al. (2015). Computational analysis of the mutations in BAP1, PBRM1 and SETD2 genes reveals the impaired molecular processes in renal cell carcinoma. Oncotarget 6, 32161–32168.

Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Auwera, G.A.V. der, Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., et al. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. BioRxiv 201178.

Prowse, A.H., Webster, A.R., Richards, F.M., Richard, S., Olschwang, S., Resche, F., Affara, N.A., and Maher, E.R. (1997). Somatic inactivation of the VHL gene in Von Hippel-Lindau disease tumors. American Journal of Human Genetics 60, 765–771.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

Rankin, E.B., Tomaszewski, J.E., and Haase, V.H. (2006). Renal Cyst Development in Mice with Conditional Inactivation of the von Hippel-Lindau Tumor Suppressor. Cancer Res 66, 2576–2583.

Raval, R.R., Lau, K.W., Tran, M.G.B., Sowter, H.M., Mandriota, S.J., Li, J.-L., Pugh, C.W., Maxwell, P.H., Harris, A.L., and Ratcliffe, P.J. (2005). Contrasting properties of hypoxia-inducible factor 1 (HIF-1) and HIF-2 in von Hippel-Lindau-associated renal cell carcinoma. Mol. Cell. Biol. 25, 5675–5686.

Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. (1997). Improved Splice Site Detection in Genie. Journal of Computational Biology 4, 311–323.

Reva, B., Antipin, Y., and Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 39, e118–e118.

Ricketts, C.J., and Linehan, W.M. (2015). Gender Specific Mutation Incidence and Survival Associations in Clear Cell Renal Cell Carcinoma (CCRCC). PLoS One 10, e0140257–e0140257.

Ricketts, C.J., De Cubas, A.A., Fan, H., Smith, C.C., Lang, M., Reznik, E., Bowlby, R., Gibb, E.A., Akbani, R., Beroukhim, R., et al. (2018). The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. Cell Reports 23, 313-326.e5.

Roberts, S.A., Lawrence, M.S., Klimczak, L.J., Grimm, S.A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G.V., Carter, S.L., Saksena, G., et al. (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nature Genetics 45, 970–976.

Roth, A., Khattra, J., Yap, D., Wan, A., Laks, E., Biele, J., Ha, G., Aparicio, S., Bouchard-Côté, A., and Shah, S.P. (2014). PyClone: Statistical inference of clonal population structure in cancer. Nat Methods 11, 396–398.

Rubin, A.F., and Green, P. (2009). Mutation patterns in cancer genomes. Proceedings of the National Academy of Sciences 106, 21766–21770.

Ruivenkamp, C., Hermsen, M., Postma, C., Klous, A., Baak, J., Meijer, G., and Demant, P. (2003). LOH of PTPRJ occurs early in colorectal cancer and is associated with chromosomal loss of 18q12-21. Oncogene 22, 3472–3474.

Sansregret, L., and Swanton, C. (2017). The Role of Aneuploidy in Cancer Evolution. Cold Spring Harb Perspect Med 7, a028373.

Scelo, G., Purdue, M.P., Brown, K.M., Johansson, M., Wang, Z., Eckel-Passow, J.E., Ye, Y., Hofmann, J.N., Choi, J., Foll, M., et al. (2017). Genome-wide association study identifies multiple risk loci for renal cell carcinoma. Nature Communications 8, 15724.

Schmid, S., Gillessen, S., Binet, I., Brändle, M., Engeler, D., Greiner, J., Hader, C., Heinimann, K., Kloos, P., Krek, W., et al. (2014). Management of von Hippel-Lindau Disease: An Interdisciplinary Review. ORT 37, 761–771.

Shankaran, V., Ikeda, H., Bruce, A.T., White, J.M., Swanson, P.E., Old, L.J., and Schreiber, R.D. (2001). IFNγ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. Nature 410, 1107–1111.

Sheltzer, J.M., Ko, J.H., Replogle, J.M., Habibe Burgos, N.C., Chung, E.S., Meehl, C.M., Sayles, N.M., Passerini, V., Storchova, Z., and Amon, A. (2017). Single-chromosome Gains Commonly Function as Tumor Suppressors. Cancer Cell 31, 240–255.

Shen, C., Beroukhim, R., Schumacher, S.E., Zhou, J., Chang, M., Signoretti, S., and Kaelin, W.G. (2011). Genetic and Functional Studies Implicate HIF1 as a 14q Kidney Cancer Suppressor Gene. Cancer Discovery 1, 222–235.

Shuib, S., Wei, W., Sur, H., Morris, M.R., McMullan, D., Rattenberry, E., Meyer, E., Maxwell, P.H., Kishida, T., Yao, M., et al. (2011). Copy number profiling in von hippel-lindau disease renal cell carcinoma. Genes, Chromosomes and Cancer 50, 479–488.

Silwal-Pandit, L., Vollan, H.K.M., Chin, S.-F., Rueda, O.M., McKinney, S.E., Osako, T., Quigley, D., Kristensen, V.N., Aparicio, S., Borresen-Dale, A.-L., et al. (2014). TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance. Clin Cancer Res clincanres.2943.2013.

Slater, A., Moore, N.R., and Huson, S.M. (2003). The Natural History of Cerebellar Hemangioblastomas in von Hippel-Lindau Disease. American Journal of Neuroradiology 24, 1570–1574.

Stefansson, O.A., Jonasson, J.G., Johannsson, O.T., Olafsdottir, K., Steinarsdottir, M., Valgeirsdottir, S., and Eyfjord, J.E. (2009). Genomic profiling of breast tumours in relation to BRCAabnormalities and phenotypes. Breast Cancer Research 11, R47.

Sukov, W.R., Hodge, J.C., Lohse, C.M., Akre, M.K., Leibovich, B.C., Thompson, R.H., and Cheville, J.C. (2012). ALK alterations in adult renal cell carcinoma: frequency, clinicopathologic features and outcome in a large series of consecutively treated patients. Mod. Pathol. 25, 1516–1525.

Suzuki, H., Aoki, K., Chiba, K., Sato, Y., Shiozawa, Y., Shiraishi, Y., Shimamura, T., Niida, A., Motomura, K., Ohka, F., et al. (2015). Mutational landscape and clonal architecture in grade II and III gliomas. Nature Genetics 47, 458.

Suzuki, Y., Ng, S.B., Chua, C., Leow, W.Q., Chng, J., Liu, S.Y., Ramnarayanan, K., Gan, A., Ho, D.L., Ten, R., et al. (2017). Multiregion ultra-deep sequencing reveals early intermixing and variable levels of intratumoral heterogeneity in colorectal cancer. Molecular Oncology 11, 124–139.

Swanton, C. (2015). Cancer Evolution Constrained by Mutation Order. N Engl J Med 372, 661–663.

Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput. Biol. 12, e1004873.

Taylor, A.M., Shih, J., Ha, G., Gao, G.F., Zhang, X., Berger, A.C., Schumacher, S.E., Wang, C., Hu, H., Liu, J., et al. (2018). Genomic and Functional Approaches to Understanding Cancer Aneuploidy. Cancer Cell 33, 676-689.e3.

The Cancer Genome Atlas Research Network (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature 499, 43.

The Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 45, 1113–1120.

Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Chambers, T., Lopez, J.I., Nicol, D., O'Brien, T., Larkin, J., Horswell, S., et al. (2018a). Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. Cell 173, 581-594.e12.

Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Horswell, S., Chambers, T., O'Brien, T., Lopez, J.I., Watkins, T.B.K., Nicol, D., et al. (2018). Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. Cell 173, 595-610.e11.

Tuupanen, S., Niittymaki, I., Nousiainen, K., Vanharanta, S., Lehtonen, R., Mecklin, J.-P., Nuorva, K., Jarvinen, H., Hautaniemi, S., Karhu, A., et al. (2008). Somatic amplification of the risk allele at colorectal cancer predisposition SNP rs6983267. Cancer Res 68, 3099.

Varshney, N., Kebede, A.A., Owusu-Dapaah, H., Lather, J., Kaushik, M., and Bhullar, J.S. (2017). A Review of Von Hippel-Lindau Syndrome. J Kidney Cancer VHL 4, 20–29.

Vinagre, J., Almeida, A., Pópulo, H., Batista, R., Lyra, J., Pinto, V., Coelho, R., Celestino, R., Prazeres, H., Lima, L., et al. (2013). Frequency of TERT promoter mutations in human cancers. Nature Communications 4, 2185.

Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Smits, A.M.M., and Bos, J.L. (1988). Genetic Alterations during Colorectal-Tumor Development. N Engl J Med 319, 525–532.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer Genome Landscapes. Science (New York, N.Y.) 339, 1546–1558.

Walter, M.J., Shen, D., Ding, L., Shao, J., Koboldt, D.C., Chen, K., Larson, D.E., McLellan, M.D., Dooling, D., Abbott, R., et al. (2012). Clonal Architecture of Secondary Acute Myeloid Leukemia. N Engl J Med 366, 1090–1098.

Walther, M.M., Choyke, P.L., Glenn, G., Lyne, J.C., Rayford, W., Venzon, D., and Linehan, W.M. (1999). Renal cancer in families with hereditary renal cancer: prospective analysis of a tumor size threshold for renal parenchymal sparing surgery. J. Urol. 161, 1475–1479.

Weaver, B.A.A., and Cleveland, D.W. (2006). Does aneuploidy cause cancer? Curr. Opin. Cell Biol. 18, 658–667.

Weinhold, N., Jacobsen, A., Schultz, N., Sander, C., and Lee, W. (2014). Genome-wide analysis of noncoding regulatory mutations in cancer. Nat Genet 46, 1160–1165.

Worrall, J.T., Tamura, N., Mazzagatti, A., Shaikh, N., van Lingen, T., Bakker, B., Spierings, D.C.J., Vladimirou, E., Foijer, F., and McClelland, S.E. (2018). Non-random Mis-segregation of Human Chromosomes. Cell Rep 23, 3366–3380.

Xi, R., Luquette, J., Hadjipanayis, A., Kim, T.-M., and Park, P.J. (2010). BIC-seq: a fast algorithm for detection of copy number alterations based on high-throughput sequencing data. Genome Biol 11, O10.

Xue, Y., Chen, Y., Ayub, Q., Huang, N., Ball, E.V., Mort, M., Phillips, A.D., Shaw, K., Stenson, P.D., Cooper, D.N., et al. (2012). Deleterious- and Disease-Allele Prevalence in Healthy Individuals: Insights from Current Predictions, Mutation Databases, and Population-Scale Resequencing. The American Journal of Human Genetics 91, 1022–1032.

Yang, H., Zhong, Y., Peng, C., Chen, J.-Q., and Tian, D. (2010). Important role of indels in somatic mutations of human cancer genes. BMC Med Genet 11, 128.

Yang, Q., Rasmussen, S.A., and Friedman, J. (2002). Mortality associated with Down's syndrome in the USA from 1983 to 1997: a population-based study. The Lancet 359, 1019–1025.

Yates, L.R., Gerstung, M., Knappskog, S., Desmedt, C., Gundem, G., Van Loo, P., Aas, T., Alexandrov, L.B., Larsimont, D., Davies, H., et al. (2015). Subclonal diversification of primary breast cancer revealed by multiregion sequencing. Nature Medicine 21, 751–759.

Ye, K., Wang, J., Jayasinghe, R., Lameijer, E.-W., McMichael, J.F., Ning, J., McLellan, M.D., Xie, M., Cao, S., Yellapantula, V., et al. (2016). Systematic Discovery of Complex Indels in Human Cancers. Nat Med 22, 97–104.

Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhang, C.-Z., Wala, J., Mermel, C.H., et al. (2013). Pan-cancer patterns of somatic copy number alteration. Nat Genet 45, 1134–1140.

Zhang, J., Fujimoto, J., Zhang, J., Wedge, D.C., Song, X., Zhang, J., Seth, S., Chow, C.-W., Cao, Y., Gumbs, C., et al. (2014). Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science 346, 256.

Zhang, L.-L., Kan, M., Zhang, M.-M., Yu, S.-S., Xie, H.-J., Gu, Z.-H., Wang, H.-N., Zhao, S.-X., Zhou, G.-B., Song, H.-D., et al. (2017). Multiregion sequencing reveals the intratumor heterogeneity of driver mutations in TP53-driven non-small cell lung cancer. Int. J. Cancer 140, 103–108.

Zimmer, M., Doucette, D., Siddiqui, N., and Iliopoulos, O. (2004). Inhibition of hypoxia-inducible factor is sufficient for growth suppression of VHL-/- tumors. Mol. Cancer Res. 2, 89–95.