Comprehensive Molecular Profiling of Cancer Progression and Rare Cancers

by

Lorena Lazo de la Vega

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular and Cellular Pathology)
in the University of Michigan
2019

Doctoral Committee:

Professor Kathleen R. Cho, Co-Chair
Adjunct Associate Professor Scott A. Tomlins, Co-Chair
Professor David G. Beer
Professor Thomas J. Giordano
Assistant Professor Elena M. Stoffel

Lorena Lazo de la Vega
llazodel@umich.edu
ORCID: 0000-0003-3609-0098

ACKNOWLEDGEMENTS

I first want to thank my thesis advisor and mentor Dr. Scott Tomlins for all of his guidance and support throughout the last five years. His love for science and research is very infectious. I will forever be grateful for the opportunity to be part of his laboratory and train under his mentorship. I am also very grateful to have been part of an amazing lab with people who became my family away from home. Their support and guidance both on a scientific and personal level were invaluable during my time here. Specifically, I would like to thank Dr. Daniel Hovelson, Andi Cani, Albert Liu, and Dr. Kei Omata for all their guidance and support. Additionally, I would like to thank Nolan Bick, Mia Samaha, Samantha Rahrig, and Sharath Kumar-Anand, who are all spectacular individuals, for helping me a tremendous amount with my projects during their years as undergraduate students.

My co-mentor, Dr. Kathleen Cho, has also been a very valuable person to have as a mentor these past couple of years. Not only was she a great resource while I was preparing one of my manuscripts but also was someone who challenged me in ways I know made be a better scientist. Additionally, I've been very lucky to have Dr. David Beer, Dr. Thomas Giordano, and Dr. Elena Stoffel on my committee guiding me throughout my years as a graduate student. Their advice was invaluable. Also, my work would not have been possible without such great collaborators. I want to thank Dr. Paul Harms, Dr. Andrew Sciallis, and Dr. Rajesh Rao for all their hard work on the projects we worked on together and for their willingness to teach me about the clinical implications of the project.

I also want to thank all of my previous educators. If it weren't for the many mentors I encountered throughout my life, I would not be here. Tim Reid and Tyler Hoxley were two high school teachers who got me excited about science. I will never forget the day Mr. Reid pulled me out of the class to suggest a summer research program that he thought I should apply to. Despite my concerns, he did everything in his power to give me access to such a great opportunity. This internship at UConn opened my eyes to the world of research and I am forever thankful for his

persistence to have me attend the program. My undergraduate institution was filled with mentors which included Professors Alison Draper, Janet Morrison, and Michelle Kovarik. They were not just amazing educators, but also great mentors who gave me consistent guidance throughout college and throughout the graduate school application process.

And lastly, I want to thank my family for their support throughout these past five years. My parents have always been there for me. They've always given me the guidance, encouragement, and strength to achieve all of my career goals even though they do not know what my work entails. Although most of my extended family is in Peru, they are always there for me. Words cannot express how lucky I am to have such as loving family.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AH- Atypical Hyperplasia
AK- Actinic Keratosis
CCP- Comprehensive Cancer Panel
CGH- Comparative Genomic Hybridization
CH- Complex Hyperplasia
CHEC- Corded and Hyalinized Endometrioid Carcinoma
CIN- Conjunctival/corneal Intraepithelial Neoplasia
CIS- Carcinoma *in situ*
CN- Copy Number
CNA- Copy Number Alteration
CNV- Copy Number Variation
CRC- Colorectal Cancer
CSCC- Cutaneous Squamous Cell Carcinoma
DAVID- Database for Annotation Visualization and Integrated Discovery
dNTP- deoxynucleotide triphosphates
Dollop- Dollo and polymorphism parsimony methods
EC- Endometrial Carcinoma
EIN- Endometrial Intraepithelial Neoplasia
FAO- flow corrected variant allele containing reads
FDP- Flow corrected read depths
FF- Fresh Frozen
FFPE - Formalin Fixed Paraffin Embedded
FISH- Fluorescence *in situ* hybridization
GoF- Gain-of-function
GSEA- Gene Set Enrichment Analysis
H&E- Hematoxylin and Eosin
HNSCC- Head and neck squamous cell carcinoma
HPV- Human Papilloma Virus
IGV- Integrated Genomics Viewer
IHC- Immunohistochemistry
indels - Insertions/deletions
LGEC- Low-Grade Endometrioid Endometrial Carcinoma
LOF- Loss-of-function
LOH- Loss of Heterozygosity
LUSC- Lung Squamous Cell Carcinoma
MCC- Merkel Cell Carcinoma
MCPyV- Merkel Cell Polyomavirus
MMR- Mismatch Repair

MSI- Microsatellite Instability
MSK-IMPACT- Memorial Sloan Kettering Integrated Mutation Profiling of Actionable Cancer Therapies
NGS- Next Generation Sequencing
NIH- National Institute of Health
OCP- Oncomine Comprehensive Panel
ONB- Olfactory Neuroblastoma
OSSN- Ocular Surface Squamous Neoplasms
PCGA- Pre-Cancer Genome Atlas
PeSCC- Penile Squamous Cell Carcinoma
Rpm- Reads per million
SCC- Squamous Cell Carcinoma
SLN- Sentinel Lymph Node
SLNB- Sentinel Lymph Node Biopsy
SNP- Single nucleotide polymorphism
SNV- Single nucleotide variant
TCGA- The Cancer Genome Atlas
VAF- Variant Allele Frequency
WES- Whole exome sequencing
WGS- Whole genome sequencing

ABSTRACT

Comprehensive molecular profiling of the genomic, transcriptomic, and epigenetic landscape of cancers is rapidly evolving due to increased accessibility to next-generation sequencing (NGS) technology and inter-institutional collaborations such as The Cancer Genome Atlas (TCGA). Currently, precision medicine approaches in oncology are being guided by NGS due to the clinical implications of genomic data on directing patient care. Here, we use targeted NGS on routine formalin-fixed, paraffin-embedded (FFPE) tissue to conduct comprehensive molecular profiling to address translational research opportunities. Targeted NGS of FFPE material provides greater access to patient samples, the ability to select for a specific tissue region to maximize tumor content in cases of small lesions, and a more cost-effective method focused on cancer-related genes. By using this approach, we were able to answer a wide-range of questions regarding tumor progression and the genomic landscape of rare cancers.

Since the molecular events driving low-grade endometrioid endometrial carcinoma (LGEC) and squamous cell carcinoma (SCC) development are incompletely understood—like in many cancers—we assessed tumor progression in these two cancers by profiling a series of presumed precursor and invasive lesions. Multi-region profiling of LGEC populations using a highly scalable approach demonstrated clinically-relevant multiclonality and intratumoral heterogeneity. Additional DNA profiling of two models of invasive SCCs and their precursors suggested that the presumed genomic complexity primarily found in invasive disease was also found at the precursor stage. Although we used transcriptomic data to compare precursors and invasive disease in LGECs and SCCs to expand our study, our conclusions were limited due to the need for a greater sample size and additional functional studies. Importantly, however, we demonstrated that our methodology is broadly scalable to enable high-throughput genomic and transcriptomic characterization of precursor and invasive cancer populations from FFPE specimens.

Additionally, we characterized three types of rare cancers where driving alterations that could have diagnostic, prognostic, and therapeutic implications were still largely unknown. First,

we focused on a rare class of soft tissue sarcomas defined by a gene fusion between *CIC* and *DUX4*, which resemble Ewing sarcomas at the histological level. Like Ewing sarcomas, we identified limited somatic driver mutations. However, we identified recurrent known chromosome 8 gain, a deleterious mutation in *ARID1A* (chr 1p36), and novel copy-number alterations (CNAs) including chromosome 1p loss, which is also the locus of *ARID1A*. Therefore, we identified genomic aberrations that can be used to refine the classification of *CIC-DUX4* sarcomas. We also focused on Merkel cell carcinoma (MCC), an aggressive cutaneous neuroendocrine carcinoma, where the development of an additional cutaneous MCC tumor can clinically be interpreted as a second primary MCC tumor or a cutaneous metastasis. Due to the important treatment and prognostic implications of this distinction, we assessed clonality. We identified cases with tumors that were non-clonal (second primary) and clonal (metastasis) and observed that tumors from the same patient arose via the same mechanism (via virus or UV-damage). Lastly, olfactory neuroblastomas (ONBs), also known as esthesioneuroblastomas, are aggressive round-cell tumors. Despite their aggressive course, molecular studies of ONBs have been limited, and targeted therapies are lacking. Here, we identified recurrent potential targetable *FGFR3* alterations associated with overexpression that may represent a novel therapeutic target in ONBs. Through these projects involving the molecular characterization of cancer progression and rare cancers, we identified important basic and clinical insights with the potential to improve patient care.

CHAPTER 1: Introduction

1.1 The Genetic Landscape of Cancer

1.1.1 Etiology of Cancer

Cancer is one of the leading causes of death. In 2015, there were approximately 8.7 million deaths associated with cancer (1). Although tumorigenesis has not been completely characterized, genetic and environmental factors have been found to play an essential role in cancer etiology. The first indication of genetic variation underlying human cancer was the discovery of single-base substitution in *HRAS* (2,3). Likewise, a study on scrotal cancers led to the discovery of carcinogenic effects of tar (4). Further investigation found that the combination of environmental and genetic factors are underlying causes of cancer.

1.1.1.1 Environmental Carcinogens

The development of certain cancers is known to be associated with a combination of external factors, such as air and water pollution, poor diet, obesity, and high concentrations of chemical substances, which can cause genetic alterations. Similarly, in some cases, alcohol and other chemicals such as aflatoxins can cause cancers such as those affecting the liver and colon due to the dysregulation of cell processes as a result of toxic metabolites (5,6). Importantly, each carcinogen is associated with different types of mutation signatures present in the tumor genome, with some factors such as smoking and UV radiation leading to high levels of mutation burden. Specifically, lung, bladder, and oral cancer can harbor a mutation signature unique to tobacco (7,8). Likewise, skin cancers, such as melanoma, basal cell carcinoma and squamous cell carcinoma, can harbor a mutation signature unique to UV induced DNA-damage caused by UVA and UVB rays (9). Mutation signature will be further discussed in Chapter 1.1.2.

1.1.1.2 Cancer Causing Viruses

Cancer causing DNA or RNA viruses result in cancer in approximately 15% of infected people (10). DNA viruses transform infected cells by integrating their DNA into the host cell's

genome, often through deregulation of cell division mediated by inactivation of *TP53* (p53) and *RB1* (pRb) (10,11). In contrast, RNA viruses reverse-transcribe their RNA before it becomes incorporated into the host cell's genome. Although relatively rare in the USA, RNA viruses include Human T-lymphotropic virus-1 (HTLV-1) and hepatitis C virus (HCV), which are associated with adult T-cell leukemia and hepatocellular carcinoma, respectively (12,13).

Certain DNA viruses are known to cause cancer by encoding proteins that bind and inactivate p53 and pRb. Specifically, high risk strains of the human papilloma virus (HPV) such as E6 and E7 can mediate p53 degradation or inhibit pRb, respectively (14,15). As detected by molecular testing, infection by high-risk HPV strains can lead to cancers such as head and neck squamous cell carcinomas (HNSCCs) and cutaneous SCCs (16,17). Importantly, while most HPV infections are self-resolving, some adults develop anogenital cancers due to infection. In fact, HPV-16 and -18 DNA is detected in approximately 70% of cervical cancers (18-20). Therefore, this prompted the development of a vaccine now commonly known as Gardasil 9, which covers 9 HPV strains including HPV-16 and HPV-18, to decrease infection by HPV in hopes of reducing the incidence of cervical cancer (21).

People infected by the herpes virus, such as the Epstein Barr virus (EBV) or human herpes virus 8 (HHV-8), can also develop cancer. The EBV increases the risk for nasopharyngeal cancer, lymphomas, possibly Hodgkin lymphoma, and some cases of stomach cancer (22). HHV-8 is associated with Kaposi sarcoma, a rare slow growing cancer just underneath the skin in patients with a weakened immune system (23). Additionally, the hepatitis B virus (HBV) can cause hepatocellular carcinoma (24).

Merkel cell carcinoma (MCC), a rare cutaneous neuroendocrine cancer later discussed, can be caused by UV-mediated DNA damage or the integration of the Merkel cell polyomavirus virus (MCPyV) which deregulates pRb (25-27). The incidence of MCC in the USA almost doubled from 2000 to 2013 and is expected to continue rising. Therefore, the rising incidence rates and estimated 33-46% disease-specific mortality due to metastasis is increasing interest in raising awareness of MCC diagnosis and management (28).

1.1.1.3 Genetic Predisposition

In a recent study, 8% of 10,389 adult cases across 33 cancer types harbored pathogenic/likely pathogenic germline mutations and CNVs in 21 genes associated with cancer (29). Some of the more common germline variants include those in *BRCA1*, *BRCA2*, *ATM*, *PALB2*,

RET, *NF1*, and *MSH6* (30,31), as well as, *TP53*, *APC*, *BRCA2*, *NF1*, *PMS2*, and *RB1* as reported by the Pediatric Cancer Genome Project (32). Although Huang *et al*. identified 435 pathogenic and 418 likely pathogenic variants in these genes, there were still 640 prioritized variants of uncertain significance (VUS), of which they nominated 47 to be likely pathogenic (29). Therefore, there are still many germline variants that have not been well characterized in terms of cancer predisposition.

Some of the most well-established cancer predisposing germline variants occur in *BRCA1* and *BRCA2*, involved in DNA double-strand break repair, which are best characterized in terms of increasing the risk of breast and ovarian cancer. While 12% of women will develop breast cancer, 72% and 69% of women with a *BRCA1* and *BRCA2* mutation, respectively, will develop breast cancer (33). Similarly, while 1.3% of women will develop ovarian cancer, 44% and 17% of women with a *BRCA1* and *BRCA2* mutation, respectively, will develop ovarian cancer (33). Other cancers can also be linked to *BRCA1* and *BRCA2* such as prostate cancer (34), and pancreatic cancers (35). Currently physicians use both single gene testing and panels for genetic sequencing to inform patients of a hereditary predisposition to certain cancers. For example, the U.S. Preventative Services Task Force recommends to conduct genetic testing for breast/ovarian cancer-related germline variants on women either diagnosed before the age of 50 or with a history of breast cancer in the family. Specifically, genetic testing would be conducted using panels including *BRCA1*/ *BRCA2*, *ATM*, *CDH1*, *CHEK2*, *PTEN*, *PALB2*, and *TP53*, among others (36,37).

Additionally, people with Lynch syndrome, associated with defects in genes involved in DNA mismatch repair such as *MLH1, MSH2, MSH6,* and *PMS2*, have an increased risk of cancers such as colorectal, endometrial, and ovarian cancer. Li-Fraumeni syndrome is caused by a germline mutation in *TP53* and is known to increase the risk of breast cancer, osteosarcoma, leukemia, adrenal gland cancer and soft tissue sarcomas. Furthermore, other syndromes that can predispose a person to cancer involve germline mutations in *PTEN*, *APC*, and *RB1* (38). In addition to germline mutations, there are also germline CNVs, most of which still need to be characterized (39,40).

1.1.2 Genomic Aberrations Characteristic of Cancer

Exploration of the cancer genome has demonstrated how characterizing genomic aberrations can impact cancer treatment. For example, breast cancers over-expressing ERBB2 (HER2) benefited from treatment with the anti-HER2 specific monoclonal antibody trastuzumab (41).

Similarly, activating mutations and fusions of other protein kinases have led to development of small-molecule inhibitors. Unfortunately, only some cancers are driven by the activation of highly targetable oncogenes, while a large fraction of cancers are driven by tumor suppressor loss of function (LOF). Since therapies for LOF alterations would require tumor suppressors to regain function, this presents challenges for therapeutic targeting.

Comprehensive molecular profiling of different cancers conducted by large-scale projects, such as The Cancer Genome Atlas (TCGA) which followed the Human Genome Project, has deepened our understanding of the cancer genome. TCGA was initiated in 2006 as a pilot program by the National Institutes of Health (NIH). While their initial goal was to characterize the genomic landscape of brain and ovarian cancers, they also wanted to test whether a national network of researchers at 20 institutions could collaborate to generate large-scale genomic data. TCGA's first pilot study conducted integrative analysis of DNA copy-number (CN), gene expression, and DNA methylation in 206 glioblastomas, of which 91 were selected for detection of somatic mutations (42). The success of this project motivated TCGA to continue to profile over 30 additional tumor types across 11,000 patients which was conducted between 2006 and 2017. This large-scale project characterized tumor types, defined pan-cancer similarities/differences, gave insight into tumor evolution, and identified novel mutations and mechanisms of therapy resistance with a foundation of next-generation DNA and RNA sequencing to characterize the cancer genome and transcriptome (43). In fact, TCGA inspired the creation of international programs, such as the International Cancer Genome Consortium (ICGC), to characterize the genomic landscape of a wide-range of cancers world-wide (44). Together, these efforts have deepened our knowledge of the genetic events at the chromosome and single-base pair level that characterize cancer.

1.1.2.1 Types of Somatic Chromosomal Abnormalities

Chromosomal abnormalities in cancer cells include structural aberrations and copy number alterations (CNAs). The discovery of the Philadelphia chromosome, t(9;22), in chronic myelogenous leukemia (CML) initiated interest in identifying aberrations at the chromosome level in cancer (45,46). Here, the identification of *BCR-ABL* gene fusions in CML led to the development of imatinib, which inhibits ABL tyrosine kinase and its activated derivatives (47). Additional studies identified similar abnormalities in other types of leukemia and lymphomas, and critically, exploitation of these gene fusions has led to the improvement of treatment and clinical outcomes (48,49).

Translocations have also been observed in solid tumors. Some examples of chromosomal rearrangements generating fusions with potential clinical implications include *TPMRSS2-ETV1*, *TMPRSS2-ETV4*, and *TMPRSS2-ETV5* fusions in prostate cancer (50,51), *RET-NTRK1* fusions in papillary thyroid carcinoma (52), and *PRCC-TFE3* fusions in a molecularly defined subtype of renal cell carcinoma (53). Soft tissue sarcomas are rare and heterogeneous due to the different types of fusions identified in each subtype such as those seen in Ewing family (54) and *CIC-DUX4* (55) sarcomas.

CNAs are also recurrently found in solid tumors. For example, colorectal cancer (CRC) has been associated with recurrent CNAs at the chromosomal level. Specifically, genomic gains affect chromosomes 7, 8q, 13, and 20q while losses are found in chromosomes 4, 8p, 17p, and 18q (56). In CRCs, recurrent CNAs at the chromosomal level have been found to be driven by a potential selection of genes in each region that are associated with colon cancer (57,58). Additionally, CNAs are also seen in focal regions. Specifically, this type of aberration is normally seen as focal amplifications of oncogenes driving cancer such as *MDM2* in liposarcoma and *MYC* and *ERBB2* in breast cancer (59,60). On the contrary, there are recurrent homozygous deletions of *CDKN2A* and *RB1* leading to the deregulation of CDK4 and p53 which induces cell cycle G1 progression (61,62).

1.1.2.2 Somatic Mutations

As previously described, cancer development is associated with a wide-range of carcinogens which initiate the deregulation of many cancer-related genes through DNA-damage in the form of somatic variants. Therefore, large-scale analysis was conducted to link the common causes of cancer to different mutation signatures. This analysis identified 30 types of recurrent mutational "signatures" linked to different mutational processes (63).

While most signatures have been linked to a particular etiology, the etiology for signatures 5, 8, 12, 14, 16-19, 21, 23, 27, 28, and 30 remain unknown (63-65). Signature 24 is found in a subset of liver cancers exhibiting high rates of C>A mutations associated with exposure to aflatoxin. Signature 11 is found in melanoma and glioblastoma which exhibit C>T substitutions associated with alkylating agents. While Signature 1 is found in all cancer types, the spontaneous deamination of 5-methylcytosine, in addition to small insertions and deletions, are correlated with age of cancer diagnosis. Signature 9 has also been associated with cancers that have undergone immunoglobulin gene hypermutation such as chronic lymphocytic leukemia and malignant B-cell

lymphomas, associated with the AID family of cytidine deaminases. Signature 22 is found in urothelial carcinomas and liver cancer which exhibit T>A substitutions associated with exposure to aristolochic acid. Bladder and cervical cancers exhibit APOBEC mutation signatures (Signature 2 and 13) which is associated with the conversion of cytosine to uracil through the process of deamination (63-65).

Two common mutation signatures are caused by exposure to UV-light (Signature 7) and smoking/tobacco chewing (Signature 4 and 29). Both are also characterized by a very high mutation burden. Specifically, skin cancers and cancers of the lip are associated with DNA-damage induced by UV radiation. Not only do these cancers have a high mutation burden, but also most of the mutations are primarily C>T and CC>TT dinucleotide variants. The smoking (and tobacco chewing) mutation signatures, are primarily composed of C>A and CC>AA dinucleotide mutations, and are found in liver cancer, esophageal cancer, head and neck cancer, cancers in the lung, and oral squamous cell carcinoma (63-65).

Additionally, there are mutation signatures associated with mutations of certain genes. Due to germline *BRCA1* and *BRCA2* mutations that predispose patients to breast and ovarian cancer, mutation Signature 3 is characterized by the failure of DNA double-strand break repair by homologous recombination. As mentioned, an important signature to note is the signature of mismatch repair deficiency (Signatures 15, 20, 26) found in stomach cancers, breast cancer, cervical cancer, and uterine carcinoma. Defective DNA mismatch repair induced by aberrations in *MLH* and *MSH* genes cause high mutation rates due to repairing errors during DNA replication. Tumors that exhibit impaired mismatch repair (MMR) can exhibit microsatellite instability (MSI) mutation signatures (Signature 6) characterized by short repetitive sequences in the genome mainly found in colorectal and uterine cancer (63-65). Similarly, recurrent *POLE* mutations (Signature 10), causes a mutation signature associated with deficient DNA proofreading. These mutation signatures are important to note due to the therapeutic implications linked to MMR and MSI status enabling us to predict response to immunotherapy (66,67).

Interestingly, somatic genomic aberrations in conjunction with germline variants in cancer-related genes can inactivate tumor suppressors (68). For example, Huang *et al*. showed cancer samples harboring a pathogenic germline variant also have a somatic mutation in the other copy of the same gene as seen in *ATM*, *BRCA2*, and *MSH6* (29). Similarly, the study also showed that there is strong loss-of-heterozygosity in cancers with a hereditary disposition. In this case, samples

with a *BRCA1* germline mutation had lost their WT allele (29). These results emphasize the importance of incorporating germline and somatic events when characterizing a set of samples.

1.1.3 Tumor Heterogeneity and Cancer Evolution

Tumors show genetic heterogeneity at multiple levels, with increasing interest in intratumoral, interpatient, intermetastatic, and intrametastatic heterogeneity; all which have clinical implications. The presence of different subclones in different metastatic lesions from one patient is intermetastatic, while multiple subclones within a single metastasis is referred to as intrametastatic heterogeneity. The idea that no two patients harbor the same set of genomic alterations is defined as interpatient heterogeneity. Although patients may share a small number of mutations that are in the same gene, many times the actual position of mutations are often different (69-71). In these cases, the recurrently altered genes would most likely harbor driver mutations in either tumor suppressors or oncogenes resulting in the deregulation of cell survival, cell fate, or genome maintenance (72).

Intratumor heterogeneity is characterized by multiple subclones within a primary tumor. Specifically, studies focused on multi-region profiling in a single tumor sample have revealed the diversity of somatic events found within a tumor in multiple cancers including pancreatic (73), ovarian (74), colorectal (75), kidney (76), and breast (77). In the first large scale study, multiregion sequencing and phylogenetic analysis of a renal carcinoma conducted by Gerlinger *et al*. showed that the only genomic alteration found at the trunk of the phylogenetic tree was *VHL*, a known driver of renal carcinoma. This event was followed by different sets of *KDM5C*, *SETD2*, and *MTOR* mutations at different branches of the tree, suggesting convergent evolution in these additional driver genes. More importantly, these studies may explain how sampling bias may lead to the challenges in the validation of biomarkers, since one sample may not capture the full complement of mutations present in the tumor and since subclonal pre-existing alterations may drive therapeutic resistance (76). As an example, patients with colorectal cancer showed resistance to anti-epidermal growth factor receptor (EGFR) therapies in the presence of *KRAS* mutations (78,79). Here, resistance to therapy may not only emerge from the selection of pre-existing *KRAS* mutant clones but also from new mutations arising as a result of continuing mutagenesis (80). Hence, *KRAS* mutation testing is recommended for these patients and future studies may aide in linking genomic data to therapy response.

The evolution of these tumors through selection may result in the outgrowth of certain clones. Due to selective pressures occurring throughout tumor development, studies suggest that while some tumors show a neutral evolution, other tumors may be undergoing positive selection of subpopulations with the greatest fitness. Although driver mutations are normally seen as clonal events since they most likely arise early in tumor evolution, subclonal expansions may occur if later genomic alterations increase tumor fitness, potentially leading to progression (81). In many clinical scenarios, it is important to understand the genetic differences that arise throughout tumorigenesis and to assess clonality by evaluation of shared somatic events, such as the presence or absence of identical mutations and CNAs. As an example, it would be essential to determine whether potentially targetable alterations are clonal, since targeting subclonal mutations would not apply selective pressure to all tumor cells.

1.2 Cancer Progression

1.2.1 Precursors

Colon, pancreatic, breast, and cervical cancers, among others, arise from histologically defined lesions known as precursors. These lesions include colorectal adenoma, pancreatic intraepithelial neoplasia (PanIN), ductal carcinoma *in situ* (DCIS), and cervical intraepithelial neoplasia (CIN), respectively. Pathologists routinely classify these lesions through assessment of abnormal histological and immunohistological/molecular features (82,83). Precise diagnosis is often critical as the presence of precursor lesions defines clinical management in several scenarios (e.g. follow-up screening is different for patients with colorectal adenomas vs. hyperplastic polyps). As discussed in Chapter 2 and 3, these lesions are classified according to a spectrum of malignancy that leads up to a diagnosis of invasive cancer.

Due to the subjectivity in assessing the histologic criteria defining precursor lesions, it can be challenging to correctly classify these lesions. Challenges in diagnosing precursor lesions due to intra-observer and inter-observer affect reproducibility of diagnosis, as seen in cases of potential endometrial carcinoma (84,85). Therefore, molecular characterization of precursors and their invasive counterparts can provide insight into the molecular differences that may improve the accuracy of diagnosis and treatment. Here, we will characterize the genomic alterations in low-grade endometrial carcinoma (LGEC) (Chapter 2) and squamous cell carcinomas (SCC) (Chapter 3), as well as in their respective precursors.

1.2.2 Examples of Recent Molecular Studies Conducted on Precursors

Recent analysis of Barrett's esophagus, a precursor to esophageal adenocarcinoma, and esophageal adenocarcinoma established that the majority of driver genes are mutated at similar frequencies in both lesions. A study using exome sequencing on tumor and adjacent Barrett's esophagus suggested that esophageal adenocarcinomas developed by acquisition of *TP53* mutations followed by genomic instability and oncogene amplification. Additionally, this study also identified cases where the Barrett's esophagus did not share somatic mutations with its associated tumor, suggesting that multi-region sampling is necessary to capture a more accurate picture of the alterations leading up to the development of invasive disease (86).

Shain *et al*. sequenced primary melanomas and adjacent precursor lesions and proposed a sequence of key events. By sequencing melanoma *in situ* and invasive melanomas, this study suggested that *BRAF* V600E and a series of events affecting the MAPK pathway are all early events. However, *TERT* mutations and *CDKN2A* biallelic inactivation is characteristic of invasive melanomas since they were found in intermediate lesions and invasive melanoma (87).

Teixeira *et al*. characterized the genomic, transcriptomic, and epigenomic landscape of precursor and invasive lung squamous cell carcinoma (LUSC). Their analysis provided one of the most comprehensive studies on precursor lesions. By distinguishing between lesions that will regress and the ones that progress, they found chromosomal instability observed through a high rate of copy number gains or losses and differences in methylation to be predictive of which lesions will regress or progress to invasive disease (88).

1.2.3 The Pre-Cancer Genome Atlas (PCGA)

Due to the initial focus on characterizing invasive and advanced disease, there have been little comprehensive efforts towards defining the molecular landscape of precursor lesions. Given that the dynamics of tumor progression can potentially be understood through characterizing precancerous lesions, the National Cancer Institute (NCI) has recently proposed to sequence normal, hyperplastic tissue, dysplastic tissue, carcinoma *in situ* and invasive cancer through the PreCancer Genome Atlas (PCGA). By understanding and predicting the molecular events that cause precursor lesions to transition to invasive cancer, we can develop biomarkers for early-detection and risk stratification (83). Therefore, the PCGA proposes to conduct deep characterization of precursor lesions to capture genomic, spatial, and structural changes involved in cancer initiation and progression (82,83,89). Through DNA and RNA sequencing, among other

methodologies, comprehensive molecular characterization can enable the detection of alterations that can help distinguish between lesions that regress or remain benign, and lesions that are likely to progress to invasive cancer. Factors to also consider involve the tumor microenvironment, since it may help mediate signaling between the tumor and stromal cells for growth and survival, as well as the interaction of tumor cells with the host immune cells (90).

1.3 Clinical Implications of Sequencing Methods

1.3.1 Evolution of Various Sequencing Platforms

In part driven by the desire to genomically characterize cancer, an increasing number of sequencing techniques has been developed over the past two decades. As such, many sophisticated DNA sequencing methods followed Sanger sequencing, the first-generation sequencing method (91). Today, sequencing platforms are cheaper, provide parallel processing of samples (i.e multiplexing), and require lower amounts of DNA/RNA compared to when they were initially developed.

For TCGA to conduct integrative analysis of the cancer genome, they applied a wide-range of techniques involving whole genome sequencing (WGS), whole-exome sequencing (WES), DNA methylation, reverse-phase protein array for protein expression (RPPA), RNA-seq for mRNA and micro-RNA gene expression, and single nucleotide polymorphism (SNP) arrays for somatic CNAs. TCGA started by using high-throughput WGS methods based on Sanger Sequencing to identify variants and SNP-based platforms (Agilent 244K Array, Affymetrix 500L SNPChip, Illumina Infinium 550K BeadChip) to define SNP and copy number variation (CNVs). They also used high-throughput methods to assess mRNA (Affymetrix HT U133A, Affymetrix GeneChip Human Exon 1.0 ST Array, Agilent 244K) and miRNA (Agilent miRNA array) gene expression, in addition to DNA methylation studies using Illumina GoldenGate BeadArray (42). As advances were made to sequencing technology, TCGA shifted their sequencing methods to more novel and efficient methods. More recent TCGA studies have used Affymetrix SNP Arrays for copy number (CN) and loss of heterozygosity (LOH) analysis and Illumina platforms for WGS/WES and RNA-seq, allowing for a more robust analysis of the cancer genome and transcriptome (92,93). Specifically, WGS targets all 3.3 GB in the human genome and provides data on non-coding regions and WES targets the 30 MB in the human exome. While WGS profiles

provide a global view of the cancer genome, WES is more cost efficient for most projects since it targets a smaller region of the human genome and requires less data storage.

While there are many different sequencing platforms available, Illumina and Ion Torrent offer two commonly used sequencing platforms that differ in multiple ways. For example, unlike Illumina which emits light to register sequencing reactions, Ion torrent measures changes in pH during sequencing. Ion Torrent sequencers release deoxynucleotide triphosphates (dNTPs) over the surface of template-bearing beads. When the dNTPs are incorporated to the growing DNA strands, ionic sensors monitor the hydrogen release since protons released during incorporation induce a change in pH of the bead-containing-well (94,95). In contrast, Illumina uses sequencing by synthesis (SBS) technology which uses four different fluorescently labelled dNTPs that are added to the nucleic acid chain. Once the dNTP is incorporated, the fluorescent tag is imaged for base calls and then cleaved to allow the next nucleotide to be added to the chain. In addition to the difference in detection methods, there are also other features that differ between each technology as shown in **Table 1.1** (96).

Table 1.1 Comparison of two common Illumina and Ion Torrent platforms (95,96)

System	Ion Torrent (Targeted NGS on Ion Proton)	Illumina (WES/ WGS Hi Seq)
Input	10-50 ng (DNA or RNA)	0.1 – 1 µg (DNA or RNA)
Amplification Method	Emulsion PCR	Bridge PCR
Detection Method	pH	Light
Average Read Length	200 bp	2x150 bp
Reads per run	40- 80 million	5-6 billion
Run Time (hours)	2-4 hours	4-72 hours
Systemic Errors	Indels in homopolymer regions (DNA)	substitutions in GGC and GGT context (DNA)

Due to the differences in sequencing technology, there are different applications and limitations for each (97,98). Specifically, one of the advantages of multiplexed PCR-based sequencing performed on Ion Torrent sequencers is its compatibility with low-input low-quality DNA/RNA **(Table 1.1)**. The DNA/RNA extracted from FFPE tissue have lower concentrations and are of lower quality due to fragmentation, degradation, and formalin fixation induced mutations when compared to DNA/RNA from fresh frozen (FF) tissue. Therefore, multiplexed PCR-based sequencing on Ion Torrent is the most suitable when profiling minute lesions from

FFPE tissue. While using FFPE tissue has other disadvantages, including the lack of standardized FFPE preparation procedures assuring the same fixation process, which are not present when using FF tissue, we find the advantages offered by the use of FFPE tissue to often outweigh the disadvantages. Unlike FF tissue, which requires expensive specialized equipment for storage to prevent sample deterioration at room temperature, FFPE blocks can be stored at room temperature and are required to be archived by pathology laboratories for at least 10 years. Therefore, the collection of well-preserved FPPE blocks is much larger when compared to the limited collection of fresh frozen tissue. Additionally, the use of FFPE tissue allows for histological analysis of hematoxylin and eosin (H&E) slides and tissue scraping through macrodissection or microdissection of consecutive unstained sections or punching from FFPE blocks. Therefore, the highly accessible, routinely available FFPE tissue provides researchers with a greater resource of tumor samples and allows for the molecular characterization of the cancer genome from a diverse set of tissues, including very minute lesions.

Hovelson *et al.* described the development and pre-clinical validation of an Ion Torrent platform based highly-scalable NGS-based assay, called the Oncomine Comprehensive Panel (OCP), which profiles previously identified common somatic genomic aberrations, as illustrated in **Figure 1.1** (99). Using our DNA panels, we generate copy-number (CN) plots and identify high confidence somatic variants by applying extensive filtering using in-house pipelines and a series of established filters conducted manually that exclude potential sequencing artifacts (99-101). Therefore, our lab routinely uses the OCP (99) and the Comprehensive Cancer Panel (CCP), now two commercially available panels. Recently, we shifted toward customizing disease specific targeted DNA and RNA panels that allow us to focus on larger sets of recurrent alterations identified in different tumor types. Due to the different targets and differences in the amount of DNA/RNA required between panels, we identify the most appropriate panel by assessing the amount of DNA and RNA isolated from each sample, and by how well the disease has been previously characterized. For example, rare cancers are more adequately profiled by the CCP which targets 409 genes since the OCP only targets 143 genes. In contrast, the OCP would be applied to cancers that have been previously identified to harbor driver alterations in the genes it profiles, such as endometrioid endometrial carcinoma with recurrent *PTEN*, *PIK3CA*, and *PIK3R1* mutations (102) and lung cancers with recurrent *TP53*, *KRAS*, and *EGFR* alterations (103). Likewise, RNA-seq panels later discussed in Chapter 2 and 4, need to be

chosen carefully. These approaches enable assessment of gene expression and fusion detection depending on panel composition.

Altogether, WES and WGS were necessary during initial studies of the cancer genome. Now that TCGA and other programs have provided tremendous insight and identified genes known to be recurrent cancer drivers as well as important mutational signatures, focus has shifted to implementing genomic testing in the clinic. Therefore, increasing accessibility, improving applicability to routine tissue samples, and decreasing costs of clinical testing could be facilitated by targeted DNA and RNA sequencing methods incorporating these fundamental characteristics of the cancer genome.

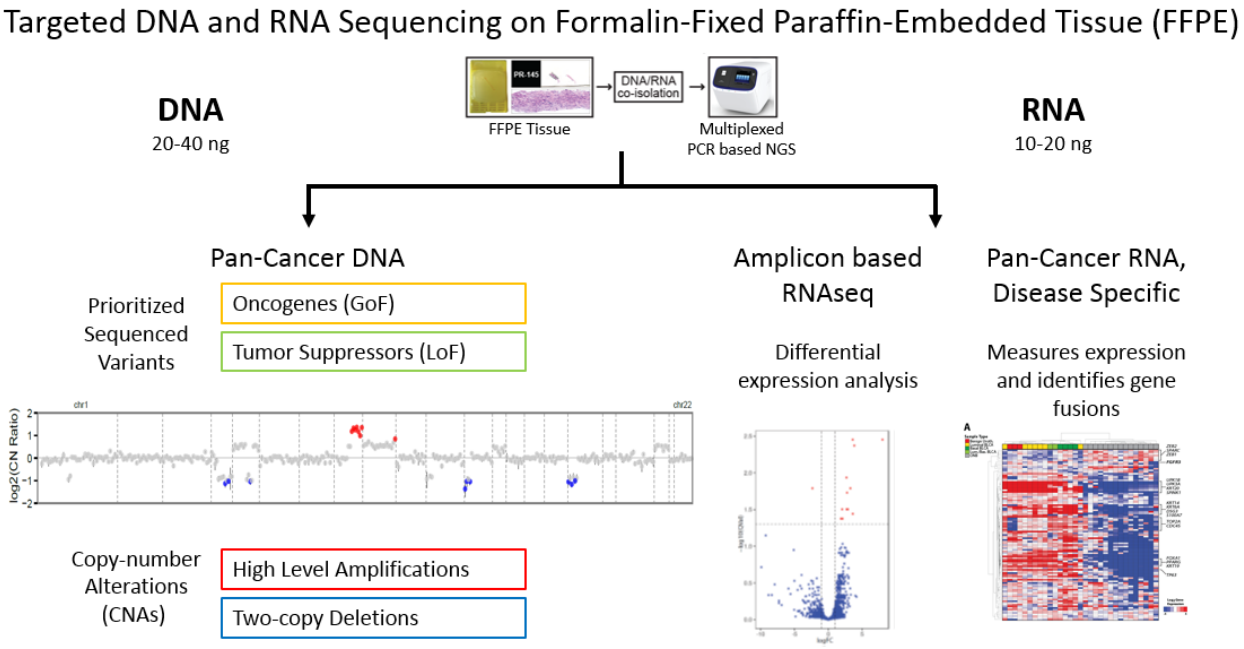

FIGURE 1.1 Summary of NGS Workflow.
Ion Torrent targeted next-generation sequencing allows for the use of very low amounts of DNA and RNA extracted from routine formalin-fixed paraffin embedded (FFPE) tissue for sequencing using a wide range of DNA and RNA panels. Figure adapted from [Hovelson *et al*. (99)].

1.3.2 Characterization of Rare Cancers Using NGS

Approximately 20% of cancer patients are diagnosed with a rare cancer, whereas the remaining cancer patients are primarily diagnosed with the more common prostate, breast, lung, and colon cancers (104). Since they are often more aggressive than common cancers, rare cancers account for about 25% of cancer related deaths (105). However, each type of rare cancer is defined

as those with fewer than 6 cases per 100,000 people per year (106). Although rare cancers have been defined by their unusual cell of origin or differentiation state or as a histologically defined subtype of a common cancer, due to the rarity of the cancer, precise and accurate diagnosis is often challenging. Clinically, this can lead to the need to see specialists, misdiagnosis, and many times delayed diagnosis. Altogether, rare cancers represent a field with a very high unmet need (107). Importantly, two of the first insights into general cancer principles were from studies of rare cancers. Percival Pott's study on scrotal cancers led to the discovery of the carcinogenic effects of tar (4), and studies of retinoblastoma discovered the *RB1* gene leading to Knudson's two-hit hypothesis (108,109).

The genomic landscape of common cancers is often biologically and clinically more complicated than rare cancers (71,110). Since rare cancers tend to have a simpler genomic landscape, these models allow us to focus on the oncogenic processes driving these cancers (111). For example, Shah *et al.* identified a novel recurrent somatic *FOXL2* p.C134W mutation in all four of the adult granulosa cell tumors sequenced (112). Although *FOXL2* was already known to be important for the development of ovaries, this study was the first time *FOXL2* mutant was linked to having a potential role in driving pathogenesis of adult granulosa cell tumors (112,113). Likewise, if rare tumors are shown to harbor a known targetable driver previously identified in more common cancers, this may support referral to a basket trial, if available, to treat the patient according to their tumor's genomic landscape (114). Not only would this nominate more patients eligible for clinical trials but also direct patients with limited therapeutic options to relevant clinical trials, as done in a recent study where 27 patients with rare gynecologic cancers were directed to clinical trials after genomic profiling (115). Profiling of rare cancers has also identified many different subtypes among the soft tissue sarcomas that harbor a different fusion such as cases that were found to harbor *BCOR-CCNB3* fusions, defining a novel sarcoma subtype (116). Altogether, these examples show that profiling rare cancer can lead to the discovery of new insights into cancer and are therefore important tumors to characterize. With research institutions, such as the University of Michigan, having an established expertise in rare cancers, such as adrenal cancer which has a poor prognosis, efforts are being made to genomically characterize and translate these findings to improve the care for patients with rare cancers.

Examples of rare cancers discussed in this thesis include *CIC-DUX4* sarcomas, Merkel cell carcinomas (MCCs), and olfactory neuroblastomas (ONBs), where we performed comprehensive

multiplexed PCR based NGS to define the genomic landscape in the context of important clinical questions.

1.3.3 Biomarkers and Clinical Applications of NGS

While traditional molecular testing approaches have involved IHC (117), FISH, and PCR (118), we are now capable of conducting comprehensive molecular profiling of the cancer genome very quickly using NGS. Through the years, TCGA has used NGS to provide us with enormous amounts of data on the genomic landscape of each type of cancer, demonstrating that each cancer can harbor its own unique set of alterations. By using multiplatform analysis, TCGA identified the somatic molecular aberrations driving cancer that have guided the development for new targeted therapies and diagnostic tests (119). However, they also show the importance of classifying cancer according to genomic profile. As seen in endometrial carcinoma, integrated genomic characterization by TCGA identified four different molecular subtypes with different treatment recommendations (102). Alike, TCGA identified molecular similarities between basal-like breast tumors and high-grade serous ovarian tumors indicating shared driving events and therefore similar therapeutic opportunities (71).

Genome-based medicine in oncology has revolutionized our approach to clinical care and is now guiding precision medicine, which the Precision Medicine Initiative describes as "an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person" (120). The role of precision medicine has been limited in the past due to the more traditional one-size-fits-all approach, in which patients are given treatment and prevention strategies based on the average person (121). However, clinicians are now recognizing that while this approach is beneficial for some patients, other patients may not benefit from standard treatment. Therefore, studies have provided new insights on how to use different types of biomarkers to provide patients the most appropriate therapy. For example, cancers with *EGFR* activating mutations and cancers with *ALK* gene translocations can be treated with either EGFR kinase inhibitors, such as erlotinib or gefitinib (122-124), or ALK inhibitors, such as crizotinib (125), respectively. In contrast, patients with lung cancers that lack these alterations do not benefit from either treatment. Instead, these treatments would be detrimental to the patient's health since the patient would only develop the toxic side-effects of the drug. Therefore, recent studies are applying multiplexed assays assessing these oncogenic drivers to guide patients to the most appropriate diagnosis and targeted therapy (126).

In prostate cancer, studies identified diagnostic markers such as the *TMPRSS2-ERG* gene fusions that offered novel therapeutic avenues for prostate cancer (50). Biomarkers can also be used to predict response and resistance to therapy. Tumor mutational burden, mismatch-repair (MMR)/microsatellite instability (MSI) status, and POLE status, can predict patient response to immunotherapy (66,67,127,128). *KRAS* mutations in colorectal cancer, as previously discussed, confers resistance to EGFR therapies (78,79).

NGS is increasingly becoming more accessible. However, routine panels for characterizing somatic aberrations in a patient's cancer has yet to be implemented as part of routine patient care except at a limited number of health care systems. Although targeted panels are significantly smaller compared to WES, they provide clinically relevant information and can identify variants with higher confidence due to greater coverage at each base, enabling detection of rare oncogenic variants. Hovelson, *et al*. showed that the OCP identified a range of 6% to 42% of samples in each cancer type profiled that harbored alterations in a gene that are potentially targetable that were not detected during routine molecular testing (99). In fact, Memorial Sloan Kettering expanded on this idea in 2017 when the FDA approved their targeted NGS panel. The Integrated Mutation Profiling of Actionable Cancer Therapies (MSK-IMPACT) provides a broadly applicable tool suitable for advancing precision oncology efforts that identifies patients eligible for basket trials through molecular profiling (129,130).

1.4 Rationale and Goals

Due to the increasing interest in cancer genomics and its clinical utility, this thesis focuses on discovering clinically useful biomarkers, as well as answering cancer relevant questions using targeted NGS. As previously discussed, targeted NGS has given us the ability to profile archival routine FFPE tissue with limited available DNA/RNA. Here, we profile the most common type of endometrial carcinoma (Chapter 2), ocular and cutaneous squamous cell carcinomas (Chapter 3), and a set of rare cancers including *CIC-DUX4* sarcomas (Chapter 4.1), Merkel Cell Carcinoma (Chapter 4.2), and Olfactory Neuroblastomas (Chapter 4.3).

Although major advances have led to a very comprehensive understanding of the cancer genome, most studies have primarily focused on late stage disease and therefore leave a gap in our understanding of cancer development and progression. Using the current available data, we can infer the sequence of genetic events in a tumor sample by distinguishing between clonal and

subclonal genomic aberrations. Clonal mutations could then be inferred to be an earlier event if present in all cancer cells. Since many driver genes appear to be clonal, there is a need to conduct comprehensive molecular profiling of precursor lesions to identify the sequence of genomic events occurring during cancer progression, including when these driving lesions occur and which lead to progression to invasive disease, as stated by multiple calls for a Pre-Cancer Genome Atlas (PCGA) (82,83,89). By identifying the genomic events occurring at earlier stages, we would be able to establish ways to predict the fate of each lesion, develop non-invasive early-detection strategies, and develop interventions at the precursor stage for each type of cancer.

Due to this interest, my first two chapters focus on two different cancers that are known to arise from well recognized precursor lesions, endometrioid endometrial carcinomas (EC) and squamous cell carcinomas (SCCs) from the conjunctiva and skin. In particular, our interest in ECs originated when two regions profiled from the same case shared clonal *PTEN* mutations but harbored discordant *CTNNB1* mutations **(Figure 1.2)** (100). Therefore, based on our observation of mutational heterogeneity in *CTNNB1* in ECs, we hypothesized that endometrial carcinomas may commonly have intratumoral heterogeneity in other known genetic driving alterations and predicted that this intratumoral heterogeneity will be observed throughout cancer progression. Our interest in SCCs originated when our lab profiled a few conjunctival intraepithelial neoplasia (precursor lesions) that showed very similar genomic characteristics as those seen in invasive SCCs. Despite SCCs being the most frequent human solid tumor at many anatomic sites, the driving molecular alterations underlying their development are poorly understood, especially in the context of photodamage. Therefore, our aim was to characterize progression of both well- (cutaneous) and poorly-studied (ocular) SCC through comprehensive molecular profiling.

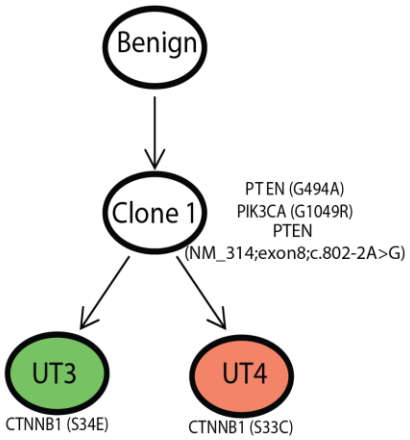

FIGURE 1.2 Endometrioid endometrial carcinoma clonality analysis.
Two endometrial carcinoma regions from one patient (UT3 and UT4) show clonal *PTEN* and *PIK3CA* mutations and heterogeneous *CTNNB1* mutations.

While my first two chapters involve the characterization of precursor lesions, Chapter 4 focuses on rare cancers, an area that also lacks thorough molecular profiling and characterization. In fact, rare cancers have a high unmet need primarily due to the late or incorrect diagnosis, lack of clinical expertise, limited number of clinical studies due to the small number of patients, few available registries and tissue banks, and challenges that come along with developing new therapies for a limited population (107,131). Due to the challenges in distinguishing between different cancer subtypes through pathological assessment, TCGA studies suggest that molecular profiling provides the most precise way to group cancer types. Hence, we apply this concept to *CIC-DUX4* sarcomas, which highly resemble Ewing sarcomas. Although very similar histologically, *CIC-DUX4* sarcomas have an aggressive course, develop rapid chemoresistance, and lack of highly efficacious therapeutic strategies (132). Since it is unclear whether *CIC-DUX4* sarcomas are similar to Ewing sarcomas at the genomic level, we aimed to identify genomic aberrations unique to *CIC-DUX4* sarcomas that can be used to guide patients toward more precise treatment options (133). In Merkel cell carcinomas (MCCs), our primary question was whether genomic profiling could distinguish the clinical scenario of patients who present with two tumors, which could represent either two primary tumors or a primary with a matched metastasis (134). Here, we aimed to use clonality studies to assess genetic relatedness since patients with two primaries versus patients with metastatic disease are treated very differently. Lastly, in our last effort to characterize rare cancers, we conduct molecular profiling of olfactory neuroblastoma to identify genetic aberrations responsible for its aggressive course as well as potential therapeutic targets (135,136).

CHAPTER 2: Multiclonality and Marked Branched Evolution of Low-Grade Endometrioid Endometrial Carcinoma

2.1 Introduction

Integrated genomic characterization of endometrial carcinoma (EC) by The Cancer Genome Atlas (TCGA) defined four groups based on histology, copy number alterations (CNAs) and mutations: *POLE* (ultra-mutated), microsatellite instability (MSI; hyper-mutated), CNA-high (serous-like) and CNA-low (endometrioid), consistent with clinical/pathological/molecular endometrial carcinoma classification as type I (usually low-grade, endometrioid [LGEC]) and type II (high-grade, non-endometrioid) (102,137). Endometrioid endometrial carcinomas are thought to develop through hyperplastic precursor lesions characterized by architectural and nuclear atypia. Although criteria differ, systems based on 1) nuclear atypia and glandular complexity (World Health Organization [WHO]) or 2) molecular genetics/morphology (endometrial intraepithelial neoplasia) are widely used (138). Lesions classified by the first as atypical hyperplasia (AH)—more specifically CAH when glandular complexity is present—and by the second as endometrial intraepithelial neoplasia (EIN) are now considered similar premalignant processes and the terms are used interchangeably in the latest WHO classification system (137,138). Endometrial hyperplasia without atypia, sometimes referred to as complex hyperplasia (CH) is thought to result from unopposed estrogen stimulation and has a lower risk of progression to LGEC than EIN/AH. LGEC and its precursors often display foci of squamous differentiation (SqD) a feature that is not typically seen in other types of endometrial carcinomas, such as serous or clear cell carcinomas. LGECs are usually CNA-low, non-hyper/ultra-mutated, lack *TP53* mutations, and frequently harbor somatic alterations affecting the PI3K, RTK/RAS, and Wnt signaling pathways (including recurrent mutations in *PTEN, PIK3R1, PIK3CA, KRAS,* and *CTNNB1)* (102).

As reflected in calls to generate a Pre-Cancer Genome Atlas (PCGA), the molecular progression of EIN/AH to endometrial carcinoma, like in many cancers, is incompletely

understood in part due to the technical challenges of profiling minute lesions/areas of interest often available only in routinely processed formalin-fixed paraffin embedded (FFPE) specimens (83). Driving *PTEN* mutations occur early in type I endometrial carcinomas since they are generally found to co-exist with other commonly mutated genes and are critical in defining EIN (139). However, whether EIN/AH usually progresses to endometrial carcinoma via linear vs. branched evolution is unresolved. Limited intratumoral heterogeneity with respect to integrative classification of endometrial carcinomas has been reported, including 96% concordance of *CTNNB1* mutational status (140), and a next-generation sequencing (NGS)- based study of three pairs of EIN/AH and CNA-low LGEC supported clonal origin in all cases (141). In contrast, substantial mutational heterogeneity, supporting branched evolution, was reported in a study of 6 cases of matched, but spatially distinct EIN/AH and CNA-low LGEC (141), as well as in a hotspot NGS-based study of endometrial carcinoma from paired uterine aspirates and multiple regions at hysterectomy (142).

Understanding intratumoral heterogeneity in LGEC development is critical for the development of prognostic biomarkers. While most patients with LGEC are cured by surgery alone, those that recur do poorly, arguing for the identification of prognostic biomarkers. Recently, Liu *et al*. and Kurnit *et al*. both reported that *CTNNB1* mutations were prognostic for shorter recurrence-free survival in patients with low-stage LGEC (143,144). We were intrigued by this finding, as we had previously observed different *CTNNB1* mutations in paired primary uterine EC (p.S45P) and tubal metastasis (p.S45F) components of a clinically type I high-grade endometrioid carcinoma that had a shared *PTEN* (p.R130X) mutation in both components (100). Likewise, we recently observed discordant *CTNNB1* mutations in the different components of a uterine EC that had areas of conventional histology, as well as, areas with variant histology referred to as "corded and hyalinized" EC (CHEC) (p.G34E and p.S33C, C.S. Carter *et al*., manuscript in preparation). Hence, to comprehensively assess intratumoral heterogeneity in LGEC development, we performed multi-region profiling of matched spatially defined EIN/AH and LGEC components from routinely processed FFPE tissue specimens using a highly scalable, comprehensive multiplexed PCR based NGS approach.

2.2 Materials and Methods

2.2.1 Cohort

With IRB approval, we retrospectively identified patients with LGEC (FIGO grade I/II) at hysterectomy using a previously described electronic medical record search engine (EMERSE) (145). We collected 14 cases with available archived FFPE tissues that had spatially and histologically distinct foci of both precursor (EIN/AH) or EC. For each case, regions of interest were identified on hematoxylin and eosin (H&E) stained slides and classified according to the WHO histologic system by board certified Pathologists (A.P.S. and S.A.T.) as: complex hyperplasia (CH), CAH, frankly invasive EC, or frankly invasive EC with squamous differentiation (ECsq). Regions were punched (1-3 punches) from the FFPE block using 21 gauge dispensing tips (0.510 mm inner diameter) followed by examination of an H&E recut to confirm localization. DNA and RNA from each punch were co-isolated using the Qiagen AllPrep FFPE DNA/RNA kit (Qiagen) and quantified using the Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA) as described (99).

2.2.2 DNA Next Generation Sequencing

We performed targeted, multiplexed PCR-based DNA next generation sequencing (NGS) essentially as described (99) using panels targeting >130 cancer related genes, including those recurrently mutated in LGEC (102). We used 20-24 ng of DNA per sample for library construction using the Ion Ampliseq library kit 2.0 (Life Technologies, Carlsbad, CA) with barcode incorporation and sequencing on the Ion Torrent Proton sequencer as described (99). Data analysis was performed essentially as described to identify high-confidence, prioritized somatic mutations and copy number alterations using validated pipelines based on Torrent Suite 5.0.4.0 (99,100,146). All high-confidence somatic variants were visualized in IGV (Integrative Genomics Viewer), with selected validation by Sanger sequencing**.** *POLE* hotspot mutation status was assessed by Sanger sequencing as they are not targeted by our panels. High-confidence somatic variants occurring at hotspots (>3 observations at that residue in COSMIC) in oncogenes, in-frame indels in oncogenes or tumor suppressor genes, or hotspot or deleterious (nonsense/frameshift/splice site altering variants) in tumor suppressor genes were considered driving variants (99,100). Case identity was confirmed in all populations by assessment of rare high-confidence SNPs.

2.2.3 Sanger Sequencing for *CTNNB1*, *POLE*, and to Validate Called Somatic Variants

Bidirectional Sanger sequencing was conducted with 10 ng of genomic DNA as template in PCR amplifications with Invitrogen Platinum PCR SuperMix Hi-Fi (Life-Technologies, Carlsbad, CA) with the suggested initial denaturation and cycling conditions. PCR products were then subjected to bidirectional Sanger sequencing for both primer pairs by the University of Michigan DNA Sequencing Core after treatment with ExoSAP-IT (GE Healthcare) and sequences were analyzed using SeqMan Pro Software (DNASTAR). Sanger sequencing was conducted for *POLE* mutations using previously published *POLE* primers (147) to sequence at least one representative sample from every case, to validate *CTNNB1* mutations, as well as other selected somatic mutations using custom designed primers made with the IDT PrimerQuest tool**.**

2.2.4 Phylogenetic Analysis

We conducted evolutionary analysis using PHYLIP v 3.695. For each tumor sample, the mutation status of non-synonymous somatic mutations was considered a character (absent or present) for this analysis. Evolutionary trees were constructed using Dollop (Dollo and polymorphism parsimony methods) using polymorphism parsimony with default parameters.

2.2.5 Amplicon-Based Whole Transcriptome Sequencing

We performed amplicon-based whole-transcriptome sequencing using the Ion Ampliseq Transcriptome Human Gene Expression Kit (Life Technologies, Carlsbad, CA) according to manufacturer's instructions with 17.5 ng of RNA per sample, allowing for interrogation of ~21,000 RNA transcripts. Library preparation was performed according to the manufacturer's instructions and as described above for DNA sequencing. Technical replicate libraries and templates were independently constructed and sequenced on separate chips. Reads were mapped and quantified using version 5.0.4.0 of TorrentSuite's (Life Science Technologies) coverageAnalysis plugin with the uniquely mapped reads option and default parameters. End-to-end reads were used for differential gene expression analysis using the R package edgeR (148,149). Principal component analysis was used on a log-normalized count table after library size normalization to detect batch effects. DEG analysis was performed using the R package edgeR (148,149). Filters used on these data prior to DEG analysis included one for non-expressed genes, and one for genes with inconsistent expression between technical replicates. The first filter required a gene to be expressed

in at least two biological samples. The second filter required at least 80% of technical replicates between samples to be concordant in either being above or below 5 reads per million (rpm). A threshold of 5 rpm was estimated to be real gene expression from plotting reads per gene between samples. The filters were applied after combining the technical replicates to represent a single sample. Library normalization using a trimmed mean of m-values method, and DEG analyses were performed using edgeR's generalized linear models. The linear models used to fit the contrasts for precursor versus cancer and *CTNNB1* wild-type versus mutant did not have an intercept term and followed the model "~0 + factor." The adenocarcinoma versus adenocarcinoma with squamous contrast controlled for *CTNNB1* status using *CTNNB1* as a blocking factor with the following model "~0 + factor + blocking factor." Volcano plots were made using the R-package ggplot and multiplicity was corrected by calculating a Q-value using Benjamini and Hochberg's false discovery rate (150).

2.2.6 Transcriptome Analysis

Functional analysis of differentially expressed transcripts between *CTNNB1* mutant vs. wild-type populations was performed using Database for Annotation Visualization and Integrated Discovery (DAVID) (151). Likewise, Gene Set Enrichment Analysis, developed by the Broad Institute (Cambridge, MA), was also done on differentially expressed genes from indicated comparisons using version 3.0 of GSEA (152,153) . The enrichment was done using a pre-ranked list with the ranking metric being the corrected p-value divided by the sign of the fold-change. The expression data was tested against the hallmark gene set to evaluate enrichment with "Wnt-B-catenin pathway (HALLMARK_WNT_BETA_CATENIN_SIGNALING: M5895)".

To determine if genes identified as differentially expressed between EC with and without squamous differentiation were consistent with known squamous transcripts, we identified the top 50 genes most over-expressed in TCGA lung squamous cell carcinoma vs. lung adenocarcinoma using batch corrected, uniformly processed TCGA data generated by Wang *et al*. (ranked by fold change, with two-sided student's t-test p-value < 0.001 and FPKM > 0 in at least 300 of the 489 lung squamous cells) (154). Significance in the overlap with the 18 genes identified by differential expression analysis as significantly over-expressed in EC with vs. without squamous differentiation was assessed by a two-sided fisher's exact test using R (from 10,882 genes expressed in our LGEC cohort).

2.2.7 Immunohistochemistry (IHC)

Polyclonal rabbit anti-amylase (AMY1A) primary antibodies (HPA045399 [562], 1:800; HPA045394 [560], 1:2000) were selected based on confirmation of expression in expected tissues (pancreas and salivary glands) in the Human Protein Tissue Atlas (155). IHC staining was performed on 4-5 um unstained FFPE slides using an automated protocol on the Ventana Benchmark XT System using ultraView Universal DAB Detection Kit (Cat no. 760-500, Ventana Medical Systems). Staining was optimized and confirmed to show expected staining in pancreas and salivary gland tissues prior to staining EC samples.

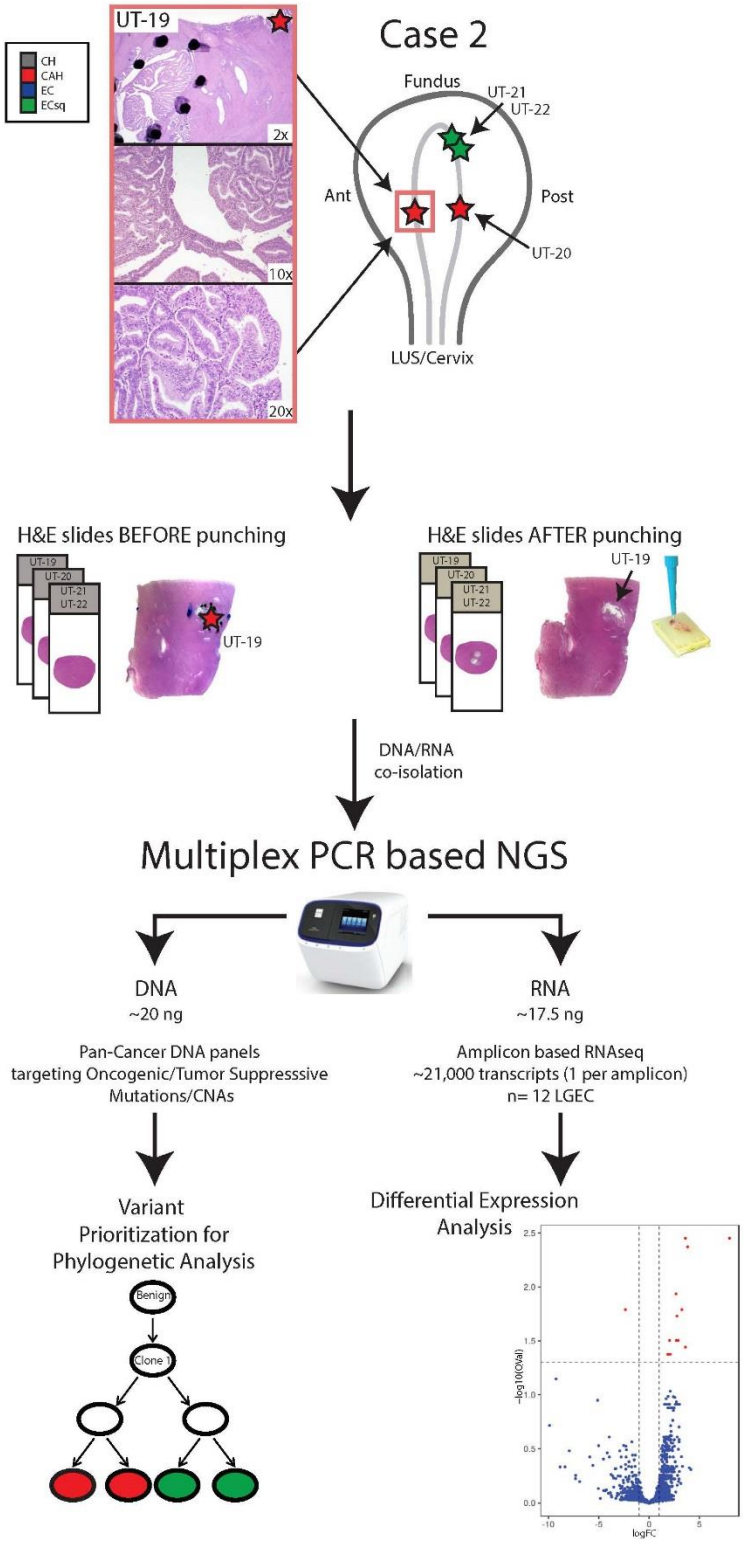

FIGURE 2.1 Comprehensive DNA and RNA profiling of low grade endometrioid endometrial carcinoma (LGEC) development from routine clinical specimens. Schematic of spatially defined uterine cell populations from a representative case (Case 2) is shown, with population type and associated histologic type indicated by the color scale (CH=complex hyperplasia, CAH=complex atypical hyperplasia, EC=frankly invasive LGEC; sq=squamous metaplasia). Histology for the one population (UT-19, CAH) is shown with original magnification indicated. Precise tissue punching was used for isolation from routine FFPE blocks and subsequent hematoxylin and eosin (H&E) stained slides were used to confirm isolation of expected populations. Multiplexed PCR based DNA and RNA sequencing was performed on ≤20ng co-isolated nucleic acids to comprehensively characterize LGEC development and intratumoral heterogeneity through driver gene alteration assessment and whole transcriptome profiling.

2.3 Results

2.3.1 Comprehensive Genomic Profiling of LGEC Development

To assess the molecular landscape of LGEC development, we performed comprehensive DNA and RNA based NGS of 14 cases of FIGO grade 1 or 2 EC with spatially defined minute precursors and/or EC components using a highly scalable approach optimized for routine FFPE material (**Fig 2.1**). We obtained between 130-1,850 ng of extracted DNA (mean 987 ng), consistent with tens of thousands of cells from the punched regions. To identify oncogenic and tumor suppressive somatic mutations and CNAs, we performed multiplexed PCR-based DNA NGS (mxDNAseq) on 70 spatially-defined, minute (~1-2mm^2 surface area) cell populations using panels targeting ≥130 genes, including essentially all recurrently altered genes in LGEC using extensively validated approaches. As described below, to validate the impact of observed histologic and somatic mutational heterogeneity, we also performed multiplexed PCR based transcriptome NGS (mxRNAseq) on co-isolated RNA from 12 cell populations.

Out of the 14 cases, one was classified as ultra-mutated. In the remaining 13 cases, no high level, focal somatic CNAs were identified in any cell populations; hence, these cases were considered CNA-low LGEC. After exclusion of populations failing QC metrics, our cohort represented the full spectrum of LGEC development, including 2, 23, 27, and 12 populations (*n*=64 total) classified as CH, CAH, invasive EC, or invasive ECsq, respectively.

Across the 13 CNA-low LGEC cases, all cell populations harbored at least one clear driving somatic mutation in *PTEN*, *PIK3R1*, or *PIK3CA* **(Fig 2.2)** consistent with the nearly universal deregulation of this pathway as a driving event in LGEC. Ten of 13 cases had at least one driving *PTEN* mutation detected in all cell populations (4 and 3 cases also had *PIK3CA* and *PIK3R1* mutations, respectively, in all cell populations) consistent with prior single gene and TCGA studies (102,156-158). In the remaining three cases, no cell populations harbored *PTEN* mutations, but two invasive EC/ECsq cases harbored somatic driving clonal *PIK3R1* mutations (Cases 6 and 9), and one CAH case had somatic driving clonal *PIK3CA* mutations (Case 7).

We also identified recurrent, driving mutations across our LGEC cases in *CTNNB1*, *FBXW7*, *KRAS* and *FGFR2*, consistent with bulk sequencing of LGEC (102). Importantly, across the 64 populations, we identified an average of 4 (range 2-5) driving somatic mutations in the above described seven genes, making LGEC an ideal system to assess clonality and intratumoral

heterogeneity using a very limited subset of the genome. Of note, no significant difference in the number of prioritized mutations was observed between CH/CAH and EC/ECsq populations (average 3.0 vs. 3.3, two-tailed unpaired t-test, p=0.18).

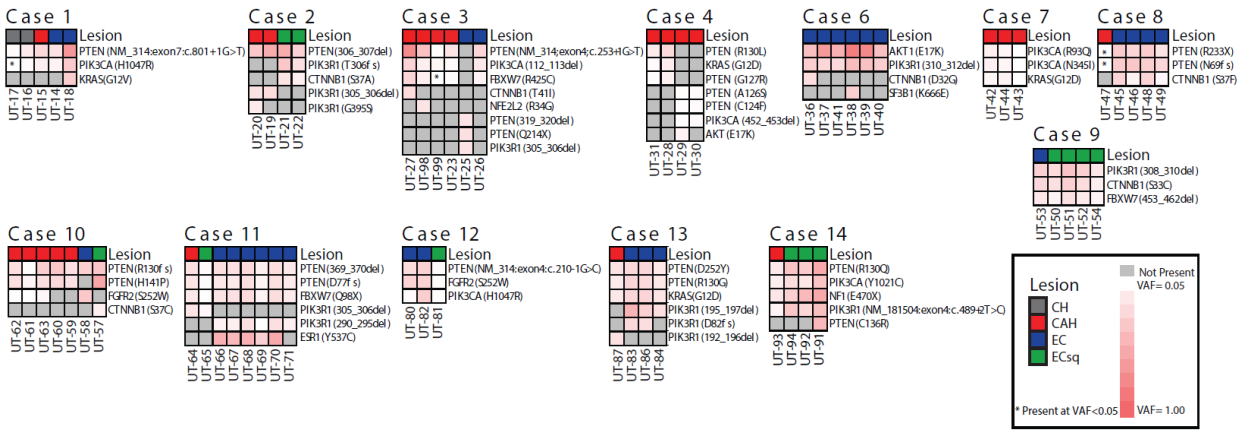

FIGURE 2.2 Somatic mutations across LGEC development.
Heatmaps showing all prioritized somatic mutations identified in endometrial cell populations, per LGEC case, with lesion histology indicated in the top row (according to the color scale at the bottom right). Individual somatic mutations are shown in rows, with the variant allele frequency (VAF) indicated by the color hue gradient at the bottom right (gray=not present, * = well supported reads on manual review and considered present but VAF<5%).

2.3.2 Multiclonality in LGEC Development

While 11/13 LGEC cases were clonal based on shared *PTEN*, *PIK3R1*, or *PIK3CA* mutations across all cell populations, two cases (Cases 3 & 4) showed clear multi-clonality in spatially distinct cell populations. In Case 3 (**Fig 2.3A**), 5 of 6 profiled cell populations (four CAH and one invasive EC; from anterior and posterior aspects of the uterus) shared driving *PTEN*, *PIK3CA* and *FBXW7* mutations; however, a population of invasive EC (UT-25) lacked these alterations but harbored two distinct driving *PTEN* mutations and a *PIK3R1* mutation. In Case 4 (**Fig 2.3B**), we profiled 4 regions of CAH from the uterine anterior, posterior and fundus. Of note, while the CAH populations from the anterior and posterior aspect (UT-28 and UT-31) harbored the same driving *PTEN* and *KRAS* mutations, the two CAH populations from the fundus (UT-29 and UT30) harbored distinct driving *PTEN* and *PIK3CA* mutations. Taken together, even with sampling of only an average of 5 spatially distinct cell populations per case, our results demonstrate that true multiclonality is relatively frequent during LGEC development.

2.3.3 Marked Intratumoral Heterogeneity in Presumed LGEC Driving Mutations

Beyond multiclonality, we observed marked heterogeneity in presumed driving somatic mutations across LGEC and precursor populations in 10 of 13 cases, particularly those with both precursor and invasive EC populations. For example, in Case 13 (**Fig 2.3C**), we profiled separate populations of CAH (UT-87) and EC (UT-83, 86, 87) from the posterior aspect. All populations shared driving *PTEN* (two mutations) and *KRAS* mutations. While all EC populations in this case also shared *PIK3R1* mutations, the CAH population harbored a distinct *PIK3R1* mutation. Likewise, a separate *PIK3R1* mutation was only present in two of the three EC populations.

Similarly, in Case 2 (**Fig 2.3D**), we profiled separate populations of CAH (UT-19, anterior; UT-20, posterior) and ECsq (UT-21 and UT-22; fundus), all of which shared a driving *PTEN* mutation. Both ECsq populations also shared driving *CTNNB1* and *PIK3R1* mutations, neither of which were present in the CAH populations. However, both CAH populations shared a different driving *PIK3R1* mutation, while one CAH population harbored an additional missense *PIK3R1* mutation of unclear pathogenicity not present in the other CAH or ECsq populations. Case 11 showed similar intratumoral heterogeneity and branched evolution between precursor and invasive EC populations.

2.3.4 Intratumoral Heterogeneity in Candidate Prognostic *CTNNB1* Mutations

As described above, a motivator of this study was our previous observation of discordance in driving *CTNNB1* mutational status in two different EC cases. Seven of 13 cases in our cohort showed no *CTNNB1* mutations in any cell population (clonally absent). In the remaining 6 cases where at least one population harbored a *CTNNB1* mutation, only one showed clonal *CTNNB1* mutations in all profiled populations (Case 9, with EC and ECsq populations). In Case 8, a shared *CTNNB1* mutation in all EC populations was not present in the CAH sample (however low tumor content in this sample precluded definitive exclusion). The remaining four cases showed: 1) a private (present in only one population) *CTNNB1* mutation in one precursor population but not in other precursor or EC populations (Case 3), 2) shared *CTNNB1* mutations in all EC/ECsq populations but not in precursor populations (Case 2), 3) private *CTNNB1* mutation in only one of 6 EC populations (Case 6), and 4) private *CTNNB1* mutation in one ECsq population but not in the EC or multiple precursor populations (Case 10) **(Fig 2.2)**. All mutations were observed at essentially clonal VAF and mutational presence/absence was confirmed by Sanger sequencing. Taken together, these results demonstrate the existence of intratumoral heterogeneity in potentially

prognostic *CTNNB1* mutations both 1) within precursor and EC populations and 2) within EC populations in a given case.

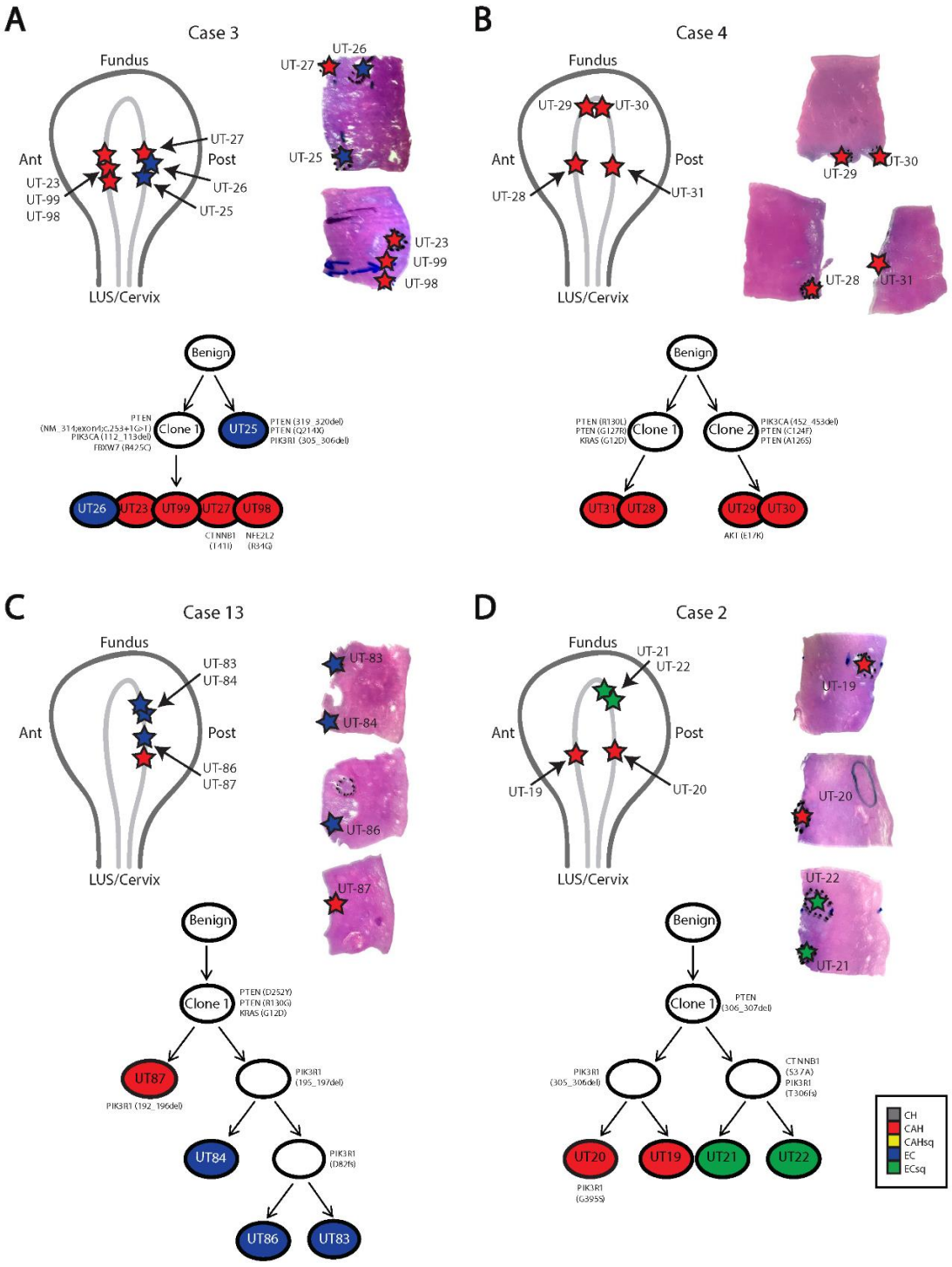

FIGURE 2.3 Multifocality and marked intratumoral heterogeneity in LGEC development.
For selected cases, location of isolated cell populations (indicated by stars) in the uterus are indicated on anatomic diagrams and corresponding H&E slides. Phylogenetic trees for these cases are shown, with shared mutations for each clone indicated. **A-B**. Cases showing multiclonal LGEC development. **C-D**. Cases showing marked intratumoral heterogeneity in LGEC precursor and invasive cell populations.

2.3.5 Clonal Sweep of Heterogeneous Mutations is Common in LGEC

Heterogeneous mutations (those present in not all cell populations from a given case) may represent 1) subclonal alterations present but variably detected in all populations due to sampling or 2) clonal alterations present and selected for in the population. Through assessment of the variant allele frequency (VAF; # variant containing reads / total # reads) of truncal PI3K pathway driving mutations which inform on the estimated tumor content (VAF ~ ½ and ~ equivalent to the tumor content for heterozygous and homozygous variants, respectively), essentially all of the homogeneous and heterogeneous driving mutations observed in our cohort were present in all cells in the population (clonal cancer cell fraction [CCF]). These results are consistent with selection of the variants due to increased fitness and "sweep" through the tumor cell population(159). Of note, the only gene that frequently showed less than clonal CCF was *CTNNB1*, further complicating its potential use as a prognostic and/or predictive biomarker. Importantly, however, the presence of numerous heterogeneous or private driving mutations in both precursor and EC populations, most at clonal CCF, indicates a fitness advantage to these mutations regardless of histologic appearance or spatial proximity. Phylogenetic analysis thus supports extensive branched evolution in both precursor and EC populations, consistent with branched evolution in LGEC development, in agreement with mechanisms in most other profiled cancers (160,161).

2.3.6 Confirmation of Intratumoral Heterogeneity in *CTNNB1* Mutation Driven Pathway Activation through Transcriptome Sequencing

We next sought to further confirm the relevance of the often heterogeneous *CTNNB1* mutations by looking for transcriptional evidence of Wnt pathway activity in populations with and without *CTNNB1* mutations. Given the challenges of performing conventional or capture based whole transcriptome RNA-seq with minute quantities of FFPE isolated RNA (162), we attempted mxRNAseq using ≤20ng of co-isolated RNA from 12 cell populations. Samples were selected to represent the spectrum of precursor vs. EC lesions with and without *CTNNB1* mutations. Across the 12 sequenced populations (with technical replicates in different batches), we generated an average of 8,891,762 end-to-end reads with the 10,882 transcripts across the cohort showing >5 RPM used for further analysis. Technical replicates showed highly correlated normalized expression (median per pair Pearson r = 0.9, range 0.92-0.99) with principal component analysis showing expected clustering of technical replicates (**Fig 2.4A**). Differential expression analysis

between profiled EC (*n*=6) and ECsq (*n*=2) populations identified 18 transcripts over-expressed in ECsq vs. EC; Comparison to transcripts over-expressed in TCGA lung squamous cell carcinoma vs. adenocarcinoma (154) confirmed significant enrichment (Odds ratio 66.9, two-sided Fisher's exact test, *p*=1.15E), including squamous epithelial specific transcripts *PRR9*, *KRT31* and *CALM3*, further supporting the validity of our approach. Likewise, we observed marked over-expression of *AMY1A* in profiled CAH (*n*=4) vs. EC/ECsq (*n*=8) populations and confirmed AMY1A protein over-expression in these samples by IHC (**Fig 2.4B**).

We thus assessed differentially expressed transcripts in profiled *CTNNB1* mutant (*n=6*) vs. wild-type (*n=6*) precursor and EC populations. Importantly, we identified 21 transcripts significantly over-expressed in *CTNNB1* mutant samples, including the known canonical Wnt target genes *NOTUM(163), CXCL14(164), GAD1(165), DKK4(166),* and *NKD1(167)* **(Fig 2.4C)**. Database for Annotation Visualization and Integrated Discovery (DAVID) functional annotation assessment(151) also identified "Wnt signaling pathway" as the most significantly enriched biological process in the over-expressed *CTNNB1* mutant gene set (p = 5.2×10^{-7}, Benjamini corrected q value = 4.2×10^{-5}). Gene set enrichment analysis (GSEA) of the hallmark gene sets also confirmed enrichment of the Wnt-β catenin signaling pathway (FDR q-value= 0.032 [NES: 2.08]). Taken together, these results support the applicability of transcriptome wide mxRNAseq to minute FFPE isolated cell populations and confirm the functional relevance of shared and private *CTNNB1* mutations in both precursor and EC populations.

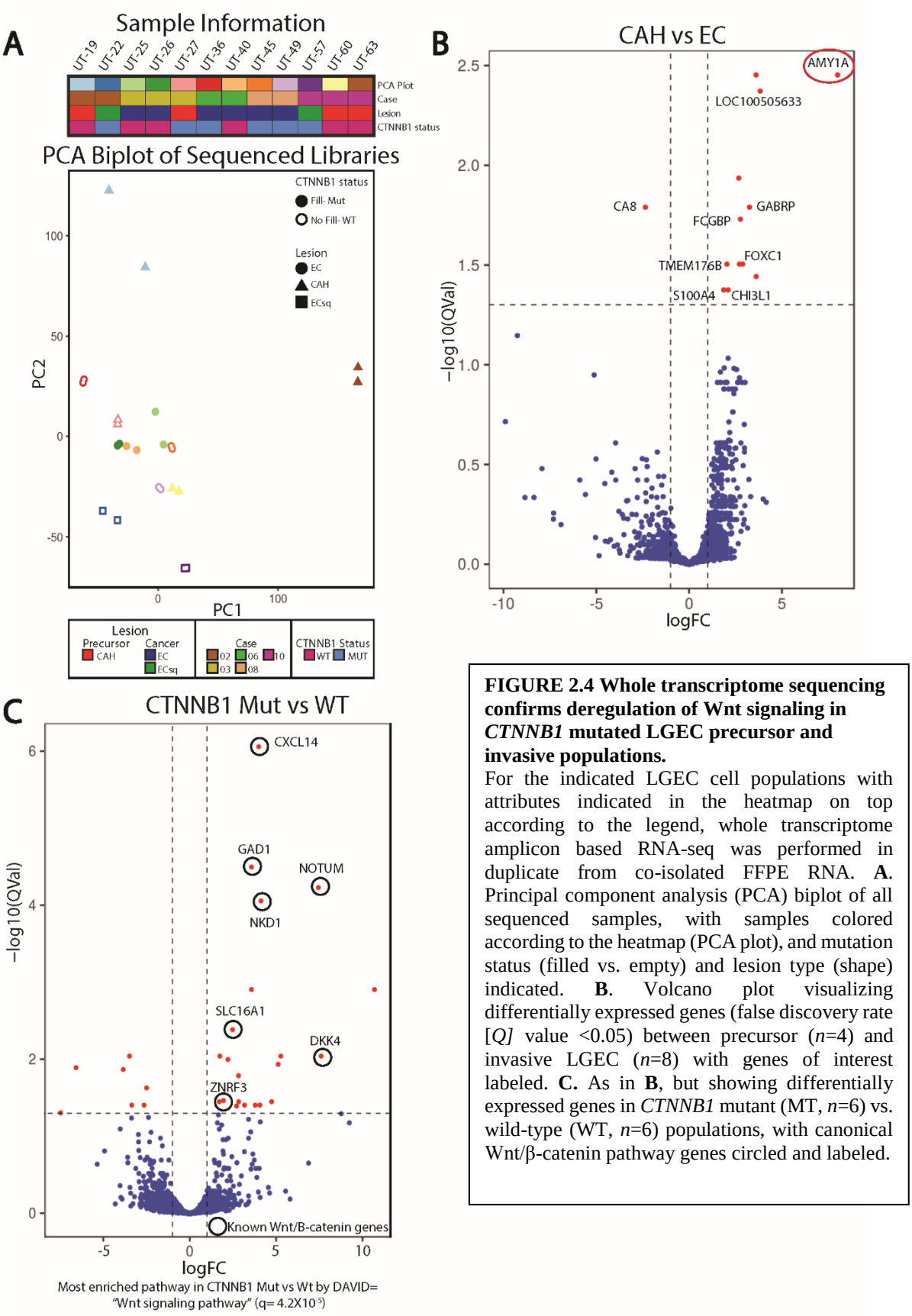

FIGURE 2.4 Whole transcriptome sequencing confirms deregulation of Wnt signaling in *CTNNB1* mutated LGEC precursor and invasive populations.
For the indicated LGEC cell populations with attributes indicated in the heatmap on top according to the legend, whole transcriptome amplicon based RNA-seq was performed in duplicate from co-isolated FFPE RNA. **A**. Principal component analysis (PCA) biplot of all sequenced samples, with samples colored according to the heatmap (PCA plot), and mutation status (filled vs. empty) and lesion type (shape) indicated. **B**. Volcano plot visualizing differentially expressed genes (false discovery rate [*Q]* value <0.05) between precursor (*n*=4) and invasive LGEC (*n*=8) with genes of interest labeled. **C.** As in **B**, but showing differentially expressed genes in *CTNNB1* mutant (MT, *n*=6) vs. wild-type (WT, *n*=6) populations, with canonical Wnt/β-catenin pathway genes circled and labeled.

2.4 Discussion

Here, to better understand the development of LGEC, we performed multi-region, comprehensive somatic molecular profiling of minute cell populations ranging from presumed precursor lesions through invasive LGEC. Through this high depth (average >1,000x coverage) approach on spatially defined populations with variable histology from 13 cases, we identified marked intratumoral mutational heterogeneity in presumed cancer driving genes in the vast majority of cases. Our work builds on two small series of LGEC precursors and invasive components, which support substantial intratumoral heterogeneity and branched evolution in LGEC development (141,142); however, our study is the first to definitively demonstrate multiclonality in both spatially defined precursor and invasive populations. Our findings have important implications for understanding LGEC development, as well as efforts to identify prognostic and predictive biomarkers, such as *CTNNB1*.

Consistent with the known role of PI3K pathway deregulation in EIN/LGEC development, all profiled cell populations harbored clear driving *PTEN*, *PIK3CA* or *PIK3R1* mutations. In three cases, all populations harbored only *PIK3CA* or *PIK3R1* mutations, demonstrating that LGEC development does not absolutely require a *PTEN* mutation, consistent with TCGA data (102). Likewise, in 7 of 10 cases with driving clonal *PTEN* mutations in all populations, we only observed *PTEN* point mutations, in-frame short deletions, or splice site mutations, consistent with the lack of sensitivity of *PTEN* immunohistochemistry for EIN identification in pathological practice (138,168). Importantly, in several cases, we observed continued selection for, and convergent evolution in, driving mutations in the above three PI3K pathway members and *AKT1* in histologically presumed precursor lesions.

Through analysis of driving PI3K pathway mutations, we identified a case that developed two clonally distinct, multifocal LGECs (Case 3) and a case with clonally distinct precursor populations (Case 4). To our knowledge, such multiclonality has not been previously described in spatially defined populations. Remarkably, in Case 3, the two clonally distinct EC populations were on the same FFPE block (< 2 cm away), with no clear morphologic distinction between the invasive EC populations **(Fig 2.3)**. In Case 4, where we could only sample superficial CAH appearing populations, distinct clones were present in the uterine fundus vs. anterior/posterior aspects. Given the relatively limited sampling performed in our study, we expect the observed rate

of 15% multiclonality to be an underestimate, with more women developing multiple transformed LGEC populations.

As described above, the two cases showing discordant *CTNNB1* mutations in paired endometrial carcinoma samples ((100) and C. Carter *et al*., manuscript in preparation) partly motivated this study. Importantly, Kurnit and Liu *et al*. recently described *CTNNB1* mutations as prognostic in low-stage LGEC (143,144). Similarly, a Phase II clinical trial of everolimus and letrozole (NCT01068249) in women with endometrial carcinoma found particularly high response rates in those with endometrioid histology and *CTNNB1* mutations (169). Remarkably, in our LGEC cohort profiled herein, we observed clonal *CTNNB1* in only one of five cases (where at least one population harbored a *CTNNB1* mutation and tumor content was sufficiently high in all samples to enable confident assessment). In the remaining cases, we saw diverse intratumoral heterogeneity, with *CTNNB1* mutations being observed privately in precursor and not EC populations (Case 3), shared in EC but not precursor populations (Case 2), and private in EC populations (Cases 6 and 10). Mutations in *CTNNB1* were also frequently subclonal in a given cell population, in contrast to essentially all other homogeneous or heterogeneous mutations in LGEC driver genes observed herein. Taken together, given that trials assessing *CTNNB1*—as well as PI3K members—as correlative biomarkers in women with EC are ongoing (e.g. NCT02228681), our results suggest that sampling and assessment strategies have the potential to substantially impact results and should therefore be carefully considered during future trial design.

In addition to comprehensive DNA-based profiling, we also conducted amplicon-based whole transcriptome sequencing on a subset of samples both to validate the approach, as well as determine whether subclonal *CTNNB1* mutations show evidence of transcriptional activation. Importantly, to our knowledge, this amplicon based whole transcriptome sequencing, which has the advantage of requiring <20ng RNA, has only been reported in a single study of FFPE tissue (170). In addition to high pairwise concordance in technical replicates supporting the validity of our transcriptome data, we also confirmed expected differential transcript expression in ECsq vs EC populations (over-expression of squamous epithelial transcripts) and *CTNNB1* mutant vs. wild-type populations (Wnt/β-catenin target genes). In an exploratory analysis of precursor vs. invasive EC populations, we identified amylase (*AMY1A*) as markedly over-expressed in precursor populations and confirmed these findings in the same samples by immunohistochemistry using two anti-AMY1A antibodies. By IHC, amylase has been reported as only occasionally expressed

in both benign secretory phase endometrial glands and well-differentiated endometrial adenocarcinomas (171-173), and hence we hypothesize that differential expression of AMY1A in paired precursor vs. invasive populations likely reflects differentiation (AMY1A expression was not diffusely present in individual precursor appearing glands across individual sections or cases) rather than a driving event in invasive EC development. Importantly, through numerous lines of validation, our results demonstrate the applicability of amplicon-based whole transcriptome sequencing to minute cell populations isolated from routine FFPE specimens, which may be particularly useful in scalable profiling of precursor lesions.

One of the major limitations of our study, which confounds most efforts to understand cancer development through precursors, is the use of concurrent presumed precursor and invasive populations to understand molecular features and drivers of invasive disease development. However, a more informative study design, where precursor populations with or without subsequent development of invasive disease are compared is confounded by the lack of clinical scenarios where precursor lesions are followed rather than completely excised. Our results herein combined with other studies highlight additional confounders including the potential of invasive disease to histologically mimic *in situ* precursors (174), the presence of multiclonal precursor and/or invasive clones, and the continued selection for driving mutations in precursor populations. Lastly, until studies profiling cases with confirmed non-progressing precursor lesions are profiled, it is unclear whether such intratumoral heterogeneity in LGEC precursors and/or invasive populations is ubiquitous, or a more specific feature of "aggressive" behavior that may be exploited for diagnosis (as has been shown feasible in uterine aspirates (142)) or prognosis (175). Supporting this concept, co-occurring *PTEN* and *PIK3CA* mutations have been reported to be extremely rare in CAH vs. invasive EC (176). However, consistent with another study assessing *PIK3CA* and *PTEN* mutation frequency in CAH from cases with co-occurring EC (177), we found co-occurrence of *PTEN* and *PIK3CA*/*PIK3R1* in at least one cell population with CAH histology in 6/8 cases (with *PTEN* mutations) with a co-occurring LGEC.

In summary, through comprehensive DNA and RNA profiling of minute, spatially defined populations from routine FFPE specimens, we demonstrate marked intratumoral heterogeneity and branched evolution in LGEC and precursors. Importantly, given this heterogeneity, sampling and sequencing depth may profoundly impact the detection of biomarkers in LGEC, including those such as *CTNNB1* that have been identified as prognostic in previous studies. Likewise, biomarker-

based studies (such as those targeting PI3K pathway members) may also need to account for this heterogeneity. Additionally, we also show relatively frequent true multiclonality, both of which have important implications for understanding LGEC development, predicting the behavior of LGEC precursors, and precision medicine for advanced LGEC. More generally, our approaches are applicable to archived FFPE samples and thus highly scalable, which may enable widespread sample contribution to efforts such as the PCGA (83), with the potential to transform the understanding of cancer precursors and early stage disease.

2.5 Future Directions

Different types of endometrial neoplasms resembling each other often affect histological analysis. Likewise, the presence of endometritis or the differences arising during the cycling of normal endometrium can make diagnosing endometrial cancer very challenging (178). Recent efforts have developed methods to detect early stages of endometrial and ovarian cancer through genomic testing of DNA in fluids obtained during a pap test. Wang *et al*. reported their ability to detect endometrial and ovarian cancer through minimally invasive sampling of the uterine cavity using a test called PapSeek. Here, they used an assay to detect aneuploidy and mutations in the 18 most commonly mutated genes and were able to detect early stage disease. In fact, their detection of endometrial cancer ranged between 81% and 93%, while their detection of ovarian cancer ranged between 33%-45%, with only up to 1.4% of false positives (179). While their approach is promising and is an example on how minimally invasive methods can be developed to detect disease, it still has limitations relating to the characterization of precursor lesions we conducted in this thesis. More importantly, since this was a retrospective study conducted on patients diagnosed with ovarian or endometrial cancer, changes need to be made as it becomes a diagnostic approach. Since we show that precursor lesions harbor the full complement of somatic alterations in invasive cancer, cells positive for mutations may not necessarily define invasive cancer. Therefore, the incorporation of transcriptomic characterization may allow for better distinction between precursor or cancer cells. To further explore transcriptomic differences, we have to address our underpowered transcriptome analysis by addition of samples to each classification to make our analysis more robust.

Furthermore, we propose assessing whether multiclonality and intra-tumor heterogeneity can predict clinical aggression. As seen in patients with prostate cancer, patients with high-risk

polyclonal tumors, as determined through phylogenetic analysis, frequently relapse (175). Since this could also be the case for LGECs, we propose increasing the number of regions sampled for each case to adequately distinguish between monoclonal ECs and ECs exhibiting multiclonality and/or intra-tumor heterogeneity. As done by Espiritu *et al.*, we propose using WGS to build comprehensive phylogenetic trees using high coverage of somatic SNVs to distinguish between polyclonal and monoclonal tumors. Altogether these analysis would also allow us to assess whether clinical outcome is linked to the evolutionary landscape of ECs (175).

Alternatively, we could try to understand cancer development through assessing the effects of sequential events that lead to endometrial adenocarcinoma. A feasible study would be the characterization of the PI3K pathway in ECs to explore the role of multiple mutations altering the PI3K pathway in a single population of cells. Compared to other cancers where we normally see pathways altered by one hit, we see multiple hits in the PI3K pathway. Therefore, we propose using *in vitro* and *in vivo* models to further clarify the role of the PI3K pathway in endometrial cancer. Future studies could include *PTEN* null cell lines where we could successively test the effects of mutations in *PIK3CA*, *PIK3R1*, and *CTNNB1*. *In vivo* studies showed that $PIK3CA^{E545K/+}$ mutation causes carcinoma in the setting of biallelic *Pten* deletion in mice. $PIK3CA^{E545K/+}$ alone did not cause endometrial carcinoma in mice (180). Therefore, we also propose using *PTEN*- null mice, to investigate the differences in cooperation of biallelic *PTEN* deletion with *FGFR2*, *CTNNB1*, and *PIK3R1* mutations.

CHAPTER 3: Invasive Squamous Cell Carcinomas and Precursor Lesions on UV-Exposed Sites Demonstrate a Concordant Level of Genomic Complexity in Known Driver Genes

3.1 Introduction

Squamous cell carcinoma (SCC) is a major cause of cancer mortality, and is the most common form of human solid tumor in many anatomic sites such as the lung (LUSC), head and neck (HNSCC), and cervix (CESC) (181). Although some SCCs are associated with human papillomavirus (HPV), a Pan-Cancer Atlas study of SCCs suggests that most SCCs are driven by recurrent somatic mutations in tumor suppressors such as *TP53* and *CDKN2A,* as well as copy-number aberrations (CNA) such as 3q, 5p and 11q arm-level gain and 9p loss (181,182). Additionally, studies focused on the molecular characterization of cutaneous squamous cell carcinoma (CSCC) found a high rate of mutations caused by UV damage, recurrent mutations in *TP53, CDKN2A,* and *NOTCH1/2*, chromosome 3q26, 7q21, and 11q22 gains, and *CDKN2A* loss (on chromosome 9p21) (183-186). In addition to well characterized tumors, SCC precursors and invasive lesions from other sites, such as the ocular surface (ocular squamous cell neoplasms [OSSN]), are not as well understood (182,187).

SCCs are thought to develop from hyperplastic precursor lesions characterized by abnormal cell growth from squamous epithelium. SCCs, such as CSCCs and OSSNs, arise in the setting of UV-induced damage or other environmental factors, transforming normal squamous epithelium into actinic keratosis (AK) on the epidermis and intraepithelial neoplasia on the ocular surface (conjunctiva and cornea, CIN). Actinic keratoses (AKs) are likely the most prevalent precancerous lesions in humans (188), generally limited to sun-exposed tissue, and histologically characterized by keratinocytic atypia involving the basal epidermis and disturbances to cell differentiation. Although some AK lesions undergo regression, a subset develop into carcinoma *in*

situ (CIS), a preinvasive stage of invasive CSCC characterized by full-thickness atypia, or invasive CSCC (189,190).

Characterization of unusual subtypes, or SCCs arising at uncommon anatomic sites, has been limited. Amongst them, is the first comprehensive molecular analysis of penile squamous cell carcinoma (PeSCC) conducted by our laboratory (191), as well as ocular surface squamous neoplasia (OSSN), a major focus of the current study. OSSN is the most common cancer of the ocular surface (cornea/conjunctiva) in the U.S. OSSN occurs in two forms: pre-invasive (i.e. conjunctival/corneal intraepithelial neoplasia [CIN] or carcinoma-in-situ [CIS]) and invasive subtypes. Risk factors include UV radiation, HIV, heavy cigarette smoking and petroleum products, immunosuppression, genetic predisposition, and injury (192,193). Some proposed risk factors for OSSN, such as HPV, remain controversial (194). The incidence of OSSN ranges from 0.2 to 35 cases/million/year (195-197). OSSN can be locally destructive and blinding. Though unusual, lethal metastatic OSSN cases have been described (198). Despite surgery and adjunctive chemotherapy (local interferon therapy) (199), OSSNs recur in a third of cases with clear surgical margins and up to 56% of cases with positive margins (195,200). This high relapse rate indicates that current treatments do not adequately control disease, which defines an important unmet need. One path toward improved therapies for OSSN would be the use of agents tailored to the unique genetic or transcriptional alterations in these tumors. No biologically targeted therapies exist because such alterations remain obscure. Despite an explosion in our understanding of the genetic mechanisms related to squamous cancers in other parts of the body (181), only one study has profiled OSSN (187). This impact of that study was limited by analysis of only CIN and CIS lesions, small cohort size, lack of treatment-naïve tumors, and failure to identify actionable targets (187). To our knowledge, neither treatment naïve, pre-invasive OSSN (CIN/CIS) nor invasive OSSN has been hitherto molecularly profiled. The lack of these studies limits our understanding in how these cancers form and hampers our ability to develop molecular therapies against these highly recurrent squamous cancers.

As shown in calls to generate a Pre-Cancer Genome Atlas (PCGA), the molecular progression of precursor SCC lesions to invasive cancer is incompletely understood, in part due to technical challenges posed by the profiling of small areas of interest often only available in routinely processed formalin-fixed paraffin-embedded (FFPE) tissues (83). To date, the genetic alterations underlying epithelial *in situ* lesions have only been comprehensively profiled in a

limited number of cancers, the largest being a recently published study regarding lung cancer (201). Obstacles to defining stepwise models for tumorigenesis in cutaneous malignancies include high burdens of passenger mutations, variability in driver mutations within a given tumor type, and lack of identifiable precursor lesions for some tumor types such as basal cell carcinoma. Despite these challenges, progression of precursor lesions to malignancy has been correlated with tumor suppressor gene inactivation events for a subset of sweat gland carcinomas and some melanomas (87,202,203). Genetic events associated with progression in cutaneous squamous neoplasms are less clear. AKs harbor mutations and methylation profiles similar to invasive SCC (186,204-206). Although AKs display chromosomal aberrations and loss of heterozygosity (LOH) events (207), these appear to be less numerous than in SCC (208,209). Some have proposed that tumor suppressor LOH might be a critical step in transition from AK to SCC (205,210); however, this hypothesis has not been rigorously addressed. Another area of uncertainty relates to genetic changes in CIS, which is premalignant but has distinct microscopic appearance and clinical management from AK. Finally, although ocular epithelium is a UV-exposed site and displays a similar spectrum of precursor neoplasms and invasive SCC, the genomic changes in these treatment naïve or invasive neoplasms remain hitherto uncharacterized.

To better understand genomic changes associated with malignant progression in UV-exposed squamous lesions, we characterized the genomic landscape of two types of SCC, CSCCs and OSSNs. In the present study, we used a highly scalable, comprehensive, multiplexed PCR-based next-generation sequencing approach to profile invasive SCC and its precursor lesions.

3.2 Materials and Methods

3.2.1 Cohort

With local IRB approval, we identified 47 cases of cutaneous squamous cell carcinoma (CSCC) or precursors, and 36 cases of ocular surface squamous neoplasia (OSSN) with available archived formalin-fixed, paraffin-embedded (FFPE) tissues. For each case, regions of interest were identified on hematoxylin and eosin (H&E) stained slides and classified as actinic keratosis (AK) or conjunctival/corneal intraepithelial neoplasia (CIN), carcinoma *in situ* (CIS), or invasive and given a histology-based tumor content by board certified pathologists (S.A.T. and P.W.H.). FFPE blocks were cut to make 4-8 10-µm sections. Although most areas with high tumor purity were macro-dissected using a scalpel, lesions classified as AK or CIN were mainly dissected under the

microscope. DNA and RNA from each sample were co-isolated using the Qiagen AllPrep FFPE DNA/RNA kit (Qiagen) using manufacturer's instructions and quantified using the Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA) as previously described (99).

3.2.2 DNA Next Generation Sequencing

Targeted multiplexed PCR-based DNA (mxDNAseq) next generation sequencing (NGS) was performed using 20 ng of DNA from each sample. DNA libraries were generated as described using the Oncomine Cancer Panel (OCP) targeting 134 cancer-related genes. We used the Ion Ampliseq library kit 2.0 (Life Technologies, Carlsbad, CA) with barcode incorporation and sequencing on the Ion Torrent Proton sequencer using the Ion PI Hi-Q Sequencing 200 Kit as described (99,101). Variant calling and identification of high-confidence, prioritized somatic mutations, in addition to copy number (CN) analysis, was performed as described using well validated pipelines (101). Sample-level variant allele frequencies (VAFs) were used to determine tumor content (TC) and were assign each variant as homozygous or heterozygous. Estimated tumor contents were then used to correct copy-number estimates to account for variability between samples.

3.2.3 cBioPortal

Selected prioritized variants for all samples were visualized using the public OncoPrinter tool available from the cBioPortal for Cancer Genomics. Additionally, the MutationMapper tool was used to map *TP53*, *RB1*, and *CDKN2A* mutations across all samples (211,212).

3.2.4 Amplicon Based Whole Transcriptome Sequencing and Analysis

We performed amplicon-based whole-transcriptome sequencing of samples in singlicate, as previously described (213). The linear models used to fit the contrasts for AK versus invasive SCC, *in situ* versus invasive CSCC, and *RB1* Mut vs WT *in situ* samples did not have an intercept term and followed the model "~0 + factor." Overlapping DEGs of AK versus invasive SCC and *in situ* versus invasive CSCC were used for heatmap visualization. Functional analysis of differentially expressed transcripts between *RB1* mutant vs. wild-type populations was performed using Gene Set Enrichment Analysis version 3.0, developed by the Broad Institute (Cambridge, MA) (152,153). The enrichment was done using a pre-ranked list with the ranking metric being the corrected p-value divided by the sign of the fold-change. The expression data was tested against the hallmark gene set.

3.2.5 RNAscope HPV

To determine HPV status of *TP53* wild-type cutaneous and OSSN cases, we used the RNAscope 2.5 HD Red Reagent Kit and target probes HPV-HR18 (pool probe of 18 high-risk HPV strains, 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 73, 82) and HPV-LR10 (pool probe of 10 low-risk HPV strains, 6, 11, 40, 43, 44, 54, 70; 69, 71, 74) (Advanced Cell Diagnostics Inc., Newark, CA), according to manufacturer's instructions. HR-HPV and LR-HPV infected warts were used as positive control samples. Positive (Hs-PPIB) and negative (DapB) control probes were also used as sample quality control and assay background control. FFPE tissue blocks were cut into 5-µm sections. After deparaffinization and pretreatments, tissue sections were hybridized with target probes, followed by a series of signal amplification steps and chromogenic staining with Fast Red dye. Stained slides were then evaluated for HR and LR HPV infection according to the staining results.

3.3 Results

3.3.1 Comprehensive NGS Profiling of Ocular and Cutaneous SCCs

We performed DNA based NGS using a highly scalable approach optimized for routine FFPE material to assess the molecular profile of SCCs and presumed precursor lesions. To identify oncogenic and tumor suppressive somatic mutations and copy-number aberrations (CNAs), we performed multiplexed PCR-based DNA NGS (mxDNAseq) on a total of 99 spatially-defined, minute cell populations using the OCP which targets 134 cancer-related genes, including nearly all genes known to be recurrently mutated or amplified/deleted in SCCs. Our cohort was composed of 56 FFPE cutaneous tissues (n= 8 AK, n= 30 CIS, n= 18 invasive SCC) from 45 cases and 43 FFPE ocular tissues (n= 2 CIN, n= 22 CIS, n= 19 invasive SCC) from 35 cases from institutions in the US and Brazil **(Table 3.1)**. Representative images CIS and invasive ocular lesions are shown in **Figure 3.1**.

Across the 99 cell populations profiled by mxDNAseq, we achieved an average of 3,183,804 mapped reads yielding 1,253x targeted base coverage across the 102 samples, enabling detection of high confidence, somatic mutations down to ~5% variant allele frequency (VAF). An average of 440 variants per high quality sample passed standard low stringency default filters, however, additional stringent filtering was applied to identify high confidence somatic alterations. All samples underwent variant analysis, with the exception of eight which were excluded due to

mutation signatures indicative of over-fixation. Additionally, we assessed CN profiles of all 99 samples using a previously validated approach (99). To account for variability in tumor content, we used the VAF based TC to adjust CN estimates of all samples, with the exception of nine which were corrected for the histology-based TC since we were unable to calculate a VAF based TC.

Table 3.1 Summary of cutaneous and ocular lesions profiled by NGS.

	Tissue Type	Number of Samples	Number of Pairs	Number of Cases	Notes
Cutaneous Lesions	Actinic Keratosis (AK)	8	0	45	Additional 2 pairs include in situ and invasive
	Carcinoma *in situ* (CIS)	30	3		
	Invasive SCC	18	4		
Ocular Lesions	Conjunctival/ Corneal Intraepithelial Neoplasia(CIN)	2	0	35	Additional 2 pairs includes in situ and invasive
	Carcinoma *in situ* (CIS)	22	3		
	Invasive SCC	19	3		

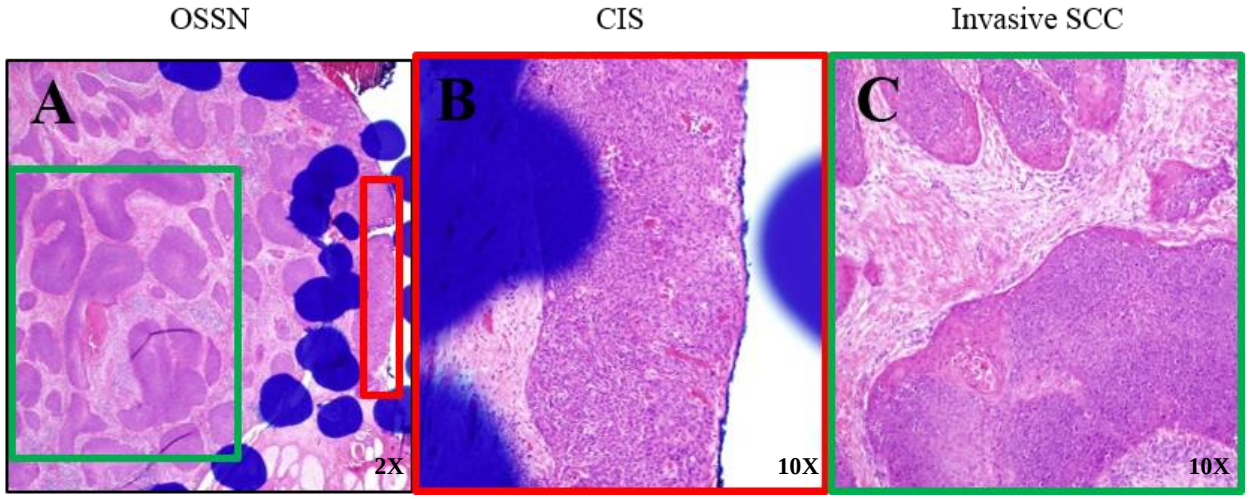

FIGURE 3.1 Representative histology of spatially defined OSSN cell populations.
Histology image of a case of (A) OSSN (2X magnification) with paired (B) CIS (10X magnification) and (C) invasive SCC (10X magnification).

3.3.2 Comparison of Chromosomal Aberrations Present in Ocular and Cutaneous SCCs

In this study, we observed a combination of alterations involving those previously reported in non UV-driven SCCs and UV-driven SCCs. CSCCs and OSSNs harbored *CDKN2A* CN loss

(n= 7/18 [39%] CSCC, n= 15/19 [79%] OSSNs) and 3q gain (n= 3/18 [17%] CSCC, n= 4/19 [21%] invasive OSSNs), with additional *SOX2* focal gains. Furthermore, we observed *CCND1* (n=3/19, 16%), *MYC* (n=4/19, 21%), and *EGFR* (n=3/19, 16%) gains in invasive OSSNs but not in invasive CSCCs **(Figure 3.2A&B)**. We also observed CN gains in chromosomes 7, and 11q as well as CN loss in chromosome 11q in both CSCCs **(Figure 3.2A)** and ocular SCCs **(Figure 3.2B)**. Although previously reported aberrations were also found to be recurrent in our cohort, we observed marked differences between CSCCs and OSSNs. Gains affecting oncogenes including *SOX2* (arm-level or focal), *MYC*, *CCND1*, and *EGFR* were more highly recurrent in OSSNs. Hence, both CSCC and OSSN display genomic loss of *CDKN2A* and 3q gains; however, amplification of other oncogenes may play a more significant role in OSSN.

3.3.3 Copy Number Aberrations Found in Invasive SCC Lesions are also Recurrent in *in situ* Lesions

After correcting for TC, we observed similar, recurrent CNAs within cutaneous and ocular lesions that were present in all three types of lesions, indicating that these precursor and invasive lesions were indistinguishable at the genomic level **(Figure 3.2A&B)**. In addition to *MYC*, *CCND1*, and *EGFR* gains also being present in precursor lesions, *CDKN2A* loss was observed in 8/8 (100%) AK and in 4/30 (13%) CIS lesions. As an example, pair 9, composed of an *in situ* (ID51) and invasive (ID50) lesion, harbors *CDKN2A* CN-loss in both components, suggesting that the loss in this case was an early event **(Figure 3.2A & 3.3A)**. A similar finding was observed with *CDKN2A* in pair 16 composed of ocular lesions IE140A and 140B **(Figure 3.2B)**. Additionally, as shown in **Figures 3.2B & 3.3B**, both *in situ* and invasive OSSNs can harbor 3q gains, an arm-level gain characteristic of SCCs. Therefore, here we show that many of the CNAs presumed to be characteristic of invasive SCC are also found prior to invasion and overt malignant cytology.

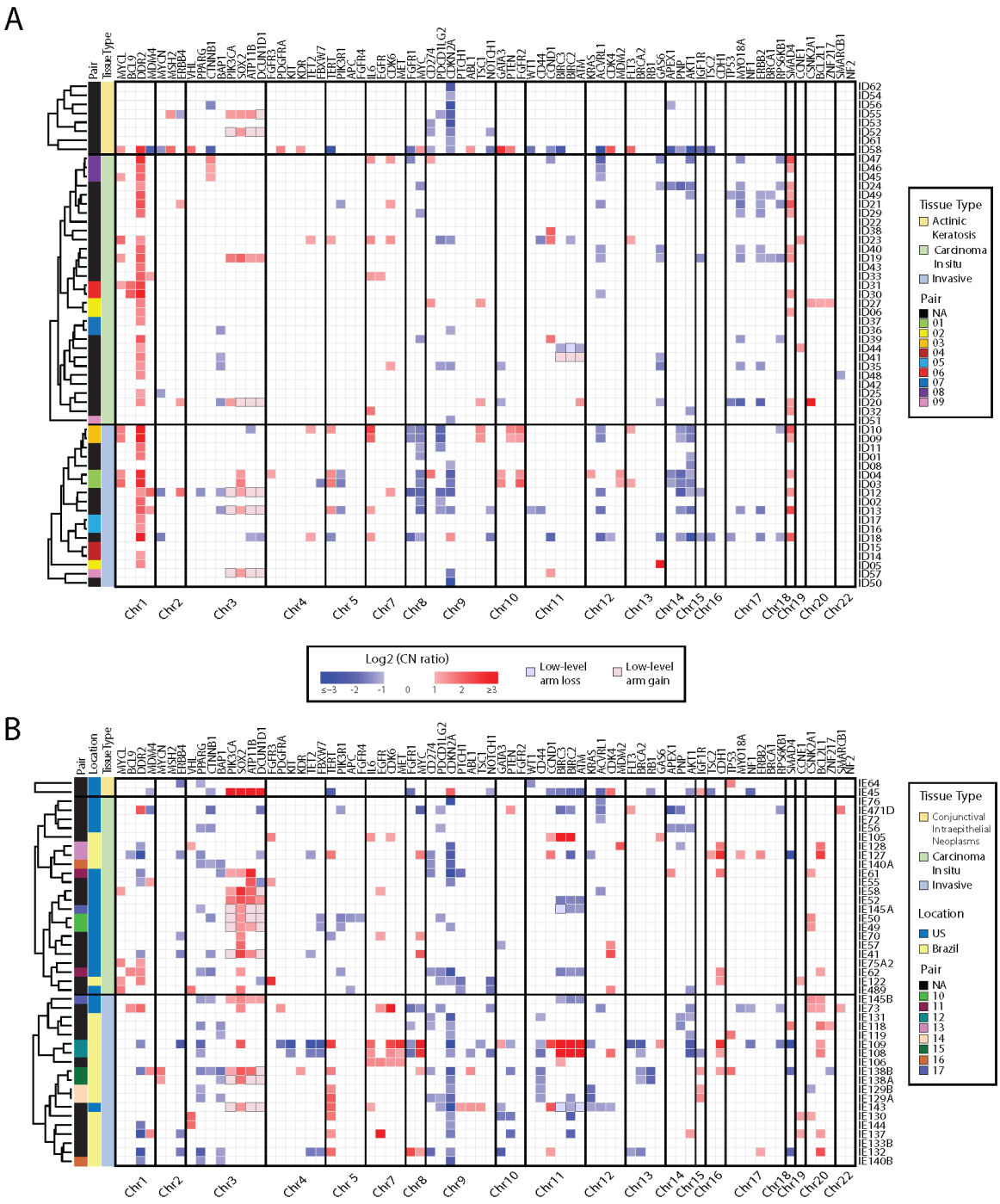

FIGURE 3.2 Somatic copy-number profiles of (A) cutaneous and (B) ocular lesions generated by targeted next generation sequencing (NGS).

Somatic, autosomal copy number profiles are presented for (A) 56 CSCCs and (B) 43 OSSN samples. Each copy-number profile was GC and tumor content corrected. Normalized read counts per amplicon were divided by those from composite normal tissue, yielding a copy-number ratio for each gene (cancer/composite normal), with red and blue indicating gain and loss, respectively, according to the log2 color scale (right). Unsupervised clustering was used on all log2 CN ratios within lesion groups. CN ratios between the range of 1 and -1 were not visualized. Genes part of low arm-level gains and losses are shown with a different shade and border. Columns represent individual targeted genes in genome order (from chromosome 1 to 22). Clinicopathologic features are indicated to the right of the heat map in the figure legend.

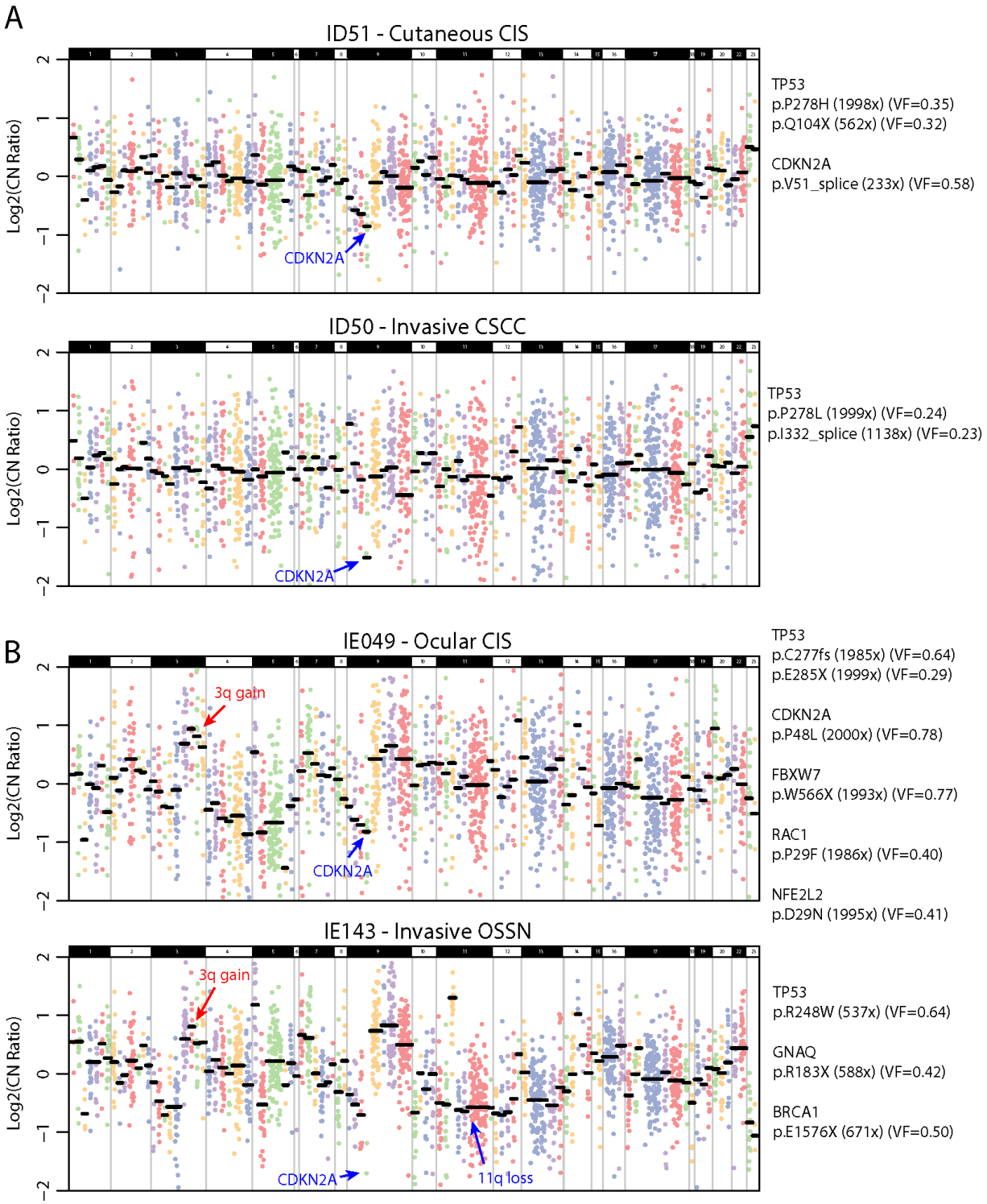

FIGURE 3.3 Copy number aberrations found in invasive SCC lesions are also recurrent in *in situ* lesions
Somatic copy number plots (log_2 copy number ratio, Log2CN) are shown for a representative (A) CSCC and (B) OSSN.

3.3.4 Prioritized Somatic Variants Recurrent in SCCs

After sequencing, we used filtering criteria as specified in the methods section to identify and prioritize somatic variants. Across all cutaneous and ocular tumor types with the exception of CIN (considering each group in aggregate), we observed a predominance of C > T transitions at dipyrimidine sites among single nucleotide variants, as well as a predominance of CC > TT tandem substitutions among dinucleotide substitutions, consistent with UV-mediated DNA damage. Despite reported association with tobacco smoking, C > A substitutions characteristic of tobacco signature mutations are rare in ocular neoplasms. Prioritized mutations were highly conserved in multiple samples collected from the same clinical lesion, supporting the clonal nature of driver mutations, with the exception of one SCCIS-SCC pair from separate blocks from a patient treated with azathioprine (ID50, ID51) which appeared to represent two clonally distinct neoplasms. As previously reported in SCC, we observed recurrent mutations in genes such as *TP53*, *CDKN2A*, *NOTCH1*, *PIK3CA*, and *EGFR* **(Figure 3.4A&B)**. *TP53* is among the most frequently mutated genes in non-HPV driven SCCs from all anatomic sites, including 83% of lung SCC, 71% of head and neck SCC, and 48% of PeSCC. However, second *TP53* mutations are infrequent to absent at other sites, identified 0% of lung SCC, 16% of head and neck SCC, and 8% of PeSCC (191,211,212). Interestingly, both of our SCC cohorts had a high frequency of *TP53* mutations but also of a second or even third mutation across all types of lesions **(Figure 3.4A&B)**. Upon mapping the location of the mutations across lesions, we confirmed that *TP53* mutations in precursor and invasive disease are not only somatic but also affect similar hotspot (e.g. p.R248, p.P278) regions across all types of tissues **(Figure 3.5)**.

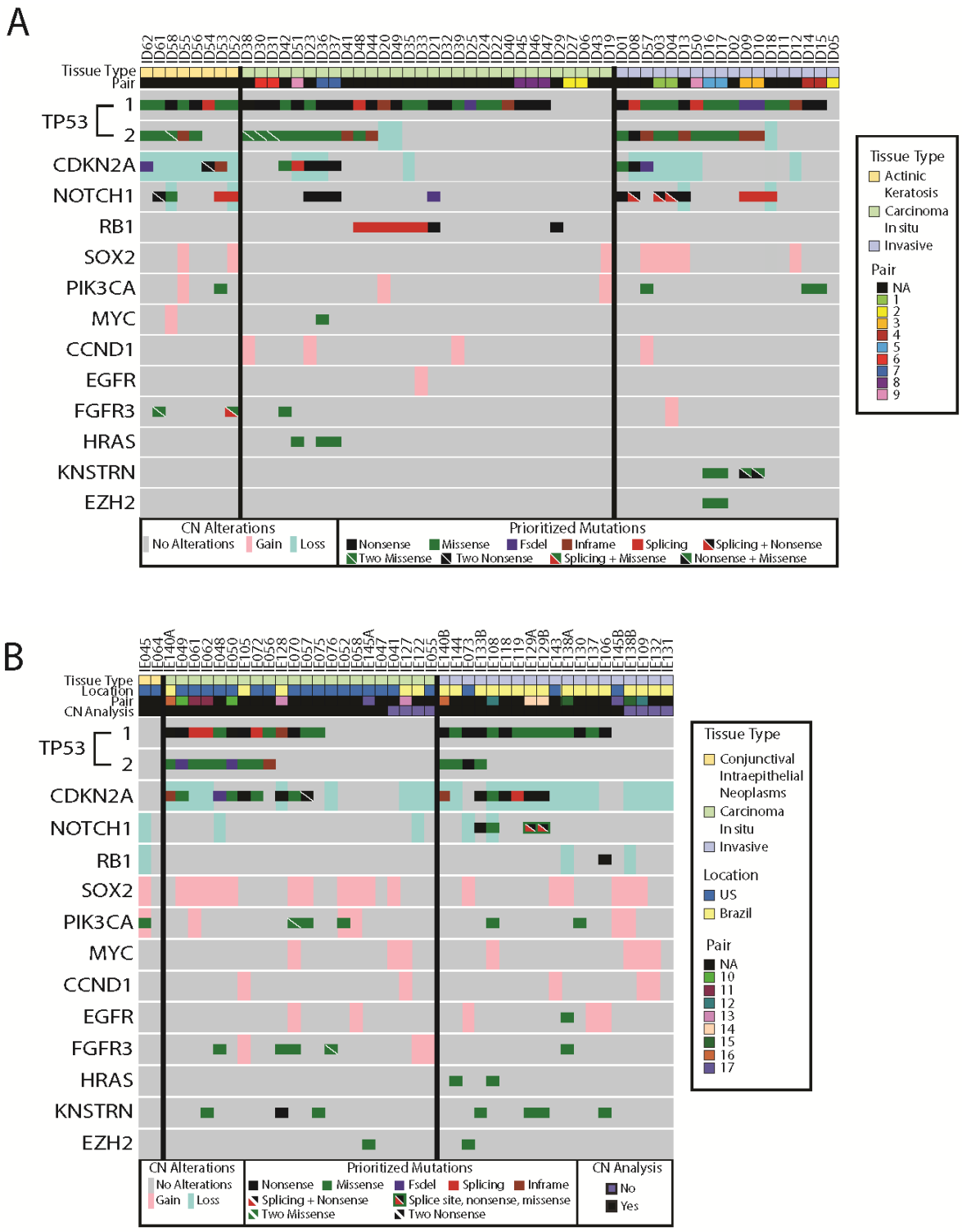

FIGURE 3.4 Integrated heatmap of prioritized mutations and copy number changes identified by comprehensive next generation sequencing.

Integrated table of prioritized nonsynonymous mutations and CNAs from (A) 56 CSCC and (B) 43 OSSN samples. Rows represent genes and columns represent individual samples. Clinicopathologic features are indicated in the figure legend. CNAs and prioritized mutation types are indicated below the table. A total of eight OSSN samples were only analyzed for CNAs are labeled.

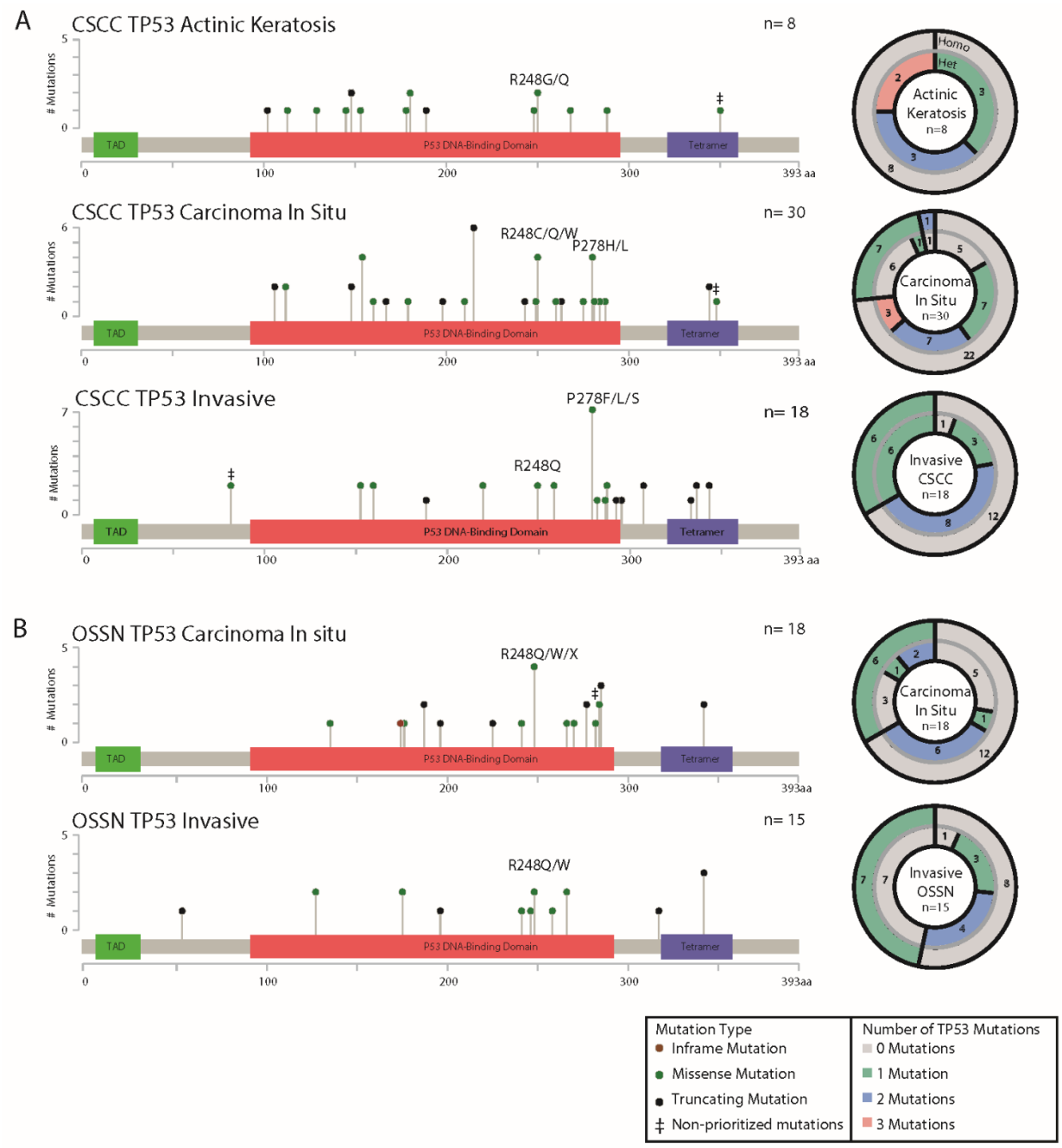

FIGURE 3.5 *TP53* variant mapping and zygosity analysis of CSCC and OSSN.
TP53 mutations in (A) CSCC and (B) OSSN arranged by amino acid location (NM_0546) and separated by histological classification. Mutation type is labeled by color according to figure legend. Concentric pie charts, to the right, show zygosity (each level) and co-occurrence (overlapping regions of the two levels) of *TP53* mutations. Outside circle gives the number of samples with a homozygous (Homo) mutations. Inside circle gives the number of samples with heterozygous (Het) mutations. The number of heterozygous/homozygous *TP53* mutations in each section is denoted by shading. Overlapping regions of the pie chart is meant to help visualize the samples with multiple mutations at homozygous and heterozygous mutations.

3.3.5 *TP53* Copy-Neutral LOH (CN-LOH) in SCC Progression

Studies on normal skin have reported the presence of multiple *TP53* mutations, consistent with multiple small clonally independent cell populations largely defined by these mutations (214). Previous studies have used the ratio of heterozygous to homozygous mutations within a region of copy neutral loss-of-heterozygosity to identify the temporal order of genomic events (215), demonstrating that in CSCC, *TP53* LOH is an early event and may gate most of the remaining mutations present in invasive tumors.

Our analysis identified homozygous *TP53* mutations in 39% (n=13/33) of invasive lesions and 25% (n= 14/56) of precursor lesions. A second or even third mutation at a heterozygous VAF was found in 6 out of the 13 invasive lesions and 4 out of the 14 precursor lesions **(Figure 3.4&3.5)**. In cutaneous lesions, this event was more frequently observed in invasive SCC than precursor lesions; however, a similar trend was not observed for ocular lesions. Most samples lacked *TP53* CN-loss, consistent with *TP53* CN-LOH through a duplication event following the initial loss of the *TP53* wild-type allele. Our data also supports either continued acquisition of *TP53* mutations after the duplication event or the presence of multiple histologically indistinguishable clonal populations, as heterozygous *TP53* mutations were found in both precursor and invasive samples with homozygous *TP53* mutations. Taken together, these results support *TP53* CN-LOH as an early event in SCC development, frequently occurring before invasion.

3.3.6 *CDKN2A* Loss of Heterozygosity in SCC Progression

As described above, *CDKN2A* CN-loss is a recurrent event in cutaneous and ocular SCCs (n= 7/18 [39%] cutaneous, n= 15/19 [79%] ocular). However, we report that *CDKN2A* CN-loss is also present in cutaneous and ocular precursor lesions **(Figure 3.6A & B)**. Interestingly, cutaneous AKs had the highest frequency (n=8/8, 100%) of CN-loss and the highest frequency (n= 2/8, 20%) of homozygous *CDKN2A* mutations. In fact, we found that most samples with a *CDKN2A* CN-loss harbor a *CDKN2A* mutation at a homozygous VAF **(Figure 3.6A)**. Similarly, this observation was also seen in the OSSNs since we see *in situ* and invasive samples with a *CDKN2A* homozygous mutation and CN-loss **(Figure 3.6B)**. Therefore, like *TP53,* our analysis supports *CDKN2A* LOH as occurring at the earliest stages of cutaneous and ocular squamous neoplasia, however *CDKN2A* LOH more frequently occurs through copy loss.

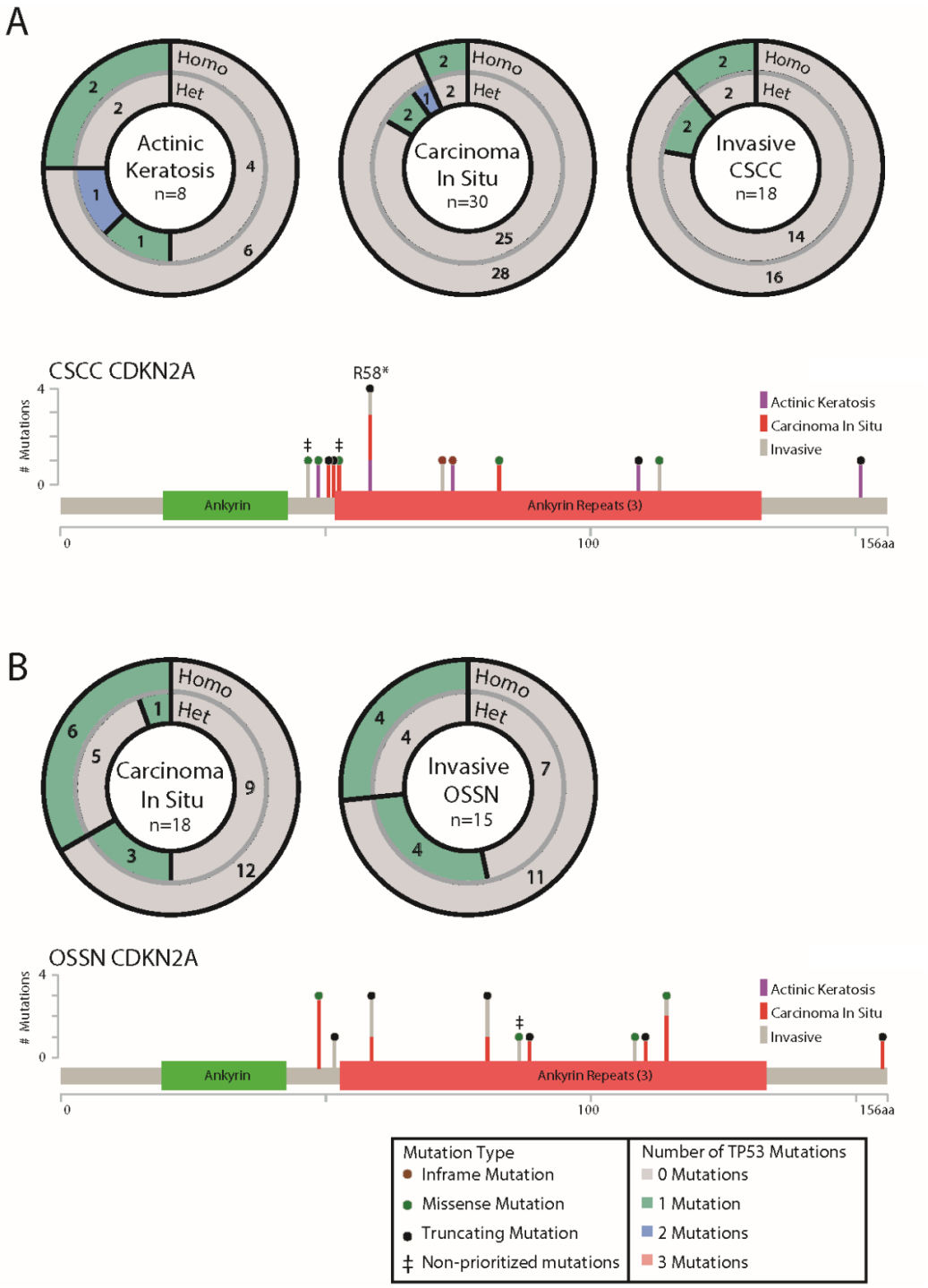

FIGURE 3.6 Two-level concentric pie charts and *CDKN2A* variant mapping.
Concentric pie charts show zygosity (each level) and co-occurrence (overlapping regions of the two levels) of *CDKN2A* mutations, as done for *TP53*. *CDKN2A* variant mapping across (A) AK, CIS, invasive SCC and B) CIS and invasive SCC. *CDKN2A* mutations in CSCC and OSSN were arranged by amino acid location (NM_0077). Histological classification is noted by the color and height of each post. Mutation type is labeled by the colored dot according to the figure legend.

3.3.7 *RB1* Nonsense and Splice Mutations Found Exclusively in Cutaneous CIS Lesions

Unexpectedly, one of the most frequent differences between cutaneous CIS and invasive SCC lesions was the frequency of *RB1* mutations **(Figure 3.4A)**, which is frequently mutually exclusive with *CDKN2A* alterations in many cancers (216,217). Not only did we observe the same mutual exclusivity, but *RB1* homozygous/heterozygous nonsense and splice site mutations were found exclusively in CIS lesions (n=0/8 AK, n= 8/30 [27%] CIS, n= 0/18 SCC,) **(Figure 3.4A & 3.7; Table 3.2)**. Comparison to two TCGA studies of invasive cutaneous SCC confirmed that driving *RB1* mutations were infrequently found in either cohort (6/68 samples with deleterious *RB1* mutations, **(Table 3.2; Figure 3.7**) (183,184). Microscopically, *RB1*-mutated CIS were morphologically heterogeneous from tumor to tumor, with no clear difference from lesions lacking *RB1* mutation. Due to the mutual exclusivity between *CDKN2A* and *RB1*, this observation suggests that *RB1* mutated lesions may display relatively less invasive potential as compared to those with *CDKN2A* mutation.

Table 3.2 *RB1* mutations present in current CSCC cohort and TCGA (183,184).

Cohort	# Samples	Samples with Mutations		Samples with	
		n	%	*n*	%
CIS Lesions	30	12	40%	8	27%
TCGA CSCC	68	11	16%	6	9%

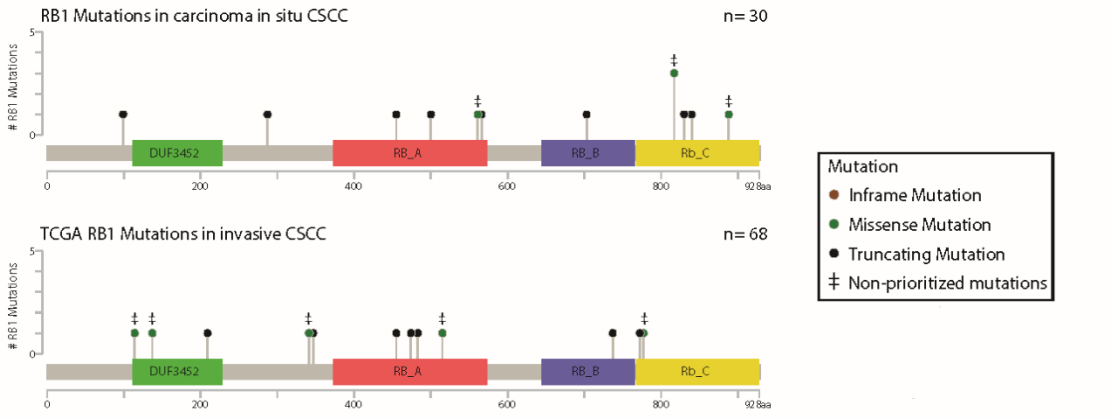

FIGURE 3.7 *RB1* mutations in our cohort versus TCGA.
RB1 mutations in carcinoma *in situ* (CIS) of CSCC (top) and CSCC from two TCGA studies (183,184) (bottom) arranged by amino acid location (top) (NM_0321). Mutation type is labeled by the colored dot and non-prioritized mutations are marked according to the figure legend.

3.3.8 Transcriptome Profiles Distinguish Precursor and Invasive Squamous Cell Neoplasia

Since cutaneous and ocular SCCs when compared to their precursor lesions did not show evidence of changes in known driver genes that lead to invasion, we pursued RNA-seq to determine whether differences arise at the transcriptome level. We found that 129 genes were differentially expressed when comparing invasive SCC against AK and CIS. These included genes previously associated with invasiveness in SCC or other cancer types, including *MMP1*, *MMP3*, *MMP9*, *LAMC2*, *LGALS1*, and *TNFRSF12A* **(Figure 3.8)**. Transcriptome profiling and gene ontology analysis of *RB1*-mutated versus wild-type CIS lesions revealed enrichment for interferon gamma/alpha, inflammatory response, and allograft rejection **(Figure 3.9)**. Although RNA-seq shows differences in expression levels, we are cautious in the interpretation since tumor content could be potentially affecting results.

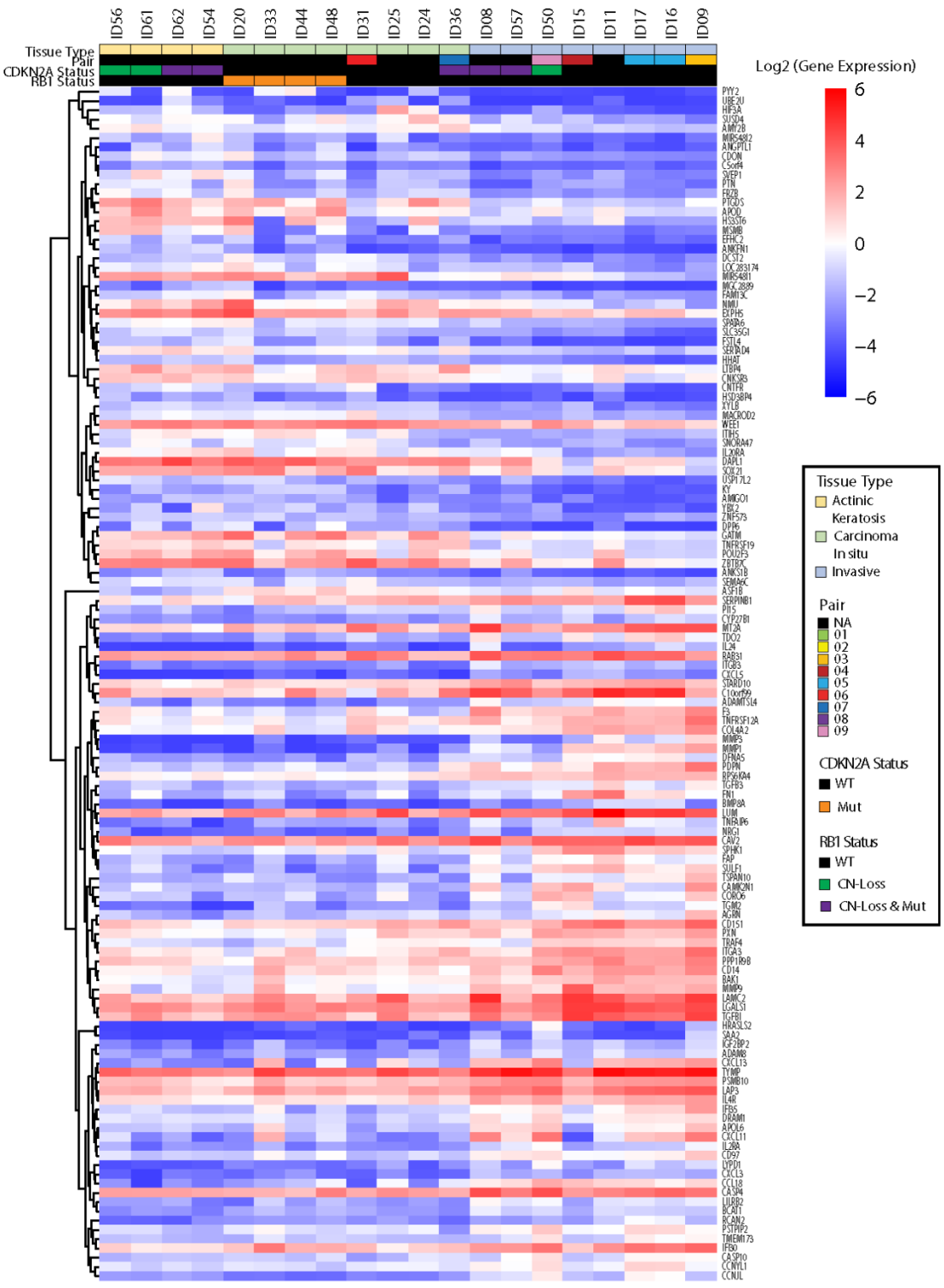

FIGURE 3.8 Gene expression heatmap generated from CSCC amplicon-based RNA-seq.
Heatmap of median-centered expression of 129 overlapping differentially expressed genes (DEGs) from the AK versus invasive SCC and *in situ* versus invasive CSCC comparison.

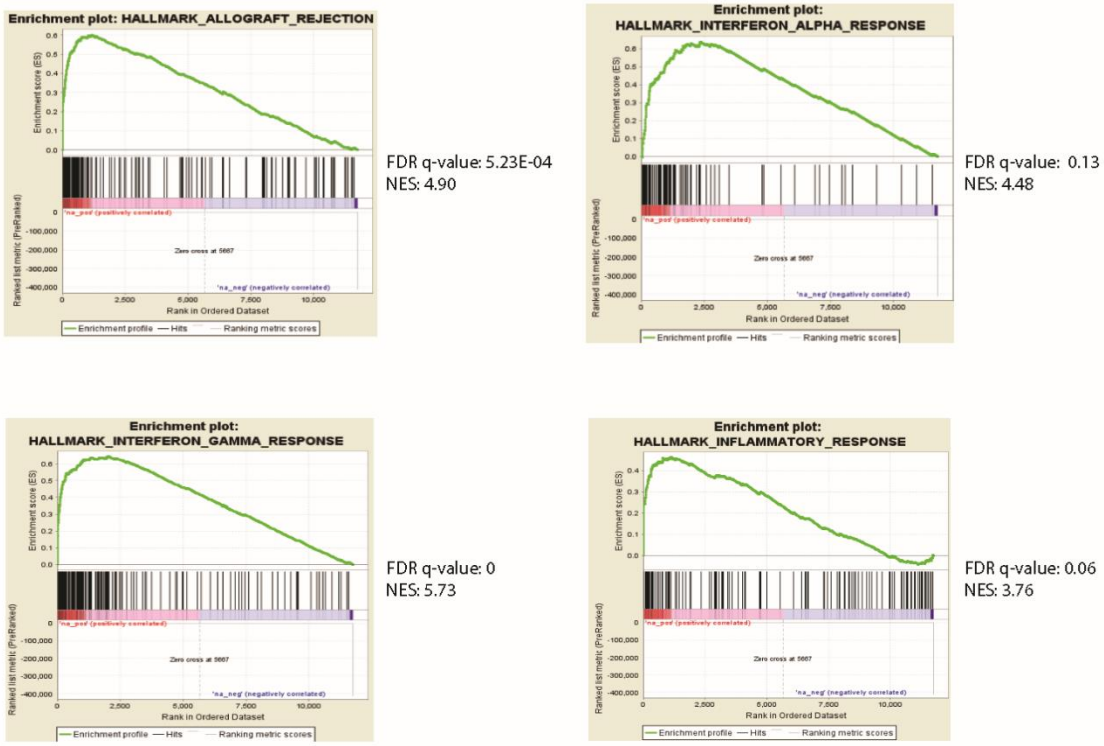

FIGURE 3.9 Assessment of *RB1* mutational signatures.
Gene set enrichment analysis (GSEA) of the entire DEG set from the *RB1* MUT vs WT comparison showed enrichment for four hallmark gene set.

3.3.9 HPV in *TP53* Wild-Type SCC

Of the subset of squamous neoplasms that lacked detectable *TP53* mutation, 3 matched samples from 1 clinical lesion (ID05, ID06, and ID27, including the only *TP53* wild-type invasive CSCC) demonstrated HPV-associated viropathic changes and were positive for high-risk HPV transcript expression. The remaining *TP53* wild-type lesions were negative for HPV and lacked characteristic HPV changes. *TP53* wild-type/HPV-negative lesions displayed amplification or mutation of at least one oncogene, with the exception of one cutaneous CIS with isolated *RB1* mutation, and two CIS (one ocular and one cutaneous) with no prioritized variants.

3.4 Discussion

Cutaneous invasive SCCs have been widely characterized but there are very limited studies on AK or CIS lesions, especially with comparisons to invasive SCCs at the genomic level. Furthermore, the molecular etiology and genomics of ocular SCC, specifically treatment-naïve CIN/CIS, and invasive OSSNs, remain almost entirely unknown. Therefore, we performed

comprehensive mxDNAseq of 134 cancer-related genes on a total of 56 cutaneous (45 cases) and 43 ocular (35 cases) AK/CIN, CIS, or invasive SCC samples. We observed that ocular SCC harbors similar genomic changes to cutaneous SCC, including frequent *TP53* mutations and evidence of UV-mediated genomic damage, but has more frequent copy number gains harboring known oncogenes. In addition, our study demonstrates that many of the previously reported recurrent alterations in SCCs—*CDKN2A* CN-loss and mutations, *TP53* mutations with LOH, as well as 3q, *EGFR*, *CCND1*, and *MYC* gains—are already present well before invasive disease, similar to reports of LUSC and precursor lesions (201).

Although *TP53* alterations are a hallmark of non-HPV driven SCCs, we report a high percentage of samples harboring a second or even third *TP53* mutation in our CSCCs and OSSNs. Multiple *TP53* mutations have also been described in normal skin and in vulvar intraepithelial neoplasias (214,218), interpreted as multiple intermingled clonal populations with distinct *TP53* mutations. Our study suggests that while the presence of multiple non-invasive/invasive tumor populations present at the CIS or invasive stage may be possible, there may also be a single clonal population harboring multiple *TP53* mutations. In fact, studies report that not only does normal skin already have a high mutation burden but also that mutations known to drive CSCC, such as *NOTCH1* and *TP53,* are already under strong positive selection. Therefore, a clone must acquire the proper combination of somatic alterations to outcompete all the other clones present in the skin for malignant transformation to begin (214). Reeves *et al*, drawing upon observations from transgenic mouse models of squamous neoplasia, suggests that while a terminally benign papilloma has a small number of subclones driving growth, a malignant tumor develops after a clonal sweep followed by the development of additional subclones originating from the progressing clone (219). Since we identify *TP53* mutations mainly at homozygous and heterozygous VAFs at all stages of squamous neoplasia, our analysis suggests that a similar selection process in photodamaged human epithelia occurs prior to the formation of microscopically identifiable neoplastic lesions.

The mechanism and sequence through which the accumulation of genomic events lead to malignancy is poorly understood. Although one of our limitations is not knowing which lesions would progress to invasive SCC (precursor lesions on the skin or ocular surface are rarely left *in situ* until the development of invasive disease), our data suggests that precursors may already harbor the full complement of genomic aberrations associated with invasive disease, and may

require other epigenetic or non-genomic events to trigger a transition to invasive disease. In support of this, we find gene expression profile differences between precursor and invasive squamous lesions. However, the mechanism for the shift in transcriptional profile remains unclear.

In addition to these observations, *RB1* mutations were identified in a subset of CIS, but not in invasive SCC. Comparison with other large studies of invasive SCC supports our observation that *RB1* mutations are infrequent in invasive SCC, and hence *RB1*-mutant CIS may represent a molecular subclass with reduced propensity for progression.

Our findings confirm and expand upon previous investigations into the mutational spectrum of UV-associated precursor lesions. Similar to SCCs, AKs have been shown to harbor mutations in tumor suppressors such as *TP53* and *CDNK2A* (186,205), as well as amplifications of *EGFR* and *MYC* (220,221). In contrast to a previous report describing intratumoral heterogeneity in pre-invasive squamoproliferative lesions (222), we found that major genomic events were relatively consistent across multiple areas of a given lesion. Despite a previous hypothesis to the contrary (205,210), we find that *CDKN2A* mutation and LOH are frequent events in AKs, and thus LOH does not represent a likely candidate driver for transition to invasiveness. Similarly, despite studies suggesting p53 inactivation to be a late event (223,224), we observed that the p53 inactivation is already being selected for at early stages of cancer development in cutaneous and ocular squamous neoplasia, more consistent with a whole genome based study of CSCC (215).

Published reports comparing expression patterns in AKs and SCCs have had mixed results (186). This may be due in part to the relatively low power of some studies. The largest such study identified significant differences in gene expression between AKs and SCCs (225). Our cohort is of similar size to Lambert *et al*, and corroborates findings from that report. Our study further benefits from correlations between gene expression and tumor mutational status, suggesting *RB1*-mutated CIS as a molecular subclass with a distinct transcriptional profile.

Previous mutational studies in ocular neoplasms have been limited, with disputed role for UV-associated *TP53* mutations (187,226). A recent exome sequencing study did not comment on mutational profiles within their cohort (187). We find that conjunctival neoplasms harbor evidence of UV mutational damage, and drivers of UV-induced squamous neoplasia are fundamentally similar in both cutaneous and conjunctival epithelia. In addition, our observations in conjunctival squamous neoplasms, although limited by inclusion of few CIN lesions, suggest that a similar

process of subclinical mutation accumulation and clonal sweep may occur in conjunctival epithelia, although we cannot confirm this for lower grades of dysplasia (CIN).

To our knowledge, this is the first next generation sequencing (NGS) study to comprehensively profile OSSNs at pre-invasive, invasive, and treatment-naïve stages. Our results demonstrate the utility of an NGS-based approach, using of small ocular surface biopsies and excisions, to nominate precision therapeutic approaches for OSSNs. The current therapies, surgical excision and topical chemotherapies (i.e. interferon-α2b, mitomycin C, 5-fluorouracil) and surgical excision, are not genetically tailored and are variably effective inasmuch as OSSN has an unusually high relapse rate, even when surgical margins are negative. These treatments can also be associated with ocular pain, limbal stem cell loss, conjunctivitis and other ocular surface toxicities. Such features create an unmet need that could potentially be addressed by currently available, or, in the future, topical, adjunctive therapies that target aberrant pathways related to genetic alterations in *EGFR, FGFR, EZH2,* and *PIK3CA* genes present in OSSNs that we have identified for the first time in this study (227-230).

Our study has several limitations. There was limited follow-up for many patients and few episodes of recurrence, precluding robust associations between genomic profiles and outcome. Our approach does not detect events that affect genes not included in our cancer panel; however, findings from previous exome-wide sequencing studies indicate that our panel encompasses the highly recurrent drivers of CSCC. Finally, our approach does not address tumor mutation burden or epigenetic alterations.

In summary, we find that ocular and cutaneous squamous neoplasms demonstrate a similar spectrum of genetic changes and hence represent parallel models for squamous neoplasia on UV-exposed epithelia. We have profiled invasive and treatment naïve preinvasive OSSNs for the first time. In both ocular and cutaneous settings, precursor lesions already possess the full complement of major genetic changes, including tumor suppressor mutations and loss of heterozygosity events, that are seen in invasive SCC. By contrast, cutaneous precursor lesions demonstrate a distinct transcriptome profile from invasive SCC. In contrast to the stepwise accumulation of mutations proposed for some other malignancies, our findings support the hypothesis that transition to invasiveness in cutaneous SCC may be driven by changes in the transcriptional program rather than acquisition of additional genomic insults. Finally, the alterations we identify here are targetable and provide crucial insights toward novel precision therapies for OSSNs, which

frequently require despite current treatment modalities of surgical resection and topical chemotherapy.

3.5 Future Directions

After characterizing cutaneous and conjunctival SCCs using our targeted approaches, we propose a range of future directions that would entail utilizing a diverse set of techniques to further the current work.

Preliminary immunohistochemistry (IHC) for p53 expression was performed in representative tumors with concurrent dominant negative and truncating *TP53* mutations. We predicted to see loss of p53 expression in cells with truncating *TP53* mutations and p53 overexpression in cells with dominant negative mutations. However, IHC did not clearly show separate cell populations, as predicted. Instead, p53 overexpression throughout the tumor suggested against the possibility of tumors having multiple clones with distinct *TP53* mutations. Hence, we would like to establish whether each type of lesion harbors different clones or whether there is a single population harboring concurrent dominant negative and truncating *TP53* mutations. Therefore, we are interested in identifying the genetic abnormalities pertinent to one cell versus a group of cells by using single cell analysis based on microfluidic methods offered by C1 Fluidigm to assess the type of selection these lesions may be undergoing (231,232). They may represent intermingled clonal populations or each lesion may be entirely composed by a single population at all stages. Alternatively, single cell analysis may also show that earlier lesions harbor more clones than invasive lesions as suggested by the multiple *TP53* clones found in normal skin (214).

Preliminary targeted and amplicon-based whole-transcriptome RNA-seq data showed increased expression of immune markers including *CCL5*, *IDO1*, *TIGIT*, and *CD8A* in invasive cutaneous lesions when compared to precursor lesions. Previous studies have shown that the mRNA profile of 18 genes in tumors can predict clinical response to anti-PD-1 therapy with inhibitors such as pembrolizumab (233). Interestingly, these 18 genes included a few of the differentially expressed genes, described above, identified in our expression analysis. However, association of immune makers to clinical response varies between cancers. PD-L1 expression in melanoma cells is lower than RCC and NSCLC cells, PD-L1 expression in stromal cells in melanoma samples is more predictive of response to anti-PD-1 therapy than expression in tumor

cells (234). Therefore, these studies along with our preliminary results calls for further investigation of the interaction between tumor cells and stromal cells (including immune infiltrating lymphocytes assessed through *TIGIT* and *CD8A* immune expressed genes) to further determine how tumor-associated inflammation can affect response to immunotherapy in cutaneous and ocular squamous cell carcinomas and assess whether there are any predictive RNA-based biomarkers for response to checkpoint blockade immunotherapy.

Since our data suggests that precursor lesions already harbor the full complement of genomic aberrations associated with invasive disease, we propose focusing on other non-genomic events, such as epigenetic modifications. A study focused on characterizing the genomic, transcriptomic, and epigenomic landscape of CIS lesions that progressed to invasive tumors and those that regressed to normal epithelium, suggested that chromosomal instability and methylation changes are progression-specific in LUSCs. Specifically, they reported at least one significant differentially methylated position in *NKX2-1, TERT, DDR2, LRIG3, CUX1, EPHA3, CSMD3, MET, ZNF479, GRIN2A, PTPRD, NOTCH1, CD74, NSD1,* and *CDKN2A*. Since *CDKN2A* CN-loss and mutations are already found in our SCC lesions, we propose examining the role of epigenetic events altering *CDKN2A* in samples with WT *CDKN2A* (88).

CHAPTER 4: Molecular Profiling of Rare Cancers

4.1 Targeted Next Generation Sequencing of *CIC-DUX4* Soft Tissue Sarcomas Demonstrates Low Mutational Burden and Recurrent Chromosome 1p Loss

4.1.1 Introduction

Soft tissue sarcomas are potentially aggressive tumors that are challenging to diagnose and classify due to the morphologic similarities between subgroups (235). Despite advancements in understanding at the morphologic and genomic level, 5% of sarcomas remain unclassifiable in clinical practice (55,236-238). Such tumors have been termed "undifferentiated soft tissue sarcomas" (USTSs) since they do not show histologic or immunohistochemical features characteristic of a specific lineage (55,236-238). Although a subset of USTSs have been characterized as "Ewing-like" tumors due to their round cell morphology and immunophenotype (239), undifferentiated round cell sarcomas (URCS) lack characteristic Ewing sarcoma fusions involving *EWSR1* and members of the *ETS* transcription factor family (132). In 2006, Kawamura-Saito *et al*. reported that some aggressive URCSs harbored fusions of *CIC* (a human homolog of *Drosophila capicua*) to *DUX 4* (*double homeobox 4*), as a result of t(4;19)(q35;q13.1) translocations (240). *CIC-DUX4* sarcomas typically have small round cell morphology, geographic necrosis, coarse chromatin, focal extracellular myxoid matrix, clear cell areas, and mild-moderate nuclear pleomorphism (132,240-242). Furthermore, most cells lack a well-defined cell border and contain vesicular nuclei with often enlarged nucleoli (133).

The *CIC-DUX4* fusion results in a chimeric protein that includes the majority of the *CIC* gene but lacks the homeodomains of *DUX4* (240). CIC is a transcription factor member of the HMG box superfamily that is involved in the development of medulloblastoma (242). *DUX4* has primarily been characterized in the context of muscular dystrophy, where aberrant *DUX4* expression due to epigenetic changes are thought to cause facioscapulohumeral muscular

dystrophy (FSHD) (243). Previous research has shown *CIC-DUX4* fusions expose the *DUX4* C-terminus, resulting in increased activation of *CIC*, even though the DNA binding of the *CIC* HMG domain is largely not affected (240). This suggests that *CIC* downstream targets, such as ETS family members, may be deregulated, supporting *CIC-DUX4* as an oncogenic transcription factor (240).

Given the rarity of *CIC-DUX4* sarcomas, molecular alterations beyond the defining translocation and their molecular relationship to Ewing sarcomas remain poorly understood (55,132,236,240-242,244-249). Previous karyotyping and fluorescence *in situ* hybridization (FISH) studies support chromosome (chr) 8 trisomy and *MYC* amplification as recurrent alterations in *CIC-DUX4* sarcomas (132,249), however a more comprehensive analysis of the genomic landscape of *CIC-DUX4* sarcomas, including assessment of somatic point mutations, small insertions/deletions (indels), and copy number alterations (CNAs), is lacking. Such an analysis is needed given the aggressive course and rapid chemoresistance of *CIC-DUX4* sarcomas and lack of highly efficacious therapeutic strategies (132). Likewise, it is unclear whether *CIC-DUX4* sarcomas are similar to Ewing sarcomas at the genomic level, as Ewing sarcomas have few recurrent point mutations/indels (most frequently involving *TP53* and *STAG2*) but several recurrent, broad, copy number alterations (CNAs). Hence, here we profiled the genomic landscape of eleven formalin fixed paraffin embedded (FFPE) *CIC-DUX4* sarcomas (including three pairs of samples) using targeted next generation sequencing (NGS) of the coding sequence from 409 cancer related genes to assess somatic mutations and CNAs.

4.1.2 Materials and Methods

4.1.2.1 Cohort

We identified eleven *CIC-DUX4* sarcoma formalin fixed paraffin embedded (FFPE) tissue samples from the University of Michigan Department of Pathology Archives **(Table 4.1)**. IRB approval was obtained to perform targeted next generation sequencing on clinical FFPE tumor material. Clinicopathological information for each sample was obtained from the medical record. Hematoxylin and eosin (H&E) stained slides were reviewed by board-certified Anatomic Pathologists (R.P and S.A.T.) to ensure sufficient tumor content. Of the eleven samples, three represented sequential pairs: Samples 5A and 5B represent a pre-treatment primary tumor and a post radiation therapy pelvic recurrence; Samples 6A and 6B represent a post systemic/adjuvant

chemotherapy treated primary tumor and a near concurrent (<1 month) brain metastasis; and Samples 7A and 7B represent an untreated primary tumor resection with no evidence of residual disease and a rapid (<3 months) local recurrence without adjuvant therapy. *CIC-DUX4* rearrangement for all samples was confirmed by RT-PCR and/or FISH as described (132) prior to inclusion in our sequencing cohort.

4.1.2.2 DNA/RNA Isolation

For each sample, 5-8 x 10um FFPE sections were cut from a single representative block and macrodissected with a scalpel to enrich for tumor content. DNA was isolated using the Qiagen AllPrep FFPE DNA/RNA kit (Qiagen, Valencia, CA) as described (101,250). DNA was quantified using the Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA).

4.1.2.3 Targeted Next Generation Sequencing

We performed targeted, multiplexed PCR based next generation sequencing (NGS) using the Ion Ampliseq Comprehensive Cancer Panel (CCP), which targets 1,688,650 bases from 15,992 amplicons representing 409 cancer genes, essentially as described (101,250). Barcoded libraries were generated from 40 ng of DNA per sample using the CCP and the Ion Ampliseq library kit 2.0 (Life Technologies, Carlsbad, CA) according to manufacturer's instructions with barcode incorporation. Templates were prepared using the Ion PI Template OT2 200 Kit v3 (Life Technologies, Carlsbad, CA) on the Ion One Touch 2 according to the manufacturer's instructions. Sequencing of multiplexed templates was performed using the Ion Torrent Proton Sequencer (Life Technologies, Carlsbad, CA) on Ion PI chips using the Ion PI Sequencing 200 Kit v3 (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions.

Table 4.1 Clinicopathologic features of *CIC-DUX4* sarcomas profiled by next generation sequencing (NGS)

Sample	Reference	Sex	Age (y)	Size (cm)	Prior treatment	Tumor location	Metastasis	Tumor content	Karyotype	NGS prioritized mutations	*CIC-DUX4* FISH	*CIC-DUX4* RT-PCR
1	New	F	21	22	None	Leg		80%	N/A	*TP53* C238Y; *TP53* D208fs	N/A	Yes
2	Choi et al	F	20	6.0	Chemo	Shoulder	–	70%	N/A		Yes	Yes
3	Choi et al	F	32	14.0	Chemo	Pelvis	–	60%	17 cells: 46-48,X,t(X;1)(q11.2;p34),del(1)(p21p36), +del(1)(p22p36),t(3;20)(p21;q13.3), t(4;19)(q35;13.1),+8,del(13)(q12.3q14), −14[cp17];3cells:9193,idemx2,+del(13)(q12.3q14)[cp3]93, idemx2,+del(13)(q12.3q14)[cp3]		Yes	Yes
4	New	M	14	17.0	None	Forearm	–	70%	46-47,XY,del(1)(q32q44),t(3;16)(p21;q22), der(4)t(4;19)(q35;q13.1),+17,del(17)(q25q25), der(19)t(19;?22)(p13;q11.2)t(4;19)(q35;q13.1), −20,der(22)t(?19;22)(p13;q11.2)[cp9]			Yes
5A	Choi et al	M	43	9.8	None	Knee	–	70%	N/A		Yes	Yes
5B					Radiation	–	Pelvis	50%	N/A			
6A	Choi et al	F	25	11.0	Chemo	Calf	–	70%	47,XX,t(4;19)(q35;q13.1),+8[15]*	*CTNNB1* E54K	Yes	Yes
6B					Chemo	–	Brain	70%				Yes
7A	New	M	13	4	None	Flank		60%	N/A			Yes
7B	New	M	14	5.5	None	Flank		80%	N/A	*ARID1A* R693X	Yes	Yes
8	New	M	17	7.4	Chemo		Inguinal LN	70%	N/A			Yes

Clinicopathologic information for *CIC-DUX4* sarcomas profiled by NGS. For each profiled sample, the case number, inclusion in prior published studies (Choi *et al*. ref 7), gender, age at original diagnosis, tumor size (greatest dimension, cm) by imaging at diagnosis, prior treatment, primary tumor location (if sequenced), metastatic/local recurrence location (if sequenced), estimated tumor content (by histology), and previously determined karyotype (*from primary, pre-chemotherapy specimen), is included. Prioritized high confidence somatic mutations (see Methods) identified by NGS in this study are also shown. *CIC-DUX4* fusion status for all cases was confirmed as indicated by FISH and/or RT-PCR.

4.1.2.4 Somatic Variant Identification

Data analysis was performed essentially as described (101,250) using validated pipelines based on Torrent Suite 4.0.2, with alignment by TMAP using default parameters, and variant calling using the Torrent Variant Caller plugin (version 4.0-r76860) with low-stringency default somatic variant settings. Variants were annotated using Annovar (251). Called variants were filtered to remove synonymous or non-coding variants, those with flow corrected read depths (FDP) ≤30, flow corrected variant allele containing reads (FAO) ≤6, variant allele frequencies (FAO/FDP) <0.10, extreme skewing of forward/reverse flow corrected reads (FSAF/FSAR <0.2 or >5), FSAF and FSAR >1, or indels within homopolymer runs >4 bases. Variants occurring exclusively in reads with other single nucleotide variants or indels and those occurring in the last mapped base of a read were excluded. Additionally, variants called in >4% of internally sequenced samples using the same panel and not having a cosmic ID were removed. Variants present in ESP6500 or 1000 Genomes (from Annovar) as well as samples with ExAC database (http://exac.broadinstitute.org) at allele frequencies greater than 0.1% were considered germ line

variants and removed. Variants reported in ExAC with observed variant allele frequencies in our data between 0.40 and 0.60 or >0.9 were also considered germ line and removed unless occurring at known somatic mutation hotspots. High confidence somatic variants passing the above criteria were then visualized in IGV. From these somatic variants, hotspots (>1 observation at that residue in COSMIC) in oncogenes, or hotspot or deleterious alterations (nonsense/frameshift variants) in tumor suppressors were then considered as prioritized variants.

4.1.2.5 Copy number analysis

Copy number analysis was performed as described using a validated approach (146,250). Briefly, normalized GC content corrected read counts per amplicon for each sample were divided by those from a composite normal male DNA sample (composed of multiple FFPE and frozen tissue, individual and pooled samples) to identify CNAs through the copy number ratio for each amplicon. Genes with less than four amplicons or a wide distribution of gene-level CN estimates across a large panel of tumor and normal samples internally sequenced on the CCP were removed from all CN analyses.

4.1.2.6 Copy Number validated with quantitative PCR

ARID1A copy numbers were assessed by quantitative reverse transcription PCR (qRT-PCR) for a subset of the cohort with sufficient DNA. Primers and probes (5' FAM; ZEN/Iowa Black FQ dual quenchers) were designed using PrimerQuest (http://www.idtdna.com/Primerquest/Home/Index, hg 19 genome assembly) and obtained from IDT (sequences available upon request). After assay specificity was confirmed using BLAST and BLAT, we excluded primers/probes in areas of SNPs. Each qPCR reaction (15ul) used 5 ng of genomic DNA per reaction, a final concentration of 0.9 uM for each primer and 0.25 uM for each probe in TaqMan Genotyping Master Mix (Applied Biosystems). Triplicate reactions were performed using 384 well plates on the Quantstudio 12K Flex (Applied Biosystems). Automatic baseline and C_t thresholds were set using QuantStudio 12K Flex Real-Time PCR System Software. Log_2 copy number of the genes were determined by the $\Delta\Delta$CT method using the average C_t of *FBXW7*, *DNMT3A*, and *IGF1R* as the reference (copy number neutral by NGS in all samples) and an unrelated FFPE isolated male genomic DNA sample (copy number neutral by NGS) as the calibrator.

4.1.2.7 Sanger sequencing to validate called somatic variants

Bidirectional Sanger sequencing was performed over the prioritized mutations with variant allele frequencies >0.15 on all tumor samples. Genomic DNA (10ng) was used as template in PCR amplifications with Invitrogen Platinum PCR Supermix Hi-Fi (Life-Technologies) with the suggested initial denaturation and cycling conditions. PCR products were subjected to bidirectional Sanger sequencing for both primer pairs by the University of Michigan DNA Sequencing Core after treatment with ExoSAP-OT (GE Healthcare) and sequences were analyzed using SeqMan Pro Software (DNASTAR).

4.1.3 Results

4.1.3.1 Targeted next generation sequencing (NGS) demonstrates a lack of recurrent driving mutations in CIC-DUX4 sarcomas

To assess the genomic landscape of FISH and/or RT-PCR confirmed *CIC-DUX4* sarcomas, we performed targeted NGS on eleven routine FFPE samples from eight patients whose clinical characteristics are presented in **Table 4.1.** Representative histology of three *CIC-DUX4* sarcomas subjected to sequencing are shown in **Figure 4.1**. Amongst the eleven samples, we sequenced two samples from one case (Samples 6A and 6B), representing a primary tumor (located on the calf) and a near concurrent, brain metastasis, both of which had been exposed to prior systemic chemotherapy. We also sequenced two samples from a second case representing a primary pre-treatment tumor (located at the knee) and a post-radiation pelvic metastasis/local recurrence (Samples 5A and 5B). The last sample pair (Samples 7A and 7B) represented a primary, untreated tumor resection from the flank (which obtained pathological confirmed lack of residual disease) and a rapid local recurrence (in the absence of adjuvant therapy).

Targeted NGS from isolated DNA on each sample generated an average of 10,309,255 mapped reads yielding 634x targeted base coverage across the eleven samples. An average of 1,347 variants per sample passed standard low stringency default filters, however after stringent filtering to identify high confidence somatic alterations, we identified a total of twenty high confidence somatic mutations across the eleven samples. Importantly, no genes were recurrently mutated across patients (Samples 7A and 7B harbor the same *KMT2D* A3318G variant at near 0.50 variant allele frequency, which although passing our somatic filtering is likely to be germline; Sample one also harbors a *KMT2D* mutation that is a known rare germline variant, but the observed variant

allele frequency (0.39) is just below our germline filtering threshold (0.40) used herein.)

Likewise, across the eleven samples, we identified only four total prioritized variants (**Table 4.1**). An activating *CTNNB1* E54K was identified in Sample 6A; however, as this variant was present at low variant allele frequency (0.11) and was not detected in the matched brain metastasis from this case (6B), this alteration likely represents a subclonal alteration. *TP53* C238Y and D208fs mutations were identified in Sample 1, with variant allele frequencies of 0.32 and 0.68, respectively, consistent with loss of *TP53* function. Lastly, a prioritized *ARID1A* R693X mutation was identified exclusively in Sample 7B (variant allele frequency 157/417=38% vs. 1/355=0.3% in 7A, **Table 4.1**). The three prioritized mutations with variant allele frequencies >15% (approximate lower limit for Sanger sequencing) were confirmed by Sanger sequencing. Taken together, these results support a lack of candidate driving somatic mutations in *CIC-DUX4* sarcomas.

4.1.3.2 Copy number profiling from NGS data demonstrates recurrent copy number alterations (CNAs) in CIC-DUX4 sarcomas

In addition to mutations, we also assessed our NGS data to identify somatic copy number alterations (CNAs) using a validated approach from NGS amplicon read counts. As shown in **Figures 4.2 and 4.3A**, in contrast to the limited mutational landscape of *CIC-DUX4* tumors, we identified CNAs in all samples, including areas of recurrent gain/loss. Consistent with previous karyotyping/FISH studies that demonstrated chromosome (chr) 8 gain and focal MYC (8q24) amplification in *CIC-DUX4* sarcomas (132,249), we observed broad, low-level chr 8 gain in 4 of 8 cases (**Figs 4.2 and 4.3A**); focal high-level MYC amplification was only seen in Sample 1, suggesting that this alteration may be subclonal. Importantly, of the three cases previously karyotyped (3,4 and 6), chr 8 gain by karyotyping (**Table 4.1**) and NGS was concordant in each case (Samples 6A and 6B showed concordant chr 8 gain by NGS).

Both broad and focal low-level gains centered on ETV4 (chr 17) were observed in 6 of 11 samples; however, this alteration may also be subclonal given that it was only observed in Samples 5B & 6A and not the matched sample from these patients (**Fig 4.3A**). As multiple groups have shown that *CIC-DUX4* sarcomas (and the *CIC-DUX4* fusion protein more directly) over-express ETS transcription factors (239,240,249), including ETV4, our results support low level/subclonal ETV4 gains as contributing to over-expression as well.

Of note, we also identified recurrent, low level deletions centered on the tumor suppressor ARID1A on chr 1p36 (**Fig 4.2 and 4.3A**), which were present in 4 of 8 profiled cases (including concordant loss in Samples 6A & B; loss in case 5A but not 5B). Importantly, case 3, which harbored this chr 1p deletion by NGS, was previously reported to harbor a deletion of chr 1p21-p36 by karyotyping (**Table 4.1**), consistent with our NGS results. In addition, qPCR analysis of genomic DNA on Samples 2-6 supported recurrent ARID1A loss across our cohort (**Fig 4.3B**). ARID1A protein expression by immunohistochemistry (IHC) was assessed on a tissue microarray (TMA) containing *CIC-DUX4* sarcomas and Ewing sarcomas; no consistent differences in ARID1A expression was observed in *CIC-DUX4* sarcomas with or without ARID1A loss or deleterious mutation by NGS (data not shown), supporting 1p single copy loss (or subclonal two copy loss) in *CIC-DUX4* sarcomas, as well as potential regulation of ARID1A protein expression by mechanisms other than genomic loss and/or mutation.

Although the pre- and post-radiation samples from case 5 (Samples 5A and 5B, respectively) showed broad low level chromosome 6p gain (**Fig 4.2**), consistent with clonality, the pre-radiation sample (5A) uniquely showed a broad chr 19 gain and chr 1p loss, while the post-radiation sample (5B) showed a broad chr 7 loss (**Fig 4.2 and 4.3A**), suggesting that this alteration may be associated with post-treatment recurrence and supporting heterogeneity between pre and post-treatment samples. In contrast, other than the low level gain of chr 17 (involving ETV4) exclusively in the primary tumor (Sample 6A), the paired post-chemotherapy primary tumor and nearly concurrent brain metastasis from case 6 showed nearly identical copy number profiles (including gain of chr 8 and loss of chr 18), supporting limited intertumoral heterogeneity (**Fig 4.2 and 4.3A**). The paired primary and untreated local recurrence in case 7 (Samples 7A and 7B, respectively) showed similar copy number profiles, with both samples showing CKS1B gain on chromosome 1 (also observed in Sample 8), PAX5 and SYK loss on chromosome 9, IRS2 gain on chr 13 (also observed in Sample 8), loss of the X chromosome and evidence of chromothripsis on chromosome 7. The recurrence sample (7B) showed unique SOX2 gain on chromosome 3, in addition to the ARID1A non-sense mutation described above.

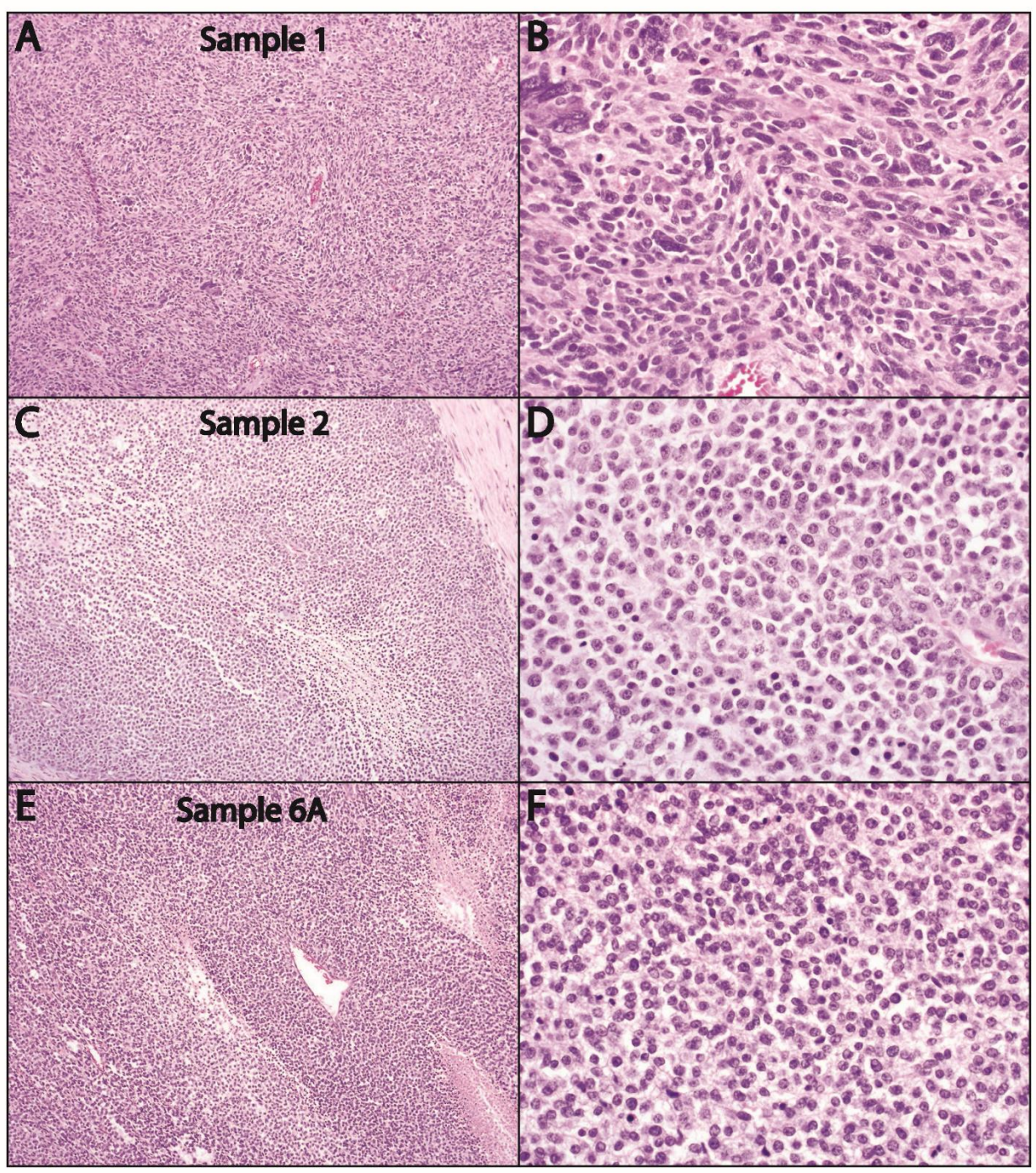

FIGURE 4.1 Histology of *CIC-DUX4* sarcomas subjected to next generation sequencing (NGS).
Low (**A, C & E**) and high power (**B, D & F**) hematoxylin and eosin stained (H&E) stained sections from three *CIC-DUX4* sarcomas, Sample 1 (**A&B**), Sample 2 (**C&D**) and Sample 6A (**E&F**), subjected to NGS. Tumors show typical small round cell morphology, geographic necrosis, coarse chromatin, and focal extracellular myxoid matrix, with Sample 1 showing more pleomorphic histology. Original magnifications 10x (**A, C & E**) and 40x (**B, D & F**).

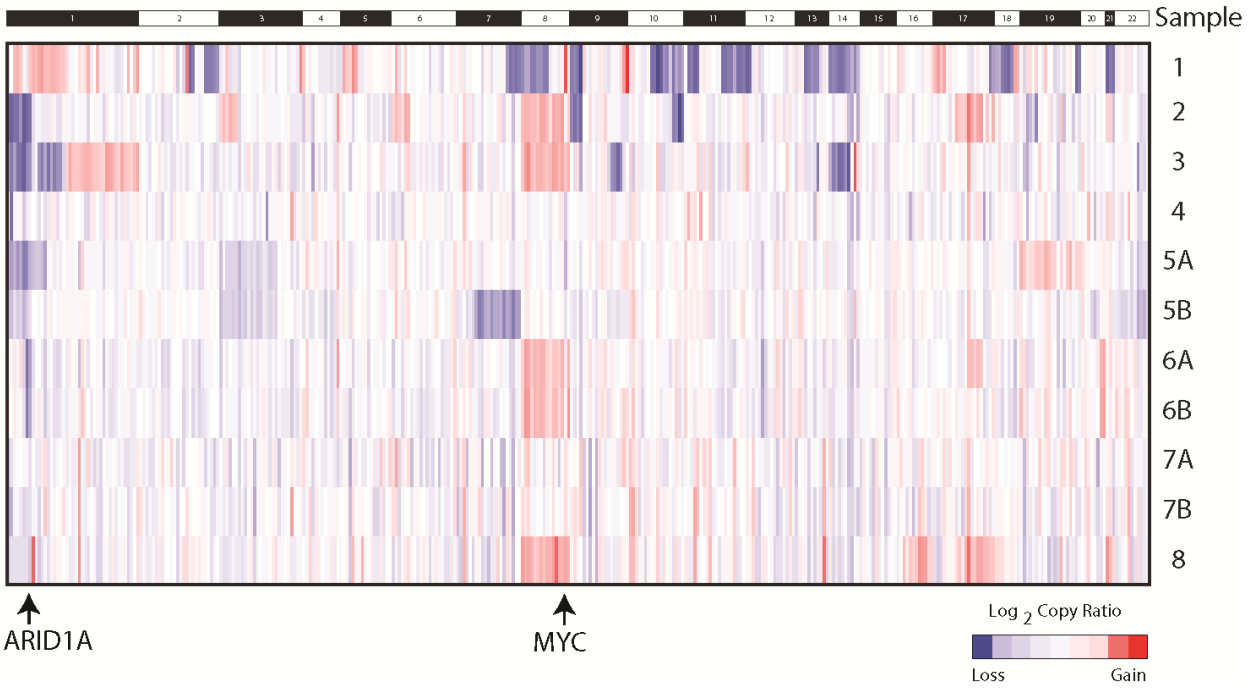

FIGURE 4.2 Somatic copy-number profiles of *CIC-DUX4* sarcomas generated by targeted next generation sequencing (NGS).

Somatic, autosomal copy number profiles are presented for the 11 *CIC-DUX4* sarcoma samples from 8 cases assessed by NGS. Gene-level copy number estimates are shown for all target genes with >= 3 amplicons across samples. Colors correspond to $\log_2$ copy number ratios (tumor to composite normal) as indicated in legend. Samples 5A&B represent a pre-treatment tumor and a post-radiation metastasis/local recurrence from the same patient. Samples 6A&B represent a primary tumor and near concurrent untreated brain metastasis. Samples 7A&7B are a primary tumor and rapid local recurrence.

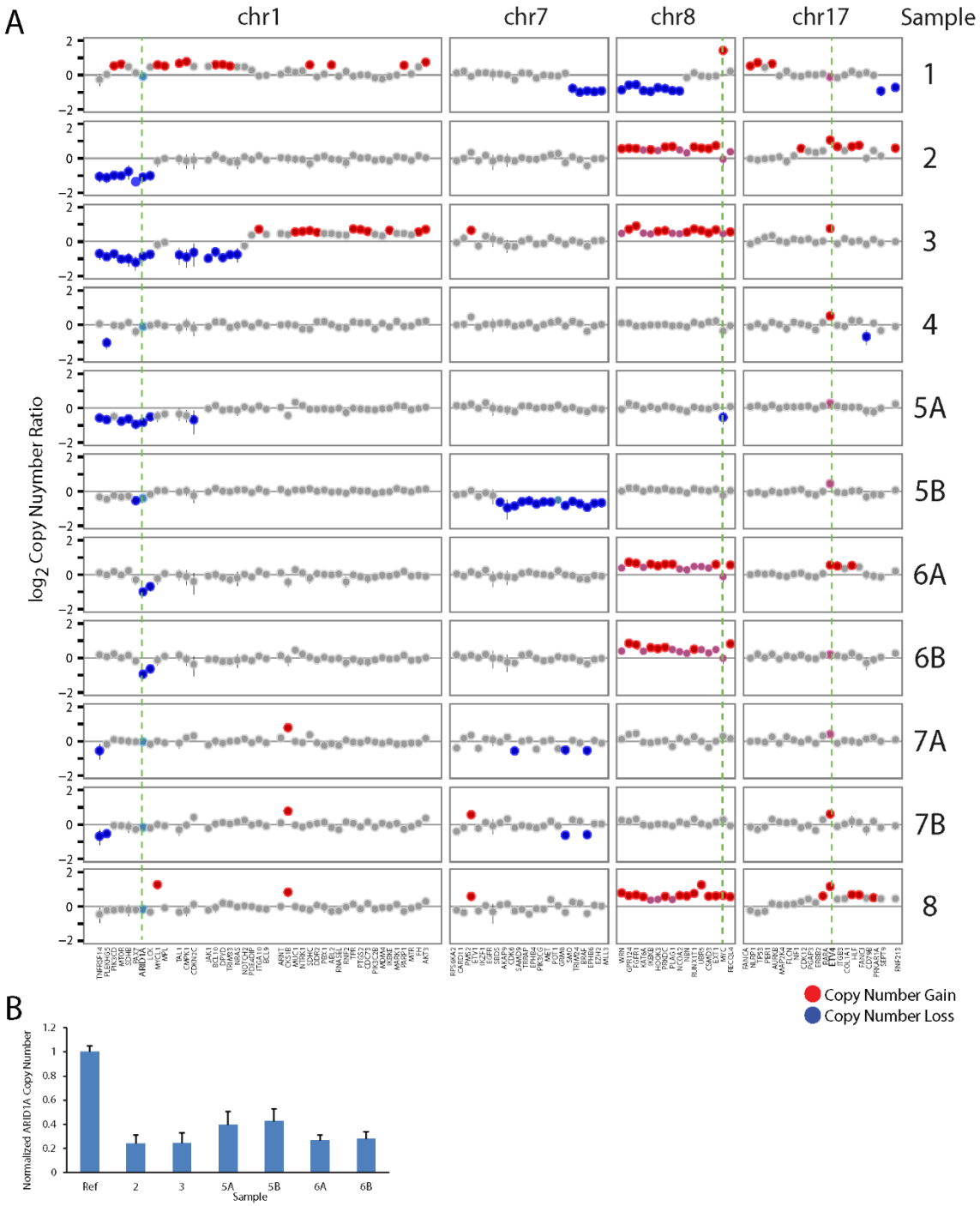

FIGURE 4.3 Recurrent copy number alterations (CNAs) identified by next generation sequencing (NGS) of *CIC-DUX4* sarcomas.

A. Log_2 copy number profiles (tumor to composite normal) for selected chromosomes from the genome wide plots in Figure 4.3 are shown. Gains and losses are shown in red and blue, respectively, with lighter shades indicating lower level alterations. Genes are indicated below the copy number profiles with *ARID1A* (chr 1p) *MYC* (chr 8q) and *ETV4* (chr 17q) bolded and indicated by dashed green lines. B. Confirmation of *ARID1A* copy number loss by quantitative PCR (qPCR). Genomic DNA was assessed by qPCR in triplicate for *ARID1A* copy number normalized to the average of three reference genes without CNA by NGS (*FBXW7*, *DNMT3A*, and *IGF1R*) from samples with available DNA. Normalized *ARID1A* copy number ratio was calibrated to an unrelated benign FFPE genomic DNA sample as the calibrator control (Con.). Mean + S.E. are shown.

4.1.4 Discussion

Although morphologically similar to Ewing sarcomas, *CIC-DUX4* sarcomas are driven by a distinct gene fusion. Given the rarity of *CIC-DUX4* sarcomas, they are much more poorly characterized at the molecular level than conventional Ewing sarcomas. Hence, here we performed targeted NGS on eleven FFPE *CIC-DUX4* samples (from eight patients) to identify somatic mutations and CNAs in *CIC-DUX4* sarcomas as well as to assess the molecular relationship to Ewing sarcomas.

Overall, we did not identify any genes with recurrent driving point mutations. However, one sample (6A) harbored a prioritized potentially driving activating oncogenic mutation (*CTNNB1* E54K, likely subclonal), while sample 1 harbored *TP53* C238Y and D208fs mutations, and Sample 7B (local recurrence) harbored an *ARID1A* R693X nonsense mutation. Our cohort included six previously treated samples (five post-chemotherapy and one post-radiation), suggesting that point mutations/indels are not major drivers of treatment resistance in *CIC-DUX4* sarcomas. Additionally, our data is consistent with Ewing sarcomas, which shows a very low rate of somatic mutations; of note, our panel did not target *STAG2*, which has been shown to be recurrently mutated in Ewing sarcomas (252-254), however case 7 showed a single copy loss of the entire X chromosome (location of *STAG2*, not shown in heatmap), which would result in complete *STAG2* loss in this male patient.

Although we identified limited focal, high level gains or losses in *CIC-DUX4* sarcomas by copy number profiling analysis of NGS data (nearly exclusively in cases 1&8), we identified areas of broad, low level CNAs, including recurrent gains of chr 8, as we have previously reported in *CIC-DUX4* sarcomas (249) and has been observed in Ewing sarcomas (254). Of note, although *MYC* has been nominated as the target of chr 8 gain (and was focally amplified in Sample 1), Sample 8 harbored a gain of chr 8 with a focal high-level gain exclusively of *UBR5* (chr 8q22), centromeric to *MYC*. We also observed low level recurrent gains of *ETV4*, which were focal in some cases, which may contribute to *ETV4* over-expression as has been observed in *CIC-DUX4* sarcomas (239,240,249). Likewise, we identified recurrent low level chr 1p deletions that included the frequently mutated tumor suppressor *AR1D1A* (255) (at 1p36), in 4 of 8 cases, in addition to the *ARID1A* R693X nonsense mutation exclusively in the local recurrence sample of case 7 (Sample 7B). Of interest, chr 1p, and specifically 1p36, has been identified as recurrently deleted in 6-22% of FISH/karyotyping based studies of Ewing sarcomas (256-259). Results from our small

series suggest that this alteration may be more frequent in *CIC-DUX4* sarcomas, and the identification of both *ARID1A* copy loss in multiple samples and a deleterious mutation in Sample 7B is intriguing. Of note, however, ARID1A protein expression was not correlated with copy number/mutation status, and the *ARID1A* nonsense mutation is estimated as heterozygous (variant allele frequency 0.38 and estimated tumor content of 80%). Likewise, 4 of the 5 samples (from 3 of 4 cases) with chr 1p loss (including *ARID1A*) were obtained after neo-adjuvant chemotherapy, although the local recurrence sample harboring the *ARID1A* nonsense mutation (Sample 7B) was treatment naïve. Hence, although *ARID1A* alterations have shown to be subclonal in other tumors through NGS (255), whether *ARID1A* has a role in *CIC-DUX4* sarcoma development or progression is unclear and requires validation in larger cohorts and through functional studies.

Notably, other recurrent copy number alterations seen in Ewing sarcomas (1q gain, 16q loss, 12q gain, and *TP53* (chr 17) deletion (258)), were not recurrent in our limited cohort. However, case 3 showed broad 1q gain, while case 1—which showed atypical morphology—harbored deleterious *TP53* mutations and *CDKN2A* two copy loss. Larger *CIC-DUX4* cohorts will need to be assessed to identify specific CNAs that may occur at different frequencies in Ewing sarcomas and *CIC-DUX4* sarcomas.

Our cohort included three sets of paired specimens, one representing near concurrent, post-chemotherapy primary tumor resection and a brain metastasis (Samples 6A&B), one representing a pre-treatment primary tumor and a post-radiation therapy pelvic recurrence (Samples 5A&B), and one representing a treatment-naïve primary tumor resection and a rapid local recurrence (Samples 7A&B). Samples 6A and 6B showed clonal copy number alterations (including chr 8 gain and 18 loss), however the primary tumor exclusively harbored a *CTNNB1* E54K mutation (at subclonal variant frequency) and low level chr 17q gain (involving *ETV4*), supporting potentially relevant intertumoral heterogeneity. Likewise, although Samples 5A and B both showed chr 6 gain, both samples had unique CNAs, including broad loss of chr 7q in the post-radiation recurrence, supporting intertumoral heterogeneity and potential chr 7q loss as an adaptive response in the radio-resistant clone. Lastly, Samples 7A and B had similar copy number profiles, with the recurrence exclusively showing *SOX2* gain and the *ARID1A* non-sense mutation.

Limitations of our study include the small cohort size, requiring validation of our findings in additional cohorts. Such studies will require intra-institutional collaborations given the rarity of *CIC-DUX4* sarcomas. Likewise, although our NGS panel was designed to assess over 400 known

cancer genes and is capable of detecting both mutations and CNAs, more comprehensive platforms will be needed to assess the existence of chromosomal rearrangements or recurrent mutations/focal CNAs in genes not targeted herein (e.g. *STAG2*). Lastly, future studies will be needed to evaluate any potential clinical implications as well as any biological links between these alterations and *CIC-DUX4* sarcoma development/progression, given the general lack of relevant cell line and animal models.

In summary, using NGS we report the somatic mutation and CNA landscape of eleven routine FFPE *CIC-DUX4* specimens from 8 patients, including three paired samples. Like Ewing sarcomas, we identify a very low mutational rate amongst a large panel of cancer related genes in *CIC-DUX4* sarcomas. Additionally, we identified known (e.g. chr 8 gain) and novel alterations in *CIC-DUX4* sarcomas, including copy number loss and a deleterious mutation in *ARID1A (*chr 1p36). Additional studies are needed to confirm these findings.

4.2 Multiple Primary Merkel Cell Carcinoma: Molecular Profiling to Distinguish Genetically Distinct Tumors from Clonally Related Metastases

4.2.1 Introduction

Merkel cell carcinoma (MCC) is a rare cutaneous neuroendocrine neoplasm. MCC presents as a red-to-violaceous nodule, typically on sun-exposed skin of aged individuals (260). At the time of diagnosis of the primary tumor, there is at least a 15-20% risk of clinically occult metastasis to the regional nodal basin (261-263). The most common site of metastasis is the regional lymph node basin, followed by skin, lung, liver, bone and other sites of distant metastasis. Evidence suggests that MCC may arise via two pathways: a virus-associated pathway mediated by the integration of the oncogenic MCPyV or a UV-damage pathway associated with a high mutation burden, UV-signature mutations, and inactivation of the tumor suppressors *RB1* and *TP53* (264-266).

The phenomenon of multiple primary tumors has been observed in melanoma, where distinct cutaneous primary tumors are clonally unrelated (267). Importantly, the designation of an additional distinct primary melanoma impacts management, as the lesion is treated as a primary melanoma with excision and possibly additional staging with sentinel lymph node biopsy, rather

than a distant cutaneous metastasis. Following a diagnosis of MCC, the development of additional cutaneous tumors, whether adjacent to the treated primary site, within the draining lymphatics or on distant skin, is usually thought to represent a local, in-transit or distant cutaneous recurrence of the original tumor. However, in rare cases, patients may present with a second cutaneous MCC that is spatially and/or temporally separated such that the lesion is clinically designated an independent primary MCC rather than a cutaneous metastasis (268-275). Given the rarity of multiple primary MCCs, genetic relatedness has been evaluated only in three cases (270,274,275). One case demonstrated genetic un-relatedness by analyzing sequences of the integrated MCPyV (270) and the other cases demonstrated clonality by comparative genomic hybridization analysis (CGH) of chromosomal copy number changes (274,275). The distinction between two primary tumors and a primary/metastasis pair has significant impact on treatment and prognosis.

Next generation sequencing (NGS), which provides a broad profile of mutations and chromosomal copy number alterations (CNAs) within a tumor, is ideally suited for clonality analysis (101). In this study, we evaluate clonal relatedness in four patients with clinically designated multiple primary MCCs using NGS.

4.2.2 Methods

4.2.2.1 Cohort

All studies were conducted according to protocols previously approved by the Institutional Review Board of the University of Michigan. Seven MCC cases (14 tumors) designated as multiple primary tumors (distant metastases not suspected) were identified from a database of 473 cases at the Multidisciplinary MCC Program at the University of Michigan from 2006 through 4/2016. Inclusion criteria were availability of paraffin blocks, adequate tumor for sequencing, and adequate quality DNA for NGS analysis; 4 cases (8 tumors) met these criteria. For case 4, a matched regional lymph node metastasis was also sequenced. Metastases in other cases did not yield adequate tumor purity or DNA quantity/quality for inclusion. Clinicopathologic features are summarized in Table 4.2. Two primary-metastasis pairs previously sequenced by the same methods were included in analysis (276).

4.2.2.2 Targeted Next Generation Sequencing

Targeted NGS assessing the complete coding sequence of 409 cancer related genes on archived FFPE (formalin-fixed paraffin embedded) material to identify somatic mutations and

CNAs was performed using the Ion Ampliseq Comprehensive Cancer Panel (CCP) (ThermoFisher Scientific, Waltham MA) as described previously (250,276).

4.2.2.3 Somatic Variant Identification

NGS data analysis was performed essentially as described (250,276) to exclude probable germline mutations and nominate high-confidence somatic single nucleotide variants and insertions/deletions (indels).

4.2.2.4 Copy Number Analysis

CNAs were identified as described (146,250,276).

4.2.2.5 Clonality Analysis

The fraction of shared somatic mutations and CNAs between samples was determined by similarity index calculation (277,278).

4.2.2.6 MCPyV PCR

Detection of MCPyV sequences in tumor DNA was performed by qPCR and PCR-Sanger as described (279-282) using novel or customized primers.

Table 4.2 Case Summaries

Case	Clinical Designation and Tumor Number	Site	MCPyV Copy Number	CNA	Mutations	CNA Similarity Index[a]	Mutation Similarity Index	Time to Second MCC Tumor, mo	Clonality	Clinical Outcome[b]
1	First primary (MCC 24)	Left finger	0.57	3	67	NA	NA	NA	NA	NA
	Second primary (MCC23)	Right finger	1.45	2	4	0	0.014	6	Nonclonal	ANED (72 mo)
2	First primary (MCC 26)	Left elbow	671.00	5	30	NA	NA	NA	NA	NA
	Second primary (MCC25)	Left thigh	471.00	4	0	0.80	0	19	Clonal	ANED (48 mo)
3	First primary (MCC28)	Right ala	260.20	3	2	NA	NA	NA	NA	NA
	Second primary (MCC29)	Left ala	589.40	2	4	0.67	0	Synchronous	Clonal	ANED (16 mo)
4	First primary (MCC 30)	Left cheek	0	1	60	NA	NA	NA	NA	NA
	Second primary (MCC32)	Right cheek	0	4	64	0	0	42	Nonclonal	Progression to visceral metastases

Similarity index was calculated as number of shared events over number of total events. MCC, Merkel cell carcinoma; MCPyV, Merkel cell polyomavirus; CNA, chromosomal copy number alteration; LN, lymph node; SLN, sentinel lymph node; N/A, not applicable. ANED: Alive with no evidence of disease. Months post-treatment: since last surgery.

4.2.3 Results

4.2.3.1 Cases

Seven cases were identified in which a second MCC tumor was clinically designated a second primary tumor rather than a recurrence or distant metastasis. Four cases had sufficient quantity and quality of DNA for clonality assessment. The clinical histories of the 4 analyzed cases are outlined below and summarized in Table 4.1.

Case 1. A man in his 70s was diagnosed with a primary MCC on the left third finger. He underwent amputation and sentinel lymph node biopsy (SLNB) revealing clear margins at the primary site and 2 negative sentinel lymph nodes (SLNs) from the left axilla. He did not undergo adjuvant radiation therapy. Six months later, biopsy of a lesion on his right (contralateral) first finger showed MCC. Re-staging imaging was negative for metastatic disease. He underwent excision and SLNB revealing clear margins at the primary site and 3 negative SLNs from the right axilla. Adjuvant radiation therapy was not indicated. He has been free of disease since treatment of the presumed second primary MCC almost 6 years ago.

Case 2. A man in his 80s was diagnosed with a primary MCC on the left elbow. He underwent excision with clear margins and left axillary SLNB interpreted as negative at an outside institution. Nineteen months later, biopsy of a cutaneous lesion on his left thigh revealed MCC, at which time he presented to the University of Michigan for further evaluation. Upon review, the initial left axillary SLNB was found to be positive for microscopic MCC in 1 of 2 lymph nodes. Therefore, at the time the newly diagnosed lesion on the thigh was presumed to be a distant cutaneous metastasis. Re-staging imaging was negative for metastatic disease. The lesion on the thigh was excised, but SLNB or adjuvant therapy was not recommended based on the presumption of stage IV disease. Four months later, he developed a nodal MCC metastasis in the left groin. Inguinal lymph node dissection revealed 2 of 26 lymph nodes positive for MCC. He did not undergo adjuvant radiation therapy. At this point, based on patterns of metastasis, the tumors were considered to represent two primary MCCs, each with regional nodal metastases. The patient has been free of disease for nearly four years since the left inguinal lymph node dissection.

Case 3. A woman in her 70s was diagnosed with two concurrent MCCs on the left and right nasal alae. Staging imaging did not demonstrate metastatic disease. The lesions were designated as 2 primary tumors and she underwent excision and SLNB for both sites. Pathology revealed clear margins at both primary sites and 2 of 4 positive SLNs in the left neck, and 3

negative SLNs in the right neck. Based on the uncertain clinical presentation, the patient underwent bilateral neck dissection revealing additional 23 and 41 negative lymph nodes in the left and right neck, respectively. She did not undergo adjuvant radiation and has been free of disease for 1.3 years after surgery.

Case 4. A man in his 70s was diagnosed with a large primary MCC on the left cheek with a concurrent nodal metastasis in the left parotid gland. He underwent excision, parotidectomy, and left neck dissection, which revealed a positive deep margin at the primary site and 9 of 20 lymph nodes positive for metastatic MCC. He underwent adjuvant radiation therapy to the left cheek and neck. He had no evidence of recurrence until 42 months later, when he developed a second MCC on the right cheek. Re-staging imaging was negative for metastatic disease. Under the presumption that the lesion represented a second primary MCC, he underwent excision and SLNB, which revealed clear margins at the primary site and 2/2 right parotid SLNs positive for MCC. He underwent right parotidectomy and neck dissection with 22 additional negative lymph nodes in the right neck. He declined adjuvant radiation therapy due to significant xerostomia from his prior contralateral treatment. Unfortunately, he had progression of disease with recurrence in the right neck and liver.

4.2.3.2 Merkel Cell Polyomavirus (MCPyV) Status

75% (6 of 8) of the primary tumors in cases 1-4 harbored MCPyV large and small T antigen sequences **(Figure 4.4, Table 4.2)**. Both tumors from a given patient had similar MCPyV copy number **(Table 4.2)**. One patient (case 4) lacked detectable MCPyV in the two primary tumors **(Figure 4.4, Table 4.2)**. In 5 tumors, adequate DNA remained for partial sequencing of the MCPyV large T antigen exon 2, revealing identical sequences across all cases, without tumor-specific mutations in the analyzed region.

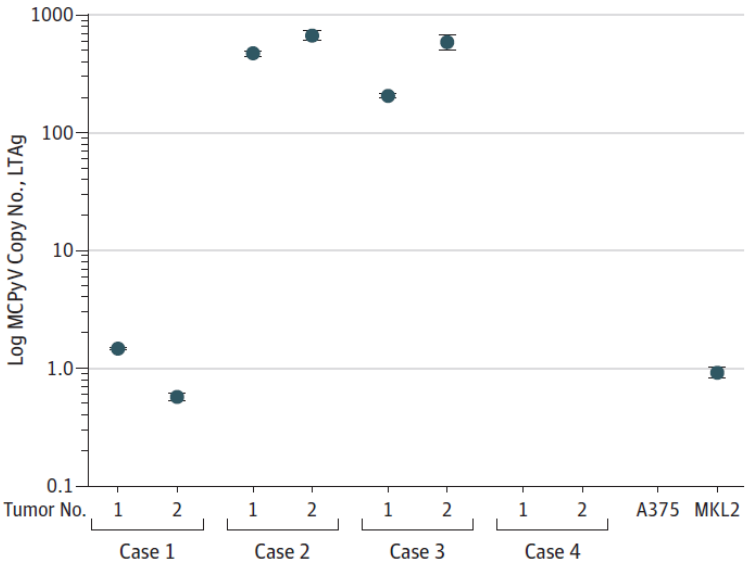

FIGURE 4.4 Merkel cell polyomavirus status of clinically designated multiple Merkel cell carcinoma primary tumors. Quantitative polymerase chain reaction was performed on tumor DNA to detect Merkel cell polyomavirus (MCPyV) sequences. Primer pairs targeted viral large T antigen (LT2) sequences. A375 indicates the negative control melanoma cell line. LTAg: large T antigen. MKL2: positive control Merkel cell carcinoma cell line.

4.2.3.3 Mutational and copy number analysis of MCC

We assessed somatic, high confidence mutations and copy number alterations in 409 cancer-related genes across the primary tumors from the 4 cases, as well as a regional lymph node metastasis in case 4. Results are summarized in Table 4.2 and 4.3. Recurrent events, which suggest clonality, were identified in a subset of paired tumors, as described in clonality analysis below **(Figure 4.5)**.

4.2.3.4 Clonality Analysis of MCC Tumor Pairs

Clonality was evaluated in the clinically-designated multiple primary MCC tumors by quantitating the fraction of shared genetic events in a pair using the similarity index calculation (278). To determine the expected range of overlap for clonally related MCC tumors, we first assessed genetic similarity in the primary MCC tumor and matched regional metastasis from case 4, as well as two primary-metastasis pairs (MCC9-MCC14 and MCC10-MCC16; cases 5 & 6) previously sequenced using an identical approach (276). Primary MCC tumor-metastasis pairs (n = 3) displayed similarity indices ranging from 0.21 to 1.0 for CNA and 0.09 to 0.91 for mutational analysis **(Figure 4.6)**. Thus, CNA and mutational similarity indices > 0.21 and 0.09, respectively, were used as the minimum scores representing likely clonal relatedness. Analysis of random pairings (primary tumors from different patients, representing 24 total pairings of cases 1-4), served as a negative control for clonality analysis and displayed no similarities (similarity index =

0 for both CNA and mutational analyses). For case 1, similarity indices were zero (not clonal) and 0.014 (not clonal), for CNA and mutational analyses, respectively. For case 2, similarity indices were 0.8 (clonal) and zero (not clonal) for CNA and mutational analyses, respectively. For case 3, similarity indices were 0.67 (clonal) and zero (not clonal) for CNA and mutational analyses, respectively. For case 4, similarity indices were zero (not clonal) for both CNA and mutational analyses. In summary, cases 1 and 4 were confirmed as independent primary tumors without genetic overlap, while cases 2 and 3 demonstrated genetic relatedness in chromosomal copy number changes and were designated as clonally related **(Figures 4.5 and 4.6)**.

To evaluate the discrepancy in the clonality results between CNA and mutational analyses, we evaluated the allele frequency of mutations. Mutations in MCPyV-negative tumors were observed at an average allele frequency of 49.2%, consistent with heterozygous mutations present in most or all tumor cells. In contrast, mutations in MCPyV-positive MCC display significantly lower allele frequencies (average 14.3%), suggesting mutations affecting only a minority of tumor cells. This pattern was consistent in an independent cohort of previously sequenced MCC tumors[5]. Mutations affecting a minority of cells in MCPyV+ tumors might not be present in tumor cells that metastasize, and/or might arise after metastasis. Hence, lack of shared mutations is not informative regarding clonality in this context. Taken together, our results support CNAs as earlier events in MCPyV-positive MCCs, with the vast majority of somatic mutations being subclonal (thus passenger mutations).

Prioritized somatic driving mutations were also evaluated in the multiple primary tumor cases determined to be clonally unrelated tumors. Prioritized driver mutations were not identified in case 1 which displayed unrelated primary MCC tumors on the left and right fingers. Prioritized driver *TP53* and *RB1* mutations were identified in case 4 which displayed unrelated primary MCC tumors on the left and right cheek and a matched nodal metastasis from the right cheek. The primary tumor on the right cheek and the matched nodal metastasis harbored the same *TP53* (Q317X and P32L) and *RB1* (Q637X) mutations. The primary tumor on the left cheek harbored distinct *TP53* (H179Y) and *RB1* (R552X) mutations. Together these results support the assessment of clonally independent tumors based on separate driving mutations.

Table 4.3 Summary of next generation sequencing results in MCC

Result	MCPyV-Negative MCC (n = 3)	MCPyV-Positive MCC (n = 6)
Mutations, mean	64.0	17.9
Tumor suppressor genes with inactivating mutations in at least 1 tumor	*NOTCH1* *RB1* *FANCC* *ASXL1* *ATM* *CREBBP* *FBXW7* *TP53*	*NOTCH1*
Oncogene activation mutations	None	None
Copy number alterations, mean	3.0	3.2

Abbreviations: MCC, Merkel cell carcinoma; MCPyV, Merkel cell polyomavirus.

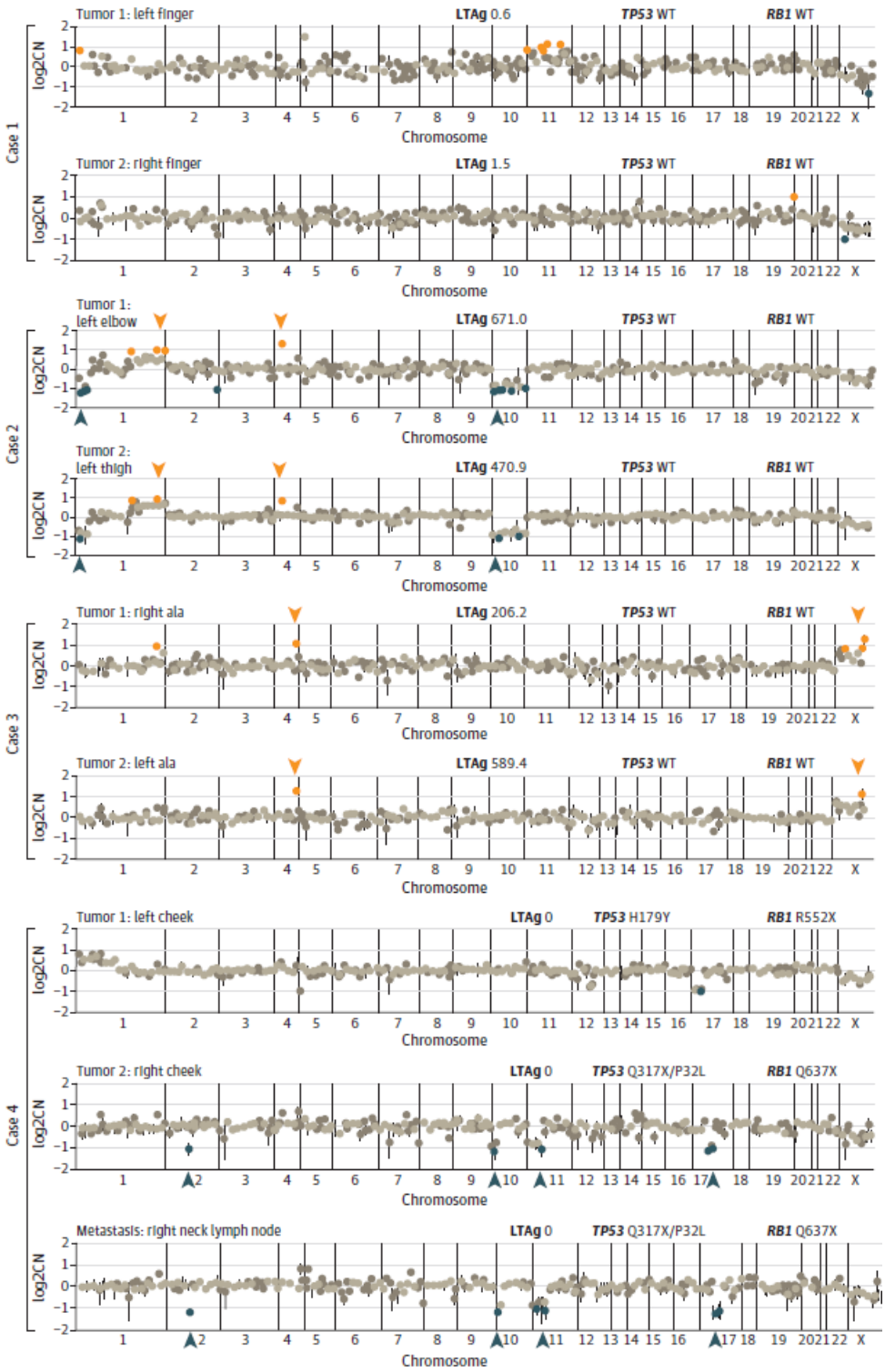

FIGURE 4.5 Copy number alterations demonstrate genetic distinct tumors in two (of four) cases and genetic relatedness in two (of four) cases of clinically-designated multiple primary MCCs.
Somatic copy number plots (log_2 copy number ratio, Log2CN) are shown for each pair of clinically-designated primary tumors. For case 4, a matched nodal metastasis from the second primary is also shown. Shared gains are highlighted by orange arrowheads, and shared losses by blue arrowheads. Genomic copy number for Merkel cell polyomavirus large T antigen (LTAg) and mutation status of *TP53* and *RB1* genes are shown at the top of each plot. Mutations are shown by resulting amino acid substitution. WT: wild type.

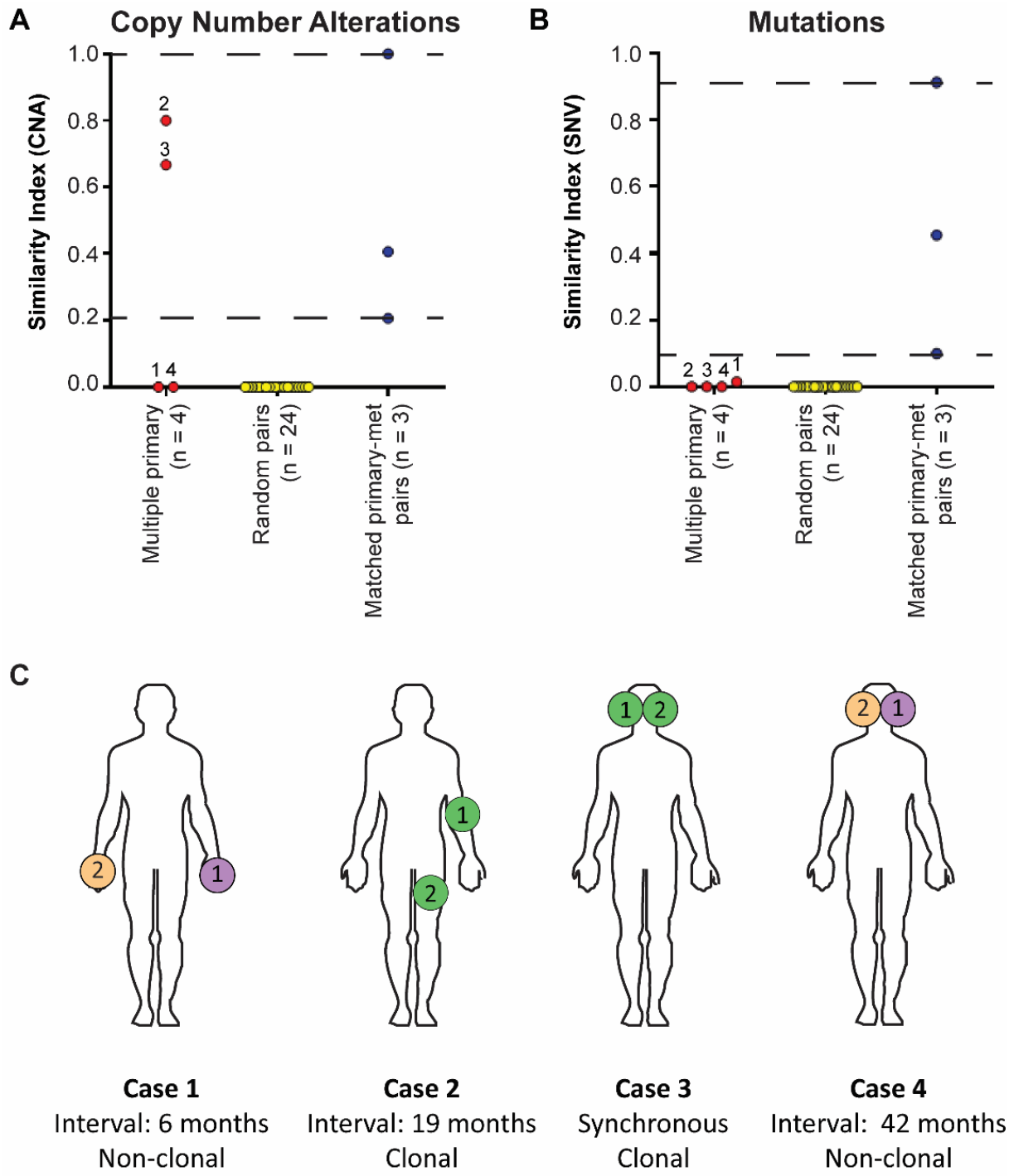

FIGURE 4.6 Clonality in clinically-designated multiple primary Merkel cell carcinomas.
(a) Similarity index analysis of copy number alterations (CNAs). Lack of shared CNAs supports independent primary MCC tumors in two cases (1 and 4). Shared CNAs support clonality in two cases (2 and 3). (b) Similarity index analysis of mutational events. Cases 1-4 lack significant overlap in shared mutational events. For (a) and (b), dashed lines indicate minimum and maximum degrees of clonal relatedness displayed by primary-metastasis pairs (positive controls for clonal relatedness). (c) Summary of findings. Considering CNA similarity indices, cases 1 and 4 are deemed non-clonal and cases 2 and 3 are deemed clonally related. Similarity index is calculated as shared events divided by the sum of shared and unique events for each pair.

4.2.4 Discussion

The clinical diagnosis of multiple primary MCC tumors is rare, but important for management and prognosis. Treatment of a second primary MCC includes excision and staging with sentinel lymph node biopsy, and prognosis depends upon sentinel lymph node status. Conversely, the diagnosis of a distant cutaneous metastasis is managed as stage IV disease. Here, we utilized NGS to analyze clonality in four cases of clinically-designated multiple primary MCC tumors. For maximal sensitivity, we compared tumor mutations, chromosomal copy number changes, and MCPyV sequence.

Multiple primary MCCs were verified in case 1 (second primary MCC arising on contralateral hand) and case 4 (second primary MCC arising on contralateral cheek) as the tumor pairs did not harbor similar copy number alterations or significant mutational overlap (Table 4.2). In case 4, tumors displayed distinct *TP53* and *RB1* mutations, consistent with the proposed role for inactivation of these tumor suppressor genes as driving, early events in MCPyV-negative MCC (264,283). Of note, in case 1, which was MCPyV-positive, a limited number of discordant CNAs drove the determination of multiple primary MCC status, as identified mutations in each sample were non-prioritized and subclonal based on variant allele frequency assessment.

Clonal relatedness was identified in case 2 (MCC on left elbow and subsequently on left thigh) and case 3 (MCCs on bilateral nasal alae). In case 2, both tumors shared multiple copy number changes that are predicted to be early events, compared to mutations arising later. Although speculative, we hypothesize that the second MCC tumor on the left thigh arose from hematogenous spread from the primary tumor or the untreated left axillary nodal metastasis and that the nodal recurrence in the left groin was from lymphatic drainage from the left thigh metastasis.

In case 3, both tumors displayed significant overlap in CNAs. Given the synchronous occurrence, these tumors likely represent a primary and in-transit metastasis pair. A less likely scenario is that both tumors represent in-transit metastases from a regressed midline primary tumor.

Rare cases of multiple primary MCCs have been reported (268-275). The designation of multiple primary MCC tumors should be assigned carefully. In our multidisciplinary MCC experience, some cases reported as multiple primary MCC tumors are more consistent with the presentation of in-transit metastatic disease, such as cases of multiple MCC tumors on the scalp

(268) and the ankle (269). Importantly, genetic relatedness has been evaluated by alternative technologies in three cases with findings similar to our current study (270,274,275). Sharma *et al.* confirmed the diagnosis of independent primary MCCs by sequence analysis of the MCPyV genomes in a patient with a MCC on the left upper arm who developed a second MCC on the right elbow six years later (270). Ahronowitz *et al.* demonstrated clonality by array CGH in a patient with a primary tumor on the right cheek and the second tumor on the left leg 2 months later (274). Nagy *et al*. used CGH to demonstrate shared and distinct molecular patterns in a primary MCC tumor on the lip and a second tumor on the palatine tonsil 7 years later (275). Nagy *et al.* conclude that the second tumor was an independent primary MCC with a field effect from the first tumor. Given the genetic overlap, we and others interpret these tumors to be clonally related (284).

MCPyV is found to be integrated in up to 80% of MCCs (26). Interestingly, we found that tumors from the same patient were consistently MCPyV-negative or MCPyV-positive, regardless of clonal relatedness. Although our cohort is small, this observation suggests that a given patient may be predisposed to development of either MCPyV-negative or MCPyV-positive MCC. Furthermore, MCPyV status alone is not a reliable indicator of clonal relatedness between MCC tumors.

This study has several limitations. The occurrence of clinically-designated multiple primary MCCs is exceedingly rare, which limits appropriate cases with tissue suitable for analysis. However, this series represents the largest to date, the first to use NGS for clonality analysis, and the only study to analyze clonality in multiple primary MCCs by multiple parameters in parallel: mutations, CNAs, and MCPyV sequence. Of these, we found CNA similarity to be a more consistent indicator of clonality than mutations in genetically related tumors, thus mutation analysis is less informative in this context. In particular, MCPyV-positive MCC has low mutational burden, with the vast majority of somatic mutations being subclonal, non-driving mutations, which may explain why examination of tumor mutations was not reliable for demonstrating clonality in these tumors. Unlike a previous study, we did not find MCPyV sequence analysis to be useful in evaluating clonal relatedness, as all MCPyV-positive tumors displayed high similarity in viral sequence. However, coverage of the MCPyV sequence was limited in many cases due to low remaining material, therefore we cannot exclude the possibility that more extensive analysis of viral sequences might be informative. Likewise, we only assessed a portion of the tumor genome, and hence are unable to assess fine subclonal structure in the related tumors. However, the purpose

of our study was to assess clonal relatedness between two apparent primary tumors, not detailed intra- or inter-tumoral heterogeneity.

4.2.5 Conclusions

Our findings show that patients with MCC may develop a second genetically distinct primary tumor, which is likely to develop through similar mechanisms of pathogenesis. Our study also supports clonality, and hence metastasis, in two cases of presumed multiple primary tumors. These findings underscore the challenge to correctly distinguish a second primary MCC from an isolated distant cutaneous metastasis, which has critically important prognostic and therapeutic implications. Our findings also support copy number analysis as more effective than mutational analysis for determining clonality in MCC, including MCPyV-positive tumors with low mutational burden. Given that clinicopathologic criteria may be imperfect, as seen in multiple primary lung carcinoma (285), copy number analysis by array comparative genomic hybridization or NGS may assist in the determination of clonality in clinically challenging cases.

4.3 Comprehensive Molecular Profiling of Olfactory Neuroblastoma Identifies Potentially Targetable *FGFR3* Amplifications

4.3.1 Introduction

Olfactory neuroblastomas (ONBs), also known as esthesioneuroblastomas, are malignant round-cell neuroectodermal tumors which represent up to 5% of sinonasal malignancies (135). Patients of all ages are affected (range 3-90 years) (136). Clinical course is variable, with some tumors displaying indolent behavior and others aggressively invading the intracranial cavity and/or displaying metastatic spread to lymph nodes, lung, or bone (135,136). 5-year survival is less than 50% for tumors with extension beyond the sinonasal cavity (136). Prognostic features are controversial (286), with patient age, tumor grade, presence of lymph node metastases, and *TP53* mutations reported as being associated with poor outcome (287). Critically, management options for aggressive disease is challenging due to a lack of recurrent, targetable oncogenic drivers, in part because molecular studies of ONBs have been limited (287).

Cytogenetic studies suggest ONBs are karyotypically complex neoplasms with chromosomal instability, however recurrent drivers or therapeutic targets have not been consistently nominated (287-293). More comprehensive molecular approaches, such as next-

generation sequencing (NGS), have only been reported for three cases of metastatic ONB (including one with matched primary tumor) (294-296). Hence, here we sought to comprehensively profile ONB through NGS to identify recurrent driving somatic alterations and potential therapeutic targets.

4.3.2 Methods

4.3.2.1 Case Selection

This study was conducted according to previously approved University of Michigan Institutional Review Board (previously only used IRB) Protocols. Retrospective ONB cases were identified by searching University of Michigan Department of Pathology databases for "olfactory neuroblastoma" or "esthesioneuroblastoma." Diagnosis was confirmed by board certified Anatomic Pathologists (PWH and JBM). Of 48 cases identified, 25 were excluded due to insufficient quality or quantity of material. Therefore, only 23 ONBs were sequenced.

4.3.2.2 Targeted multiplexed PCR based (mxPCR) NGS

An H&E stained section was used as a guide for dissection from a minimum of 4 formalin-fixed, paraffin-embedded (FFPE) 10-micron sections by a board-certified pathologist (P.W.H.) to obtain a minimal estimated tumor purity of 60%. The areas with highest tumor purity were macrodissected using a scalpel. DNA and RNA were co-isolated using the Qiagen Allprep FFPE DNA/RNA kit (Qiagen, Valencia, CA) and the Qiagen QIAcube (Qiagen, Valencia, CA), according to the manufacturer's instructions and were quantified using the Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA).

Targeted mxPCR based NGS was performed as previously described (99,101,297). Forty ng of DNA per sample was used to generate libraries using the Ion AmpliSeq Library Kit 2.0 (Life Technologies, Carlsbad, CA) and targeted multiplexed PCR with barcode incorporation using the Comprehensive Cancer Panel (CCP), which targets 1,688,650 bases from 15,992 amplicons representing 409 cancer genes (http://tools.invitrogen.com/downloads/cms_103573.csv). Template preparation was performed using the Ion PI-Hi-Q Template OT2 200 Kit (Life Technologies, Carlsbad, CA) and the Ion OneTouch ES Instrument (Life Technologies, Carlsbad, CA). NGS of multiplexed templates was then performed on Ion Proton P1 chips using the Ion P-HiQ Sequencing 200 Kit according to the manufacturer's instructions.

4.3.2.3 Variant Calling and Prioritization

Data analysis was performed using in-house developed, previously validated pipelines using Torrent Suite 4.0.2, with alignment by TMAP and variant calling using the Torrent Variant Caller plugin (298). Annotated variants were filtered to remove synonymous or non-coding variants, poorly supported calls/sequencing artifacts and germ line alterations. Any variant present in the 1000 Genomes (Phase II), Exome Sequencing Project callset, or the ExAC database (http://exac.broadinstitute.org) at population allele frequencies greater than 0.1% was considered germline and excluded. Additionally, variants that were present in ExAC with a variant fraction between 40 and 60% (or >90%) were also excluded unless occurring at a well-supported somatic mutation hotspot in COSMIC. All retained variants had flow corrected variant allele containing reads (FAO) counts ≥ 6 and flow corrected read depth (FDP) ≥ 20, with overall variant fraction (FAO/FDP) ≥ 10%. Additionally, any variants called in >4% of internally sequenced samples using the same panel and not reported in COSMIC as well as variants with extreme skewing of forward/reverse flow corrected reads (FSAF/FSAR <0.2 or >5) were removed. High confidence somatic variants passing the above criteria were then visualized using the Integrated Genomics Viewer (IGV) for read level confirmation. After filtering and visual confirmation, potential driving alterations were prioritized using the COSMIC Database, with prioritization of recurrently reported variants (>2 occurrences) in oncogenes, and recurrent or deleterious (non-sense, splice site, frame-shifting insertions/deletions [indels]) variants in tumor suppressors.

4.3.2.4 Copy number analysis

Normalized, GC-content corrected read counts per amplicon for each sample were divided by those from a pool of normal male genomic DNA samples (FFPE and frozen tissue, individual and pooled samples), yielding a copy number ratio for each amplicon. Gene-level copy number estimates were determined as described previously by taking the coverage-weighted mean of the per-probe ratios, with expected error determined by the probe-to-probe variance (146,297). Genes with a $\log_2$ copy number ratio estimate of <-1 or >0.80 were considered to have high level loss (deletion) and gain (amplification), respectively.

4.3.2.5 Mi-Oncoseq

Two ONB cases were also identified from the Mi-Oncoseq program at the University of Michigan, which performs comprehensive germline and somatic sequencing for adult patients with advanced cancers or unusual presentations to facilitate clinical trial enrollment and guide precision

medicine approaches. Clinical grade, hybrid capture-based exome or targeted (1,711 cancer related genes) sequencing of tumor and normal tissue to identify somatic mutations, fusions, and copy number alterations was performed as described (299). Potential driving somatic mutations were classified as for targeted sequencing described above.

4.3.2.6 Expression profiling by RNAseq

Targeted mxPCR based NGS of RNA was performed on co-isolated RNA from all 23 ONB samples assessed by mxPCR based NGS of DNA (ONB1-23) as previously described using a custom Ion Ampliseq panel assessing 103 target genes and 8 housekeeping genes relevant for urothelial carcinoma (101), except template preparation and sequencing was performed as described above for DNA mxPCR. Data analysis was performed essentially as described using the Coverage Analysis RNA Plug-in (101). Samples with low quality data (<150,000 total mapped reads or <50% full-length reads) were excluded from all further analyses. For high quality samples, for each amplicon, full length read counts were $\log_2$ transformed (read count + 1). Then, to determine normalized expression for each target gene, the $\log_2$ count was normalized to the median of the $\log_2$ counts (geometric mean) of five housekeeping genes expressed across normal urothelium, urothelial cancer and ONB (*CTCF*, *TARDBP*, *HNRNPK*, *TRHAP3*, *SAFB*). Data from previously sequenced normal urothelium (n=3) and urothelial carcinoma (n=11) (101) meeting the above quality criteria, as well as urothelial carcinoma with extensive squamous differentiation *(n=2*; Hovelson *et al*., manuscript in preparation) were processed exactly as for ONB samples. Unsupervised centroid linkage hierarchical clustering of genes and samples (with median centering of genes) was performed using Cluster 3.0 and visualized using Java TreeView. Median normalized expression of *FGFR3*, *FOXA1* (luminal subtype marker) and *KRT6A* (basal subtype marker) was compared across luminal subtype bladder cancers, basal subtype bladder cancers and ONB by the Kruskal-Wallis test with pairwise comparisons if results were statistically significant (p<0.05) using MedCalc v14.12.

4.3.2.7 Immunohistochemistry

Immunohistochemical staining for p53 and β-catenin was performed on a Ventana automated stainer**.**

4.3.3 Results

4.3.3.1 ONB cohort

We identified a cohort of 23 ONBs with sufficient available archived FFPE tissue for targeted mxPCR NGS. Seven and eight samples showed poor quality DNA and RNA sequencing data, respectively, and were excluded from subsequent analysis, resulting in a final cohort consisted of 18 tumors from unique patients. The mean patient age was 53 years (median 55, range 26-72 years), with samples from 14 men and 4 women. The tumors were mainly primary from the sinonasal region (*n*=9), however there were also recurrent (*n*=5), as well as locoregional metastases (4). There was a prior history of radiotherapy for 7 tumors (3 recurrences and 4 metastases). We also identified two ONBs profiled as part of the Mi-Oncoseq clinical sequencing program at the University of Michigan to identify precision medicine opportunities for patients with advanced cancer. Both samples were locoregional recurrences from male patients aged 40-59 who had been treated with radiotherapy (one with radiotherapy and chemotherapy [cisplatinum/etoposide]). Considering both archival and Mi-Oncoseq cases, the final cohort size was 20 tumors.

4.3.3.2 NGS demonstrates minimal recurrent driving somatic mutations in ONBs.

Targeted mxPCR NGS of DNA assessing the coding sequence of 409 cancer related genes from the 16 informative archived samples generated an average of 4,529,551 mapped reads yielding 263x targeted base coverage. After stringent filtering, we identified an average of 7 high confidence somatic mutations (range 0-21) and <1 high confidence prioritized driving somatic mutation (range 0-2) across the 16 samples**.** In the two Mi-Oncoseq profiled ONBs, we identified 16 (from >1,700 targeted genes) and 45 somatic mutations (from full exome capture), respectively, with no prioritized driving somatic mutations. All prioritized somatic mutations from the 18 total informative samples are shown in **Figure 4.7.**

No genes showed recurrent prioritized somatic mutations, and 6 of the 7 prioritized somatic mutations were in tumor suppressors. For example, ONB16 (a case from a patient who died of disease) harbored a *TP53* V272M mutation and a deleterious *PTEN* Y16X mutation (**Figure 4.7**). We confirmed the expected p53 protein over-expression in this case by immunohistochemistry. Likewise, we also identified inactivating mutations of the lysine methyltransferase genes *KMT2C* Y366X (MLL3) in ONB 15 and *KMT2A* S3702X (MLL1) in ONB6 (**Figure 4.7**). ONB6 harbored an activating *CTNNB1* S37F mutation, and we confirmed the expected nuclear beta catenin

expression by immunohistochemistry (**Figure 4.7).** Taken together, these results suggest that driving somatic mutations in known cancer associated genes are infrequent in ONB.

4.3.3.3 ONBs harbor highly recurrent chromosome/arm level copy number alterations and amplifications in FGFR3 and CCND1

Although driving somatic mutations were rare in our cohort, we found that all ONBs harbored multiple whole chromosome (or arm level) CNAs, including highly recurrent gains and losses (**Figure 4.8A and 4.8B**). Chromosome 20 gains were the most frequent CNA gains observed in our cohort, present in 16 of 18 (89%) ONBs. Gains of chromosomes 5, 7, and 11 were similarly common in our cohort (occurring in 78%, 61%, and 67% of tumors, respectively). Losses of chromosomes 12, 8, 3 and 1 were also frequently observed. In addition to these single-copy chromosome level gains and losses, we also identified recurrent low-level amplifications of known oncogenes in our ONB cohort (**Figures 4.7 and 4.8)**, including *CCND1* in 4/18 (22%) samples, respectively. Most intriguingly, we identified focal low-level amplifications of the targetable receptor tyrosine kinase *FGFR3* in 5/18 (28%) ONBs (**Figure 4.8**). Two samples harbored co-amplification of *CCND1* and *FGFR3* (**Figure 4.7 and 4.8C**), and *FGFR3* amplification was also observed in the absence of broader chromosome 4 gain (**Figure 4.8C**). Focal, high level deletions were rare in our cohort, with the only prioritized two-copy deletion in our cohort occurring in ONB2194, which harbored a focal, two copy deletion of *CDKN2A* (and *CDKN2B*), as shown in **Figure 4.8B**.

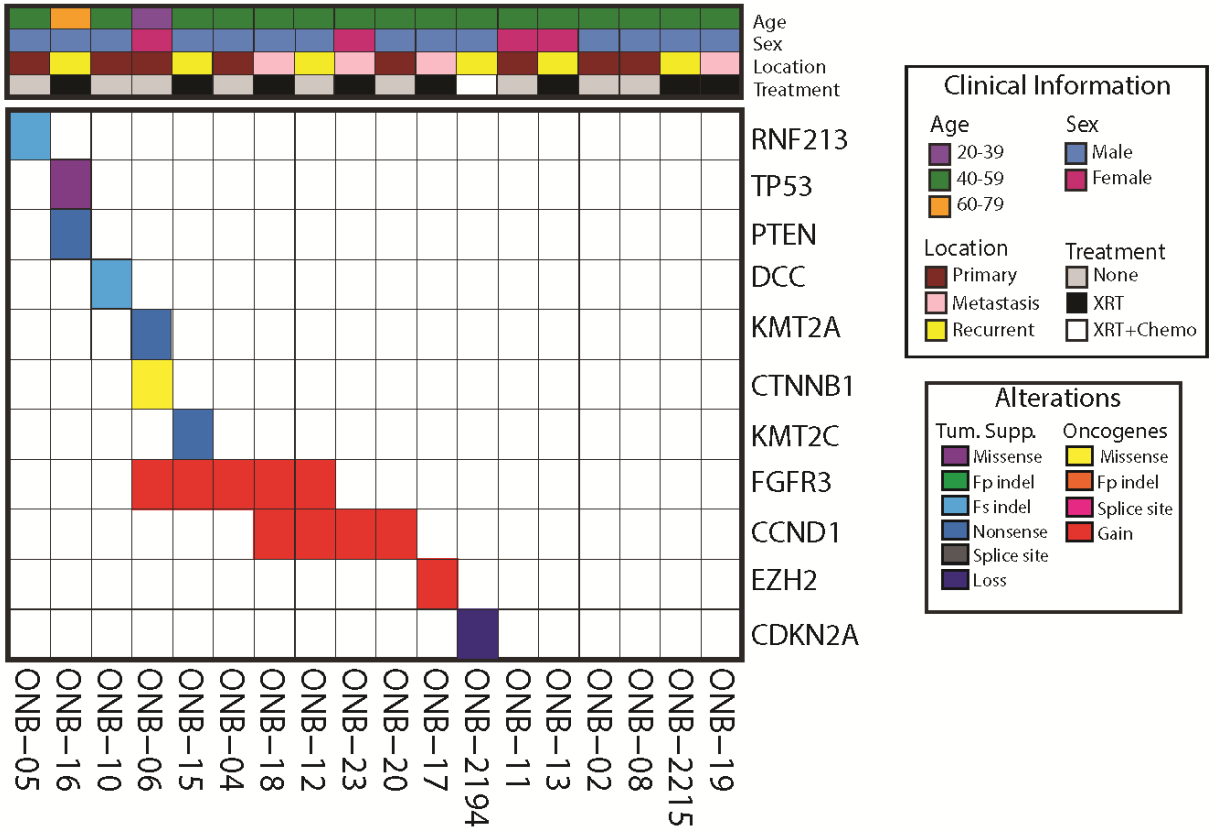

FIGURE 4.7 Integrated heatmap of prioritized mutations and copy number changes in olfactory neuroblastomas (ONB) identified by comprehensive next generation sequencing.
Integrated table of high-confidence, prioritized somatic mutations and copy number alterations (CNAs) identified by NGS of >400 cancer related genes across a cohort of 18 ONBs. Clinicopathological information for each sample is shown above the heatmap according to the legend (XRT = radiation therapy; Chemo = chemotherapy). All prioritized somatic mutations and high level CNAs are shown in the heatmap with alteration type indicated by cell color according to the legend. FP indel: frame-preserving insertion/deletion. FS indel: frameshift insertion/deletion. Loss: homozygous copy number loss. Gain: copy number gain.

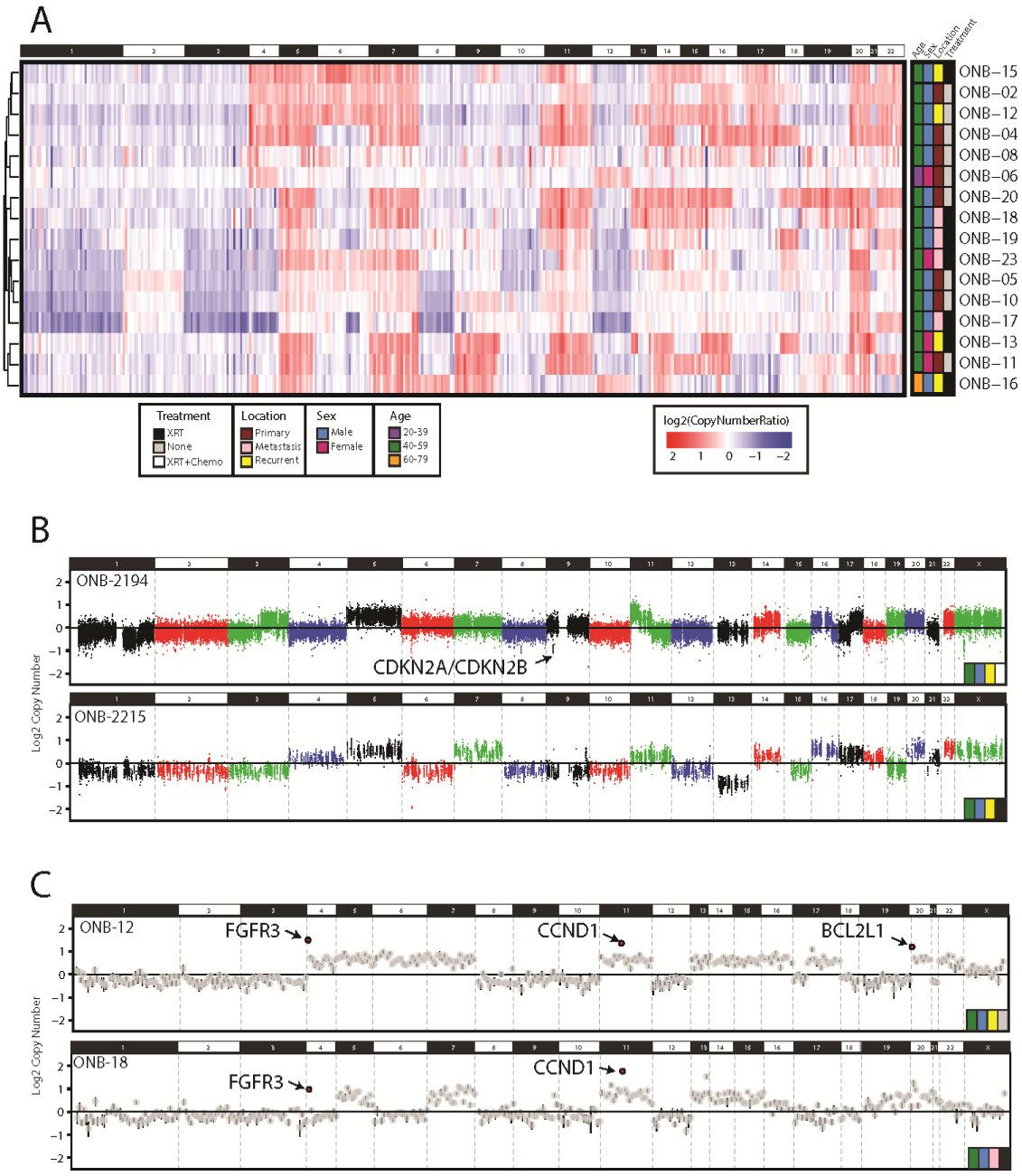

FIGURE 4.8 Recurrent copy number changes in ONBs include gains of chromosomes 5, 7, 11, and 20, and focal amplification of the potential therapeutic target *FGFR3.*

Genome wide copy number profiles were by multiplexed PCR (mxPCR) based NGS (**A**) or comprehensive hybridization-capture based NGS (**B**). Heatmap of autosomal, gene-level, log base 2 (Log_2) copy number ratios (vs. a composite normal sample) for 371 targeted genes. Genes are sorted in order of genomic position and samples are clustered by unsupervised hierarchical clustering. Gains and losses are indicated according to the color scale. Clinicopathological information for each sample is indicated. **B**. Genome-wide copy number profiles for two ONBs profiled by comprehensive hybrid capture based NGS (ONB2194 profiled by exome sequencing, ONB2215 profile by sequencing of over 1,700 cancer related genes). Individual exon-level probe Log_2 copy number ratio (vs. a matched normal sample) is plotted for each targeted gene, with data colored by chromosome. Prioritized high level amplifications and deletions are indicated. **C**. Recurrent amplifications in the potential therapeutic target *FGFR3* were identified in 5/18 profiled ONBs. Genome-wide autosomal copy number profiles for two samples with *FGFR3* amplifications are shown. Data is as in **A**, except Log_2 copy number ratios for each gene are plotted. High level amplifications and deletions are indicated.

4.3.3.4 Expression profiling of ONBs confirms high FGFR3 expression in amplified cases

To assess the expression of relevant genes in ONB, most notably *FGFR3,* we profiled co-isolated RNA from the 23 ONBs subjected to DNA mxPCR sequencing by targeted mxPCR RNAseq using a previously described custom Ampliseq panel assessing 103 target genes (and 8 housekeeping genes) relevant for urothelial carcinoma (101). Importantly, this panel assesses *FGFR3,* and would enable us to compare the expression profiles of our current ONB cohort to our previously profiled luminal (which is defined in part by high *FGFR3, FOXA1* and *KRT19* expression (300)) and basal subtypes of urothelial carcinoma (defined in part by high KRT6A and KRT14 expression (300)). Fifteen of the 23 ONB generated high quality mxPCR RNAseq sequencing results, with ONB04, ONB15, and ONB10 only having high quality DNA sequencing data, and ONB14 and ONB01 only having high quality RNA sequencing data. Across the 15 ONB samples, we generated an average of 2,593,832 total mapped reads and 2,132,101 full length reads (77%). Unsupervised hierarchical clustering of the high quality ONB samples and previously profiled urothelial samples shown unsurprisingly showed clear separation of ONB from all urothelial samples (n=3 normal urothelial samples, n=7 luminal bladder cancers, n=2 luminal/basal bladder cancers and n=4 basal bladder cancers), as shown in **Figure 4.9A**. Likewise, ONB showed essentially no expression of nearly all canonical luminal and basal subtype markers (**Fig 4.9A and 4.9B**). For example, normalized $\log_2$ expression of *FOXA1* (luminal marker) differed substantially across luminal bladder cancer, basal bladder cancer and ONB (Kruskal Wallis test p=0.0001, median $\log_2$ expression 3.4, -0.2 and -9.2, respectively; all pairwise comparisons p<0.05; **Fig 4.9B**), as did expression of the basal marker *KRT6A* (Kruskal Wallis test p=0.0002, median $\log_2$ expression 2.4, 10.8 and -5.6, respectively; all pairwise comparisons p<0.05; **Fig 4.9B**). Strikingly, however, ONB showed non-significantly greater *FGFR3* expression than luminal or basal bladder cancers (median $\log_2$ expression 0.1 vs. -1.2 and -1.4, respectively; Kruskal Wallis test p=0.21; **Fig 4.9B**), and the highest *FGFR3* expression in the cohort was observed in ONB12 (**Fig 4.9A**), which harbored a high level *FGFR3* amplification by DNA mxPCR based sequencing. Likewise, ONB18, which also had an *FGFR3* amplification by DNA sequencing, showed the fifth highest FGFR3 expression across the 31 total samples and had the second highest *FGFR3* expression in the ONB cohort (**Fig 4.9A**). Lastly, although *FGFR3* is activated in some tumor types by both hotspot mutations and gene fusions (301), we did not observe any activating mutations by DNA sequencing and the Mi-Oncoseq cases, which underwent capture transcriptome sequencing, did

not harbor any driver gene fusions (including those involving *FGFR3*). Taken together, these results demonstrate relatively high expression of *FGFR3* across ONB, and confirm that the highest expression is observed in cases with focal *FGFR3* amplification.

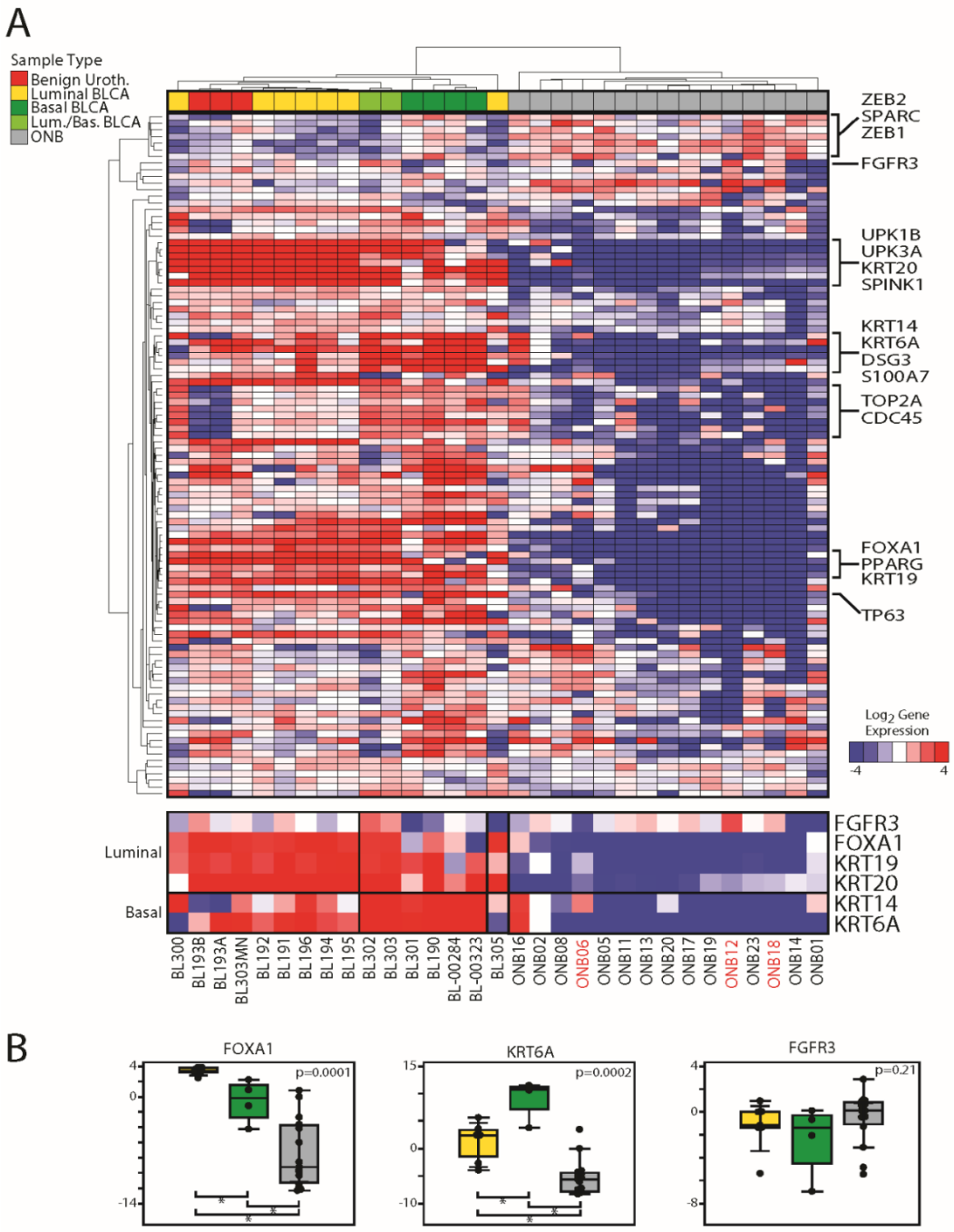

FIGURE 4.9 Targeted multiplexed PCR (mxPCR) based RNA sequencing (RNAseq) confirms relatively high expression of *FGFR3* in ONB and high level expression in *FGFR3* amplified samples.
RNA was co-isolated from all FFPE ONB samples subjected to mxPCR based DNA NGS (**Figures 1 & 2**), and 20ng was subjected to RNAseq using a custom AmpliSeq panel assessing 103 target genes (relevant to bladder cancer) and 8 housekeeping genes. **A**. Unsupervised hierarchical clustering using of target gene expression (median centered, log_2 normalized expression according to the color scale) for all high quality ONB samples (gray) and similarly processed previously profiled normal urothelium samples (red), luminal bladder cancer samples (yellow), basal bladder cancer samples (green) and luminal/basal bladder cancer samples (light green). Genes of interest are shown to the right of the heatmap. Expression of *FGFR3*, *FOXA1*, *KRT19* and *KRT20* (luminal subtype markers) and *KRT14* and *KRT6A* (basal markers) are shown below the heatmap. FGFR3 amplified ONB samples are named in red. **B**. Box and whisker plots of log_2 expression for all luminal bladder cancer samples (yellow), basal bladder cancer samples (green) and ONB

samples (gray) from **A** are shown for *FOXA1*, *KRT6A* and *FGFR3*. P values from Kruskal Wallis tests across the three classes are shown. For statistically significant distributions ($p<0.05$), all statistically significant ($p<0.05$) pairwise comparisons are indicated by *.

4.3.4 Discussion

Olfactory neuroblastomas are aggressive tumors for which accurate prognostication is challenging and targeted therapeutic strategies are lacking. Here, we utilized multiple NGS approaches on a cohort of 22 total ONBs to determine the genomic landscape, identify potential therapeutic strategies and performed limited transcriptional profiling to confirm DNA findings. We are aware of three case reports sequencing ONBs (3 metastatic ONBs, one with a matched primary). In the study of a matched primary and metastatic tumor, whole genome sequencing demonstrated a deleterious somatic *TP53* mutation, with additional non-hotspot mutations of unknown significance affecting *KDR* and *MYC*, among others, in the metastasis but not the primary (294). Whole exome sequencing of another case of metastatic ONB also demonstrated mutations in *TP53* and *CDKN2C*, as well as other genes (296). An additional exome sequencing study of one case of metastatic ONB found mutations including *EGFR* R521K (associated with response to cetuximab), as well as mutations of unknown significance in *KDR* and *FGFR2* (295). Our results are consistent with these findings, supporting limited recurrent somatic mutations in known cancer genes in ONBs. Importantly, we did identify *TP53* and *PTEN* mutations in one case with an aggressive course. However, the presence of *TP53* mutation is not likely to be a sensitive finding for poor outcome, as many tumors with aggressive course in our cohort were wild type for *TP53*, consistent with a previous targeted study of 18 ONBs (including aggressive cases) that failed to identify *TP53* mutations (302). Likewise, as we performed only targeted NGS of tumor tissue in the majority of our samples, our study was not designed to identify novel recurrently mutated genes in ONB or identify contributing germline alterations.

In contrast to low and variable mutation rates, ONBs displayed high rates of recurrence in copy number aberrations affecting whole chromosomes (5, 7, 11, 20) and focal amplifications including *CCND1* and *FGFR3* were identified in subsets of patients. In some tumors, copy number gain of an oncogene represented the only prioritized driver event. In others, copy gain of oncogenes was accompanied by deleterious mutation of tumor suppressor genes. These findings represent a major expansion upon previous knowledge of ONBs. Prior molecular analyses have predominantly focused on copy number alterations in ONB using lower resolution approaches (288-293). Our

findings agree with previous reports that ONBs are cytogenetically complex tumors harboring multiple chromosomal aberrations (typically arm/chromosome level events). Comparing previous studies, a consistent pattern of specific cytogenetic changes in ONB is not found (287). However, similar to our findings, gains have been reported to involve chromosome 11 (including *CCND1*)(288,292,293) and 20q (including *BCL2L1*) (288-290). Although gains of chromosome 4 have been reported previously (289,292), ours is the first to specifically nominate *FGFR3* as a recurrent oncogenic driver in a subset of ONBs. Importantly, multiple FGFR specific or non-selective tyrosine kinase inhibitors that inhibit FGFR are in various stages of clinical development (303), including ponatinib, a multi-tyrosine kinase inhibitor (including FGFR1-4) that is FDA approved for leukemia treatment and has shown potent pan-FGFR inhibition in preclinical models (304). Unfortunately, cell line models for ONB are not widely available to assess preclinical evidence for FGFR3 inhibition in this disease, however, if confirmed in other studies, our results suggest that a subset of patients with ONB requiring medical therapy may be candidates for ongoing basket studies evaluating FGFR inhibitors in patients with advanced solid tumors (e.g. NCT01752920 and NCT02272998, clinicaltrials.gov, accessed 12/7/16).

By targeted RNAseq (RNA-seq used throughout) using bladder cancer as a comparison (where *FGFR3* activation and expression is part of the defining signature distinguishing luminal vs. basal subtypes), we demonstrated that ONB had relatively high *FGFR3* expression (comparable to bladder cancer), but essentially lacked expression of all other luminal or basal markers. Importantly, 2 of the 3 *FGFR3* amplified ONBs with high quality RNAseq data had the highest *FGFR3* expression across our ONB cohort, further supporting the potential therapeutic relevance of *FGFR3* amplifications in this disease. Although the targeted RNAseq panel was designed to profile bladder cancer, limiting more general conclusions from our ONB cohort, we observed that ONB had very low expression of canonical cell cycle/proliferation genes (see cluster containing *TOP2A* and *CDC45* in **Figure 4.9A**), with relatively high expression of the "mesenchymal" type gene expression module (300) (see cluster containing *ZEB2, SPARC* and *ZEB1* in **Figure 4.9A)**. Lastly, in addition to amplification, *FGFR3* is known to be activated by gene fusions and mutations in other types (305). No ONB samples profiled herein harbored activating *FGFR3* mutations. Likewise, while our targeted RNAseq panel did not assess for the canonical *FGFR3-TACC3* and *FGFR3-BAIAP2L1* fusions seen in bladder cancer and other cancer types (306), representing a limitation of our study, comprehensive capture transcriptome

sequencing did not identify any driving genes fusions in the two samples (both of which lacked *FGFR3* amplifications) profiled by Mi-Oncoseq.

In summary, we report the largest NGS based comprehensive molecular profiling study of ONB to date. We found that ONBs are characterized by multiple chromosomal copy number alterations, including potential drivers such as *CCND1* amplification, as well as potentially targetable *FGFR3* amplifications. ONBs harbored a low mutational burden, with rare but heterogeneous deleterious tumor suppressor mutations and no highly recurrently mutated genes. Although our findings suggest that ONBs are largely copy number driven tumors with heterogeneity in deregulated pathways, a subset may be driven by potentially targetable *FGFR3* amplifications resulting in *FGFR3* over-expression.

4.4 Future Directions

Due to the rarity of the cancers discussed, we were primarily limited by sample size and availability of model systems to further investigate the genomic alterations identified. Therefore, as a way to improve our studies, we propose a multi-institutional study where sample processing and data analysis is consistent. This approach would allow us to have access to more samples and use the genomic data obtained at different locations to fully characterize each type of cancer. Additional sequencing of other rare cancers, such as choriocarcinomas, is also necessary since it can help guide patients toward a targeted therapy, increase the number of patients eligible for therapies in clinical trials, and expand our knowledge of cancer biology as it has been throughout the years (4,307).

Specifically, we propose to conduct RNA expression analysis to gain a deeper understanding of how different fusions affect the transcriptomic landscape since there were very little somatic events identified through targeted NGS. In this case, we would profile *CIC-DUX4* sarcomas and other sarcomas driven by gene fusions such as Ewing, synovial, and other undifferentiated small round cell sarcomas. In Merkel cell carcinomas, there are still questions concerning differences between virus positive (VP) and virus negative (VN) MCCs. Although MCC tumors involve cells that resemble sensory Merkel cells, MCC tumors are frequently found in the dermis but can also arise from any layer of the skin. In conjunction with the genetic differences between VP and VN MCCs we observe in our study, these tumors may have distinct cells of origin and therefore a different mechanism of tumor initiation. Therefore, identifying cell

of origin through genetic lineage marking studies and exploring the effect of the tumor microenvironment contributing to the pathogenesis of MCCs would be valuable and beneficial to modelling this disease (308). In ONBs, *FGFR3* amplification should be characterized in an ONB cell line; however, a cell line of this type does not currently exist. Future studies would ideally involve creating an ONB cell line or patient-derived xenograft with endogenous *FGFR3* amplification in which we could assess response to therapeutic targeting using FGFR inhibitors, such as ponatinib (304).

CHAPTER 5: Summary, Conclusions, and Precision Medicine

5.1 Summary and Conclusions

The overall objective of this dissertation was to answer important questions about cancer development and progression through characterizing a range of cancers, including rare cancers, by a validated targeted NGS approach on routine clinical FFPE tissues.

In LGECs, we performed multi-region, comprehensive somatic molecular profiling of routinely processed FFPE material from 13 cases of LGEC totaling 64 minute, spatially defined cell populations ranging from presumed precursor lesions through invasive LGEC. Shared driving *PTEN, PIK3R1,* or *PIK3CA* mutations support clonal origin of the samples in each case, except for two cases with clonally distinct neoplastic populations, consistent with unexpected multiclonality in LGEC development. Although substantial heterogeneity in driving somatic alterations was present across populations in nearly all cases, these alterations were usually clonal in a given population, supporting continued selection and clonal sweeping of driving alterations in populations with both precursor and LGEC histology. Importantly, *CTNNB1* mutational status, which has been proposed as both prognostic and predictive in LGEC (143,144), was frequently heterogeneous and subclonal, occurring both exclusively in precursor or cancer populations in different cases. Whole-transcriptome profiling of co-isolated RNA from 12 lesions was robust and confirmed histologic and molecular heterogeneity, including activated Wnt signaling in *CTNNB1*-mutant versus wild-type populations. Taken together, we demonstrate clinically relevant multiclonality and intratumoral heterogeneity during LGEC development with important implications for diagnosis, prognosis, and patient care

Additionally, we used squamous cell carcinomas as our second model characterizing precursor lesions. Here, we used high depth, targeted next-generation sequencing to characterize the somatic genomic landscape of 56 cutaneous lesions and 43 ocular surface lesions. Ocular and cutaneous squamous neoplasms displayed a predominance of UV signature mutations. Precursor

lesions had highly similar somatic genomic landscapes to invasive SCCs, including chromosomal gains of 3q involving *SOX2*, and highly recurrent mutation/loss of heterozygosity events affecting tumor suppressors *TP53* and *CDKN2A*. Among *TP53* wild-type tumors, human papillomavirus (HPV) transcript was detected in one matched pair of cutaneous CIS/invasive SCC. A subset of cutaneous CIS displayed an *RB1* mutation that was not observed in invasive SCC. Transcriptome analysis demonstrated significant upregulation of pro-invasive genes in invasive SCC relative to precursor lesions. Ocular and cutaneous neoplasms demonstrate similar alterations, supporting a common pathway for squamous neoplasia on UV-exposed epithelia. However, oncogene amplification appears more frequent in ocular lesions. Precursor lesions possess the full complement of major genetic changes seen in SCC. However, precursor and invasive SCC lesions demonstrate distinct transcriptome profiles, supporting non-genomic drivers of invasiveness.

We also performed comprehensive molecular profiling of three different types of rare cancers with little to no genomic characterization, *CIC-DUX4* sarcomas, MCCs, and ONBs. Although we did not identify recurrent somatic mutations in *CIC-DUX4* sarcomas, CN analysis showed recurrent, broad CN alterations, including gain of chromosome 8 and loss of 1p. In one sample pair (untreated primary and local recurrence resections), we identified similar CN profiles and a somatic *ARID1A* p.R963X nonsense mutation exclusively in the local recurrence sample. In another sample pair (pre- and post-radiation treatment specimens), we observed single-copy loss of chromosome 7q exclusively in the post-treatment recurrence sample, supporting it as an acquired event after radiation treatment. In the last sample pair (near-concurrent, post-chemotherapy primary and distant metastasis), molecular profiles were highly concordant, consistent with limited intertumoral heterogeneity. In summary, NGS identified limited somatic driver mutations in *CIC-DUX4* sarcomas. However, we identified novel, recurrent CNAs, including chromosome 1p, which is also the locus of *ARID1A*. In MCCs, mutations, CN alterations, and MCPyV sequence were analyzed and compared between clinically designated multiple primary MCCs to characterize genetic relatedness and hence assess clonality in 4 cases. Cases 1 and 4 were verified as genetically distinct primary tumors and did not harbor similar CNAs or demonstrate significant mutational overlap. Cases 2 and 3 were designated as clonally related based on overlapping CNAs. In clonally related tumors, chromosomal CN changes were more reliable than mutations for demonstrating clonality. Regardless of clonality, we found that MCPyV status was concordant for all tumor pairs and MCPyV positive tumors harbored predominantly

subclonal mutations. To identify potential oncogenic drivers and targetable pathways in ONBs, we characterized 20 ONBs. Somatic mutations were infrequent in our cohort, with 7 prioritized nonsynonymous mutations in 5 of 18 (28%) ONBs, and no genes were recurrently mutated. We detected arm/chromosome-level copy-number alterations in all tumors, most frequently gains involving all or part of chromosome 20, chromosome 5, and chromosome 11. Recurrent focal amplifications, often but not exclusively in the context of arm-level gains, included *CCND1* and the targetable receptor tyrosine kinase *FGFR3*. Targeted RNA NGS confirmed high expression of *FGFR3* in ONB (at levels equivalent to bladder cancer), with the highest expression observed in *FGFR3*-amplified ONB cases.

In summary, we characterized cancer progression in two cancer models which suggested that precursor and invasive lesions harbor similar genomic aberrations. Therefore, we hypothesize that invasion is most likely not driven by a series of genomic events but rather by downstream transcriptomic or epigenetic events. We have also addressed an area of unmet need by characterizing three different types of rare cancers addressing three major issues, including clonality assessment to correctly stage cancer, molecular characterization to improve diagnostic challenges, and identification of new cancers eligible for targeted therapies. We concluded that patients with MCC may develop a second genetically distinct primary tumor; in which the subsequent tumor is likely to develop through similar mechanisms of pathogenesis, either MCPyV-mediated or ultraviolet light-mediated. By distinguishing multiple primary MCCs from clonal MCC tumors clinically resembling multiple primaries, we are able to conduct appropriate staging of the patient to guide treatment. In ONBs, recurrent FGFR3 amplifications may suggest that a subset of patients with these aggressive tumors can be directed to targeted therapies, such as ponatinib. Although we did not identify targetable alterations in *CIC-DUX4* sarcomas, molecular profiling identified genomic aberrations not previously seen in Ewing sarcomas. Therefore, this cohort showed that genomic characterization can improve diagnoses in cancers that resemble each other. Taken together, this dissertation has utilized the power of targeted NGS to gain a deep knowledge of the somatic genomic aberrations involved in cancer progression and rare cancers to improve diagnoses and cancer therapies.

Since the main focus of the projects was on the somatic genomic landscape of selected cancer types, we propose incorporating adjacent normal tissue as a way to expand our studies to have a more complete analysis of genomic aberrations that lead up to invasive cancer. Studies on

normal skin show *TP53* mutations are present at very low variant allele frequencies (VAFs) (214). Therefore, normal tissue would allow us to identify the level of genetic variation found in the normal skin and allow us to compare it to its precursor and invasive counterparts. By assessing the VAFs and number of *TP53* mutations of adjacent normal tissue in our CSCC study, we would have a better understanding on the level of UV damage the tissue was exposed to before certain cell populations became malignant. Similarly, in our endometrial study, we would add not only adjacent normal samples but also a greater spectrum of precursor lesions which would include premalignant lesions from simple hyperplasia to complex atypical hyperplasia.

Additionally, due to studies suggesting that somatic mutations in the setting of germline variants are more likely to increase the likelihood of developing cancer and are therefore predisposing a population to cancer (29,309), we believe that incorporating normal tissue would give us a more comprehensive view of the genomic aberrations leading to each cancer. While we've used filters and databases such as EXAC (310), the 1000 Genome project (311), and dbSNP (312) to remove germline mutations, differentiating between a germline and a somatic mutation can be challenging. In cases with a ~50% tumor content, homozygous mutations could be misinterpreted as a heterozygous germline mutation. The incorporation of normal would allow us to call germline variants and somatic mutations more easily, and potentially more accurately, particularly outside of well characterized cancer driver genes. We would also be able to identify rare germline variants, and/or incorporate germline and somatic variants to our analysis to determine whether there are any recurrent germline alterations that can potentially predispose to the development of the cancers studied, particularly in our rare cohort since there are very limited studies pertaining to these cancers. Specifically, it would be necessary to report germline variants in the case it has therapeutic implications, as has been suggested with *BRCA1/2* germline variants and PARP inhibitors (313,314). Therefore, we believe that by sequencing normal tissue and having access to germline variants and potentially even germline CNVs, we could gain a deeper knowledge of the genomic variations found in the cancers we profiled.

5.2 Early Detection, Diagnostics, and Resistance

Although individual future directions have been addressed at the end of each Chapter, here is a brief discussion of future applications of NGS to advance the field of precision oncology.

Cancer has been proposed to arise from a series of events in cancer-related genes that can occur throughout decades. Although many early-stage and localized cancers can be successfully treated with surgery, most cancers are not detected throughout the first 90% of its lifetime when potentially treatable. Instead, many cancers are detected once metastasis develops, which is usually a few years before death (72,73). Therefore, most cancer-related deaths are due to metastasis that already occurred before the onset of symptoms. Now, there is a shift of focus from treatment of late disease to prevention and early detection, which is predicted to reduce cancer deaths by more than 75% (315). Ideally, the goal is to take advantage of the large window between cancer initiation and progression to late-stage disease through the development of new minimally invasive detection methods (73,83).

Although there are established screening methods such as Pap smears, colonoscopies, mammograms, and PSA screening for cervical cancer, colon cancer, breast, and prostate cancer, respectively, there is still much room for improvement. Pa[tests have enabled early detection of cervical cancer and dramatically decreased mortality rate (316). However, it cannot detect other cancers such as endometrial and ovarian cancer, which account for a total of about 25,000 deaths each year and are together the third leading cause of cancer-related mortality in women in the United States (317). The transvaginal ultrasound (TVUS) offers a way to check for gynecologic abnormalities. However, this method is mostly ineffective and cannot distinguish between benign and malignant lesions which can therefore lead to over-diagnosis and overtreatment, a major limitation of early detection methods (318-320).

The longitudinal deep characterization of precursor lesions via the PCGA will potentially allow us to gain in-depth knowledge of the genetic events involved in cancer initiation, progression, and maintenance. Additionally, we would view genomic variation in the context of organization and migration of tumor clones as they invade the surrounding tissues through spatial analysis. Following comprehensive characterization of precursor lesions from all tissue types, the PCGA hopes to improve and/or develop early detection methods, as well as provide strategies to accurately assess the likelihood of a precancerous lesion to progress versus regress (83,89). Therefore, we support PCGA's proposal to continue assessing pre-cancer in other models to facilitate the implementation of these early detection methods.

While the PCGA will lead to a new era focused on early detection and prevention, current trials are still making huge strides in precision oncology. These trials include the National Cancer

Institute Molecular Analysis for Therapy of Choice (NCI-MATCH) (321) and the American Society of Clinical Oncology Targeted Agent and Profiling Utilization Registry (ASCO TAPUR) (322). In addition to these efforts, Memorial's MSK-IMPACT FDA-approved targeted panel will also have huge implications to the field of precision oncology. This panel will guide treatment decisions, help assess patient response to therapy, and identify genetic aberrations linked to relapse (129). Importantly, in March 2018, Medicare beneficiaries with solid tumors became eligible for NGS testing with FoundationOne CDx™, an FDA-approved assay for all solid tumors targeting select genes, gene rearrangements, and genomic signatures such as MSI and mutation burden. Now, Medicare patients with recurrent, relapsed, refractory, metastatic, or advanced cancer can receive testing at no cost (323).

Current efforts are also focused on developing targeted disease specific panels, especially after the completion of TCGA. An area of personal interest that represents a less well studied area is in pediatric cancers. In fact, the pediatric cancer genome landscape is currently being characterized through large sequencing initiatives including BASIC3 (324), PEDS-MIONCOSEQ (325), iCAT (326), and INFORM (327), which focus on profiling pediatric cancers primarily using WES platforms. All four studies together sequenced approximately 500 patients by 2016, which, in comparison to the 11,000 samples conducted by TCGA (43), is very low. Unfortunately, 89 of those patients profiled by iCAT used two targeted panels (326), OncoMap and OncoPanel, that are not pediatric specific. Therefore, the patients sequenced under this study may have had alterations that were missed since the panel did not capture the full spectrum of genomic alterations potentially unique to driving pediatric cancers. Although additional studies have also used targeted sequencing panels and found it very useful in identifying potential targetable alterations (118,328-330), their methods do not completely characterize pediatric cancers. For example, one study compiled data of recurrently altered genes in pediatric cancers and designed a panel specifically to guide clinical management of children with solid tumors (331). However, they were unable to completely validate their detection methods for gene fusions, which are found in most pediatric cancers, and also failed to include genes previously identified in rare cancers, such as germline *PMS2* mutations in duodenal cancer and mutations in MMR genes (332,333). Following a similar pattern to what is suggested for adult cancers, we propose using a recently developed pediatric panel, OncoKids (334), to identify recurrent genes important in the brain, solid tumors, and liquid cancers to generate even more cost-efficient targeted panels. Additionally, we would be able to generate

molecular subgroups that can help guide treatment, as seen in the four groups of medulloblastoma (335).

Overall, this work shows the power of using NGS-based approaches compatible with FFPE tissue for molecular profiling of minute precursor lesions and rare cancers. NGS provided us with multiple tools for charactering a wide range of cancers and for answering clinically relevant questions demonstrating the value of molecular profiling that can ultimately be applied to routine patient care.

References

1. Global Burden of Disease Cancer C, Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, *et al.* Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. JAMA oncology **2017**;3(4):524-48
2. Tabin CJ, Bradley SM, Bargmann CI, Weinberg RA, Papageorge AG, Scolnick EM, *et al.* Mechanism of activation of a human oncogene. Nature **1982**;300(5888):143-9
3. Reddy EP, Reynolds RK, Santos E, Barbacid M. A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. Nature **1982**;300(5888):149-52
4. Kipling MD, Waldron HA. Percivall Pott and cancer scroti. British journal of industrial medicine **1975**;32(3):244-6
5. Boffetta P, Hashibe M. Alcohol and cancer. The Lancet Oncology **2006**;7(2):149-56
6. Wogan GN. Impacts of chemicals on liver cancer risk. Seminars in cancer biology **2000**;10(3):201-10
7. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, *et al.* Mutational signatures associated with tobacco smoking in human cancer. Science **2016**;354(6312):618-22
8. . How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General, Publications and Reports of the Surgeon General. Atlanta (GA)2010.
9. Sinha RP, Hader DP. UV-induced DNA damage and repair: a review. Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology **2002**;1(4):225-36
10. zur Hausen H. Viruses in human cancers. Science **1991**;254(5035):1167-73
11. Sambrook J, Botchan M, Hu SL, Mitchison T, Stringer J. Integration of viral DNA sequences in cells transformed by adenovirus 2 or SV40. Proceedings of the Royal Society of London Series B, Biological sciences **1980**;210(1180):423-35
12. Gallo RC, Kalyanaraman VS, Sarngadharan MG, Sliski A, Vonderheid EC, Maeda M, *et al.* Association of the human type C retrovirus with a subset of adult T-cell cancers. Cancer research **1983**;43(8):3892-9
13. Fattovich G, Giustina G, Degos F, Tremolada F, Diodati G, Almasio P, *et al.* Morbidity and mortality in compensated cirrhosis type C: a retrospective follow-up study of 384 patients. Gastroenterology **1997**;112(2):463-72
14. Eichten A, Westfall M, Pietenpol JA, Munger K. Stabilization and functional impairment of the tumor suppressor p53 by the human papillomavirus type 16 E7 oncoprotein. Virology **2002**;295(1):74-85

15. Gonzalez SL, Stremlau M, He X, Basile JR, Munger K. Degradation of the retinoblastoma tumor suppressor by the human papillomavirus type 16 E7 oncoprotein is important for functional inactivation and is separable from proteasomal degradation of E7. Journal of virology **2001**;75(16):7583-91
16. Karagas MR, Nelson HH, Sehr P, Waterboer T, Stukel TA, Andrew A, *et al.* Human papillomavirus infection and incidence of squamous cell and basal cell carcinomas of the skin. Journal of the National Cancer Institute **2006**;98(6):389-95
17. Brandsma JL, Abramson AL. Association of papillomavirus with cancers of the head and neck. Archives of otolaryngology--head & neck surgery **1989**;115(5):621-5
18. de Sanjose S, Quint WG, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, *et al.* Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. The Lancet Oncology **2010**;11(11):1048-56
19. Durst M, Gissmann L, Ikenberg H, zur Hausen H. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. Proceedings of the National Academy of Sciences of the United States of America **1983**;80(12):3812-5
20. Boshart M, Gissmann L, Ikenberg H, Kleinheinz A, Scheurlen W, zur Hausen H. A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer. The EMBO journal **1984**;3(5):1151-7
21. Merck. 2017 Gardasil 9. <https://www.gardasil9.com/>. Accessed 2019.
22. Ko YH. EBV and human cancer. Experimental & molecular medicine **2015**;47:e130
23. Sunil M, Reid E, Lechowicz MJ. Update on HHV-8-Associated Malignancies. Current infectious disease reports **2010**;12(2):147-54
24. Tan YJ. Hepatitis B virus infection and the risk of hepatocellular carcinoma. World journal of gastroenterology **2011**;17(44):4853-7
25. Agelli M, Clegg LX, Becker JC, Rollison DE. The etiology and epidemiology of merkel cell carcinoma. Current problems in cancer **2010**;34(1):14-37
26. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science **2008**;319(5866):1096-100
27. Borchert S, Czech-Sioli M, Neumann F, Schmidt C, Wimmer P, Dobner T, *et al.* High-affinity Rb binding, p53 inhibition, subcellular localization, and transformation by wild-type or tumor-derived shortened Merkel cell polyomavirus large T antigens. Journal of virology **2014**;88(6):3144-60
28. Paulson KG, Park SY, Vandeven NA, Lachance K, Thomas H, Chapuis AG, *et al.* Merkel cell carcinoma: Current US incidence and projected increases based on changing demographics. J Am Acad Dermatol **2018**;78(3):457-63 e2
29. Huang KL, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, *et al.* Pathogenic Germline Variants in 10,389 Adult Cancers. Cell **2018**;173(2):355-70 e14
30. Lu C, Xie M, Wendl MC, Wang J, McLellan MD, Leiserson MD, *et al.* Patterns and functional implications of rare germline variants across 12 cancer types. Nat Commun **2015**;6:10086
31. Southey MC, Goldgar DE, Winqvist R, Pylkas K, Couch F, Tischkowitz M, *et al.* PALB2, CHEK2 and ATM rare variants and cancer risk: data from COGS. J Med Genet **2016**;53(12):800-11

32. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, *et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. The New England journal of medicine **2015**;373(24):2336-46
33. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Blom MJ, *et al.* Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. Jama **2017**;317(23):2402-16
34. Levy-Lahad E, Friedman E. Cancer risks among BRCA1 and BRCA2 mutation carriers. British journal of cancer **2007**;96(1):11-5
35. Ferrone CR, Levine DA, Tang LH, Allen PJ, Jarnagin W, Brennan MF, *et al.* BRCA germline mutations in Jewish patients with pancreatic adenocarcinoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology **2009**;27(3):433-8
36. Force USPST. Risk Assessment, Genetic Counseling, and Genetic Testing for BRCA-Related Cancer in Women: Recommendation Statement. American family physician **2015**;91(2):Online
37. 2019 Evaluating the Utility of Genetic Panels. <http://blue.regence.com/trgmedpol/geneticTesting/index.html>. Accessed 2019.
38. Lynch HT, Shaw TG, Lynch JF. Inherited predisposition to cancer: a historical overview. American journal of medical genetics Part C, Seminars in medical genetics **2004**;129C(1):5-22
39. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, *et al.* Large-scale copy number polymorphism in the human genome. Science **2004**;305(5683):525-8
40. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, *et al.* Global variation in copy number in the human genome. Nature **2006**;444(7118):444-54
41. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, *et al.* Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. The New England journal of medicine **2001**;344(11):783-92
42. Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature **2008**;455(7216):1061-8
43. The future of cancer genomics. Nature medicine **2015**;21(2):99
44. International Cancer Genome C, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, *et al.* International network of cancer genome projects. Nature **2010**;464(7291):993-8
45. Nowell PC, Hungerford DA. Chromosome studies on normal and leukemic human leukocytes. Journal of the National Cancer Institute **1960**;25:85-109
46. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature **1973**;243(5405):290-3
47. Druker BJ. Translation of the Philadelphia chromosome into therapy for CML. Blood **2008**;112(13):4808-17
48. Rampal R, Levine RL. Leveraging cancer genome information in hematologic malignancies. Journal of clinical oncology : official journal of the American Society of Clinical Oncology **2013**;31(15):1885-92
49. Rowley JD. Chromosomal translocations: revisited yet again. Blood **2008**;112(6):2183-9
50. Tomlins SA, Mehra R, Rhodes DR, Smith LR, Roulston D, Helgeson BE, *et al.* TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. Cancer research **2006**;66(7):3396-400

51. Kumar-Sinha C, Tomlins SA, Chinnaiyan AM. Recurrent gene fusions in prostate cancer. Nat Rev Cancer **2008**;8(7):497-511
52. Wells SA, Jr., Santoro M. Targeting the RET pathway in thyroid cancer. Clinical cancer research : an official journal of the American Association for Cancer Research **2009**;15(23):7119-23
53. Kauffman EC, Ricketts CJ, Rais-Bahrami S, Yang Y, Merino MJ, Bottaro DP, *et al.* Molecular genetics and cellular features of TFE3 and TFEB fusion kidney cancers. Nature reviews Urology **2014**;11(8):465-75
54. Delattre O, Zucman J, Plougastel B, Desmaze C, Melot T, Peter M, *et al.* Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature **1992**;359(6391):162-5
55. Yoshimoto M, Graham C, Chilton-MacNeill S, Lee E, Shago M, Squire J, *et al.* Detailed cytogenetic and array analysis of pediatric primitive sarcomas reveals a recurrent CIC-DUX4 fusion gene event. Cancer genetics and cytogenetics **2009**;195(1):1-11
56. Ried T, Knutzen R, Steinbeck R, Blegen H, Schrock E, Heselmeyer K, *et al.* Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. Genes Chromosomes Cancer **1996**;15(4):234-45
57. Firestein R, Bass AJ, Kim SY, Dunn IF, Silver SJ, Guney I, *et al.* CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. Nature **2008**;455(7212):547-51
58. Camps J, Pitt JJ, Emons G, Hummon AB, Case CM, Grade M, *et al.* Genetic amplification of the NOTCH modulator LNX2 upregulates the WNT/beta-catenin pathway in colorectal cancer. Cancer research **2013**;73(6):2003-13
59. Crago AM, Singer S. Clinical and molecular approaches to well differentiated and dedifferentiated liposarcoma. Current opinion in oncology **2011**;23(4):373-8
60. Fridlyand J, Snijders AM, Ylstra B, Li H, Olshen A, Segraves R, *et al.* Breast tumor copy number aberration phenotypes and genomic instability. BMC cancer **2006**;6:96
61. Ozenne P, Eymin B, Brambilla E, Gazzeri S. The ARF tumor suppressor: structure, functions and status in cancer. International journal of cancer **2010**;127(10):2239-47
62. Manning AL, Dyson NJ. pRB, a tumor suppressor with a stabilizing presence. Trends in cell biology **2011**;21(8):433-41
63. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, *et al.* Signatures of mutational processes in human cancer. Nature **2013**;500(7463):415-21
64. Alexandrov LB, Stratton MR. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. Current opinion in genetics & development **2014**;24:52-60
65. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. Nature reviews Genetics **2014**;15(9):585-98
66. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science **2017**;357(6349):409-13
67. Johanns TM, Miller CA, Dorward IG, Tsien C, Chang E, Perry A, *et al.* Immunogenomics of Hypermutated Glioblastoma: A Patient with Germline POLE Deficiency Treated with Checkpoint Blockade Immunotherapy. Cancer Discov **2016**;6(11):1230-6
68. Knudson AG. Two genetic hits (more or less) to cancer. Nat Rev Cancer **2001**;1(2):157-62
69. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, *et al.* The genomic landscapes of human breast and colorectal cancers. Science **2007**;318(5853):1108-13

70. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, *et al.* The landscape of cancer genes and mutational processes in breast cancer. Nature **2012**;486(7403):400-4
71. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature **2012**;490(7418):61-70
72. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. Science **2013**;339(6127):1546-58
73. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature **2010**;467(7319):1114-7
74. Bashashati A, Ha G, Tone A, Ding J, Prentice LM, Roth A, *et al.* Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. The Journal of pathology **2013**;231(1):21-34
75. Thirlwell C, Will OC, Domingo E, Graham TA, McDonald SA, Oukrif D, *et al.* Clonality assessment and clonal ordering of individual neoplastic crypts shows polyclonality of colorectal adenomas. Gastroenterology **2010**;138(4):1441-54, 54 e1-7
76. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. The New England journal of medicine **2012**;366(10):883-92
77. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, *et al.* Tumour evolution inferred by single-cell sequencing. Nature **2011**;472(7341):90-4
78. Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, *et al.* Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology **2008**;26(10):1626-34
79. Karapetis CS, Khambata-Ford S, Jonker DJ, O'Callaghan CJ, Tu D, Tebbutt NC, *et al.* K-ras mutations and benefit from cetuximab in advanced colorectal cancer. The New England journal of medicine **2008**;359(17):1757-65
80. Misale S, Yaeger R, Hobor S, Scala E, Janakiraman M, Liska D, *et al.* Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. Nature **2012**;486(7404):532-6
81. McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. Science translational medicine **2015**;7(283):283ra54
82. Srivastava S, Ghosh S, Kagan J, Mazurchuk R, National Cancer Institute's HI. The Making of a PreCancer Atlas: Promises, Challenges, and Opportunities. Trends Cancer **2018**;4(8):523-36
83. Campbell JD, Mazzilli SA, Reid ME, Dhillon SS, Platero S, Beane J, *et al.* The Case for a Pre-Cancer Genome Atlas (PCGA). Cancer Prev Res (Phila) **2016**;9(2):119-24
84. Trimble CL, Kauderer J, Zaino R, Silverberg S, Lim PC, Burke JJ, 2nd, *et al.* Concurrent endometrial carcinoma in women with a biopsy diagnosis of atypical endometrial hyperplasia: a Gynecologic Oncology Group study. Cancer **2006**;106(4):812-9
85. Zaino RJ, Kauderer J, Trimble CL, Silverberg SG, Curtin JP, Lim PC, *et al.* Reproducibility of the diagnosis of atypical endometrial hyperplasia: a Gynecologic Oncology Group study. Cancer **2006**;106(4):804-11

86. Stachler MD, Taylor-Weiner A, Peng S, McKenna A, Agoston AT, Odze RD, *et al.* Paired exome analysis of Barrett's esophagus and adenocarcinoma. Nature genetics **2015**;47(9):1047-55
87. Shain AH, Yeh I, Kovalyshyn I, Sriharan A, Talevich E, Gagnon A, *et al.* The Genetic Evolution of Melanoma from Precursor Lesions. The New England journal of medicine **2015**;373(20):1926-36
88. Teixeira VH, Pipinikas CP, Pennycuick A, Lee-Six H, Chandrasekharan D, Beane J, *et al.* Deciphering the genomic, epigenomic, and transcriptomic landscapes of pre-invasive lung cancer lesions. Nature medicine **2019**;25(3):517-25
89. Srivastava S, Ghosh S, Kagan J, Mazurchuk R. The PreCancer Atlas (PCA). Trends Cancer **2018**;4(8):513-4
90. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell **2000**;100(1):57-70
91. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol **1975**;94(3):441-8
92. Robertson AG, Kim J, Al-Ahmadie H, Bellmunt J, Guo G, Cherniack AD, *et al.* Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. Cell **2017**;171(3):540-56 e25
93. Cancer Genome Atlas Research Network. Electronic address edsc, Cancer Genome Atlas Research N. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. Cell **2017**;171(4):950-65 e28
94. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, *et al.* An integrated semiconductor device enabling non-optical genome sequencing. Nature **2011**;475(7356):348-52
95. Ion Torrent Next-Generation Sequencing Technology.
96. 2019. Illumina Sequencing Methods. <https://www.illumina.com/techniques/sequencing.html>. 2019.
97. Lahens NF, Ricciotti E, Smirnova O, Toorens E, Kim EJ, Baruzzo G, *et al.* A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. BMC Genomics **2017**;18(1):602
98. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics **2012**;13:341
99. Hovelson DH, McDaniel AS, Cani AK, Johnson B, Rhodes K, Williams PD, *et al.* Development and validation of a scalable next-generation sequencing system for assessing relevant somatic variants in solid tumors. Neoplasia **2015**;17(4):385-99
100. McDaniel AS, Stall JN, Hovelson DH, Cani AK, Liu CJ, Tomlins SA, *et al.* Next-Generation Sequencing of Tubal Intraepithelial Carcinomas. JAMA oncology **2015**;1(8):1128-32
101. Warrick JI, Hovelson DH, Amin A, Liu CJ, Cani AK, McDaniel AS, *et al.* Tumor evolution and progression in multifocal and paired non-invasive/invasive urothelial carcinoma. Virchows Archiv : an international journal of pathology **2015**;466(3):297-311
102. Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, *et al.* Integrated genomic characterization of endometrial carcinoma. Nature **2013**;497(7447):67-73

103. Campbell JD, Alexandrov A, Kim J, Wala J, Berger AH, Pedamallu CS, *et al.* Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nature genetics **2016**;48(6):607-16
104. DeSantis CE, Kramer JL, Jemal A. The burden of rare cancers in the United States. CA: a cancer journal for clinicians **2017**;67(4):261-72
105. Greenlee RT, Goodman MT, Lynch CF, Platz CE, Havener LA, Howe HL. The occurrence of rare cancers in U.S. adults, 1995-2004. Public health reports **2010**;125(1):28-43
106. Gatta G, van der Zwan JM, Casali PG, Siesling S, Dei Tos AP, Kunkler I, *et al.* Rare cancers are not so rare: the rare cancer burden in Europe. European journal of cancer **2011**;47(17):2493-511
107. Lu E, Shatzel J, Shin F, Prasad V. What constitutes an "unmet medical need" in oncology? An empirical evaluation of author usage in the biomedical literature. Seminars in oncology **2017**;44(1):8-12
108. Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM, *et al.* A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature **1986**;323(6089):643-6
109. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proceedings of the National Academy of Sciences of the United States of America **1971**;68(4):820-3
110. Budinska E, Popovici V, Tejpar S, D'Ario G, Lapique N, Sikora KO, *et al.* Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. The Journal of pathology **2013**;231(1):63-76
111. Jamshidi F, Nielsen TO, Huntsman DG. Cancer genomics: why rare is valuable. Journal of molecular medicine **2015**;93(4):369-81
112. Shah SP, Kobel M, Senz J, Morin RD, Clarke BA, Wiegand KC, *et al.* Mutation of FOXL2 in granulosa-cell tumors of the ovary. The New England journal of medicine **2009**;360(26):2719-29
113. Pisarska MD, Bae J, Klein C, Hsueh AJ. Forkhead l2 is expressed in the ovary and represses the promoter activity of the steroidogenic acute regulatory gene. Endocrinology **2004**;145(7):3424-33
114. Garralda E, Dienstmann R, Piris-Gimenez A, Brana I, Rodon J, Tabernero J. New clinical trial designs in the era of precision medicine. Molecular oncology **2019**;13(3):549-57
115. Rodriguez-Freixinos V, Lheureux S, Mandilaras V, Clarke B, Dhani NC, Mackay H, *et al.* Impact of somatic molecular profiling on clinical trial outcomes in rare epithelial gynecologic cancer patients. Gynecologic oncology **2019**
116. Pierron G, Tirode F, Lucchesi C, Reynaud S, Ballet S, Cohen-Gogo S, *et al.* A new subtype of bone sarcoma defined by BCOR-CCNB3 gene fusion. Nature genetics **2012**;44(4):461-6
117. Matos LL, Trufelli DC, de Matos MG, da Silva Pinhal MA. Immunohistochemistry as an important tool in biomarkers detection and clinical practice. Biomark Insights **2010**;5:9-20
118. Carter JH, McNulty SN, Cimino PJ, Cottrell CE, Heusel JW, Vigh-Conrad KA, *et al.* Targeted Next-Generation Sequencing in Molecular Subtyping of Lower-Grade Diffuse Gliomas: Application of the World Health Organization's 2016 Revised Criteria for Central Nervous System Tumors. The Journal of molecular diagnostics : JMD **2017**;19(2):328-37
119. Pant S, Weiner R, Marton MJ. Navigating the rapids: the development of regulated next-generation sequencing-based clinical trial assays and companion diagnostics. Frontiers in oncology **2014**;4:78

120. What is precision medicine? <https://ghr.nlm.nih.gov/primer/precisionmedicine/definition>. Accessed 2019.
121. Health NIo. 2/27/2019. What is precision medicine? <https://ghr.nlm.nih.gov/primer/precisionmedicine/definition>. 2/27/2019.
122. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. The New England journal of medicine **2009**;361(10):947-57
123. Sequist LV, Yang JC, Yamamoto N, O'Byrne K, Hirsh V, Mok T, *et al.* Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. Journal of clinical oncology : official journal of the American Society of Clinical Oncology **2013**;31(27):3327-34
124. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, *et al.* Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. The Lancet Oncology **2012**;13(3):239-46
125. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, *et al.* Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. The New England journal of medicine **2010**;363(18):1693-703
126. Kris MG, Johnson BE, Berry LD, Kwiatkowski DJ, Iafrate AJ, Wistuba, II, *et al.* Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. Jama **2014**;311(19):1998-2006
127. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, *et al.* Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nature genetics **2019**;51(2):202-6
128. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, *et al.* Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. Annals of oncology : official journal of the European Society for Medical Oncology **2019**;30(1):44-56
129. Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, *et al.* Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. The Journal of molecular diagnostics : JMD **2015**;17(3):251-64
130. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nature medicine **2017**;23(6):703-13
131. Mathoulin-Pelissier S, Pritchard-Jones K. Evidence-based data and rare cancers: The need for a new methodological approach in research and investigation. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology **2019**;45(1):22-30
132. Choi EY, Thomas DG, McHugh JB, Patel RM, Roulston D, Schuetze SM, *et al.* Undifferentiated small round cell sarcoma with t(4;19)(q35;q13.1) CIC-DUX4 fusion: a novel highly aggressive soft tissue tumor with distinctive histopathology. The American journal of surgical pathology **2013**;37(9):1379-86
133. Antonescu C. Round cell sarcomas beyond Ewing: emerging entities. Histopathology **2014**;64(1):26-37

134. Eluri M, Feneran A, Bordeaux JS, Ruben B, Ostrowski S, Bastian BC, *et al.* Multiple Merkel cell carcinomas: Late metastasis or multiple primary tumors? A molecular study. JAAD Case Rep **2017**;3(2):131-4
135. Rimmer J, Lund VJ, Beale T, Wei WI, Howard D. Olfactory neuroblastoma: a 35-year experience and suggested follow-up protocol. The Laryngoscope **2014**;124(7):1542-9
136. Thompson LD. Olfactory neuroblastoma. Head and neck pathology **2009**;3(3):252-9
137. Kurman RJ, International Agency for Research on Cancer., World Health Organization. WHO classification of tumours of female reproductive organs. Lyon: International Agency for Research on Cancer; 2014. 307 p. p.
138. Sanderson PA, Critchley HO, Williams AR, Arends MJ, Saunders PT. New concepts for an old problem: the diagnosis of endometrial hyperplasia. Hum Reprod Update **2017**;23(2):232-54
139. Levine RL, Cargile CB, Blazes MS, van Rees B, Kurman RJ, Ellenson LH. PTEN mutations and microsatellite instability in complex atypical hyperplasia, a precursor lesion to uterine endometrioid carcinoma. Cancer research **1998**;58(15):3254-8
140. van Esterik M, Van Gool IC, de Kroon CD, Nout RA, Creutzberg CL, Smit V, *et al.* Limited impact of intratumour heterogeneity on molecular risk assignment in endometrial cancer. Oncotarget **2017**;8(15):25542-51
141. Russo M, Broach J, Sheldon K, Houser KR, Liu DJ, Kesterson J, *et al.* Clonal Evolution in Paired Endometrial Intraepithelial Neoplasia/Atypical Hyperplasia and Endometrioid Adenocarcinoma. Human pathology **2017**
142. Mota A, Colas E, Garcia-Sanz P, Campoy I, Rojo-Sebastian A, Gatius S, *et al.* Genetic analysis of uterine aspirates improves the diagnostic value and captures the intra-tumor heterogeneity of endometrial cancers. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc **2017**;30(1):134-45
143. Kurnit KC, Kim GN, Fellman BM, Urbauer DL, Mills GB, Zhang W, *et al.* CTNNB1 (beta-catenin) mutation identifies low grade, early stage endometrial cancer patients at increased risk of recurrence. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc **2017**;30(7):1032-41
144. Liu Y, Patel L, Mills GB, Lu KH, Sood AK, Ding L, *et al.* Clinical significance of CTNNB1 mutation and Wnt pathway activation in endometrioid endometrial carcinoma. Journal of the National Cancer Institute **2014**;106(9)
145. Hanauer DA, Mei Q, Law J, Khanna R, Zheng K. Supporting information retrieval from electronic health records: A report of University of Michigan's nine-year experience in developing and using the Electronic Medical Record Search Engine (EMERSE). J Biomed Inform **2015**;55:290-300
146. Grasso C, Butler T, Rhodes K, Quist M, Neff TL, Moore S, *et al.* Assessing copy number alterations in targeted, amplicon-based next-generation sequencing data. The Journal of molecular diagnostics : JMD **2015**;17(1):53-63
147. Billingsley CC, Cohn DE, Mutch DG, Stephens JA, Suarez AA, Goodfellow PJ. Polymerase varepsilon (POLE) mutations in endometrial cancer: clinical outcomes and implications for Lynch syndrome testing. Cancer **2015**;121(3):386-94
148. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **2010**;26(1):139-40

149. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic acids research **2012**;40(10):4288-97
150. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met **1995**;57(1):289-300
151. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc **2009**;4(1):44-57
152. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature genetics **2003**;34(3):267-73
153. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America **2005**;102(43):15545-50
154. Wang Q, Armenia J, Zhang C, Penson AV, Reznik E, Zhang L, *et al.* Unifying cancer and normal RNA sequencing data from different sources. Scientific data **2018**;5:180061
155. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, *et al.* A pathology atlas of the human cancer transcriptome. Science **2017**;357(6352)
156. Oda K, Stokoe D, Taketani Y, McCormick F. High frequency of coexistent mutations of PIK3CA and PTEN genes in endometrial carcinoma. Cancer research **2005**;65(23):10669-73
157. Urick ME, Rudd ML, Godwin AK, Sgroi D, Merino M, Bell DW. PIK3R1 (p85alpha) is somatically mutated at high frequency in primary endometrial cancer. Cancer research **2011**;71(12):4061-7
158. Cheung LW, Hennessy BT, Li J, Yu S, Myers AP, Djordjevic B, *et al.* High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability. Cancer Discov **2011**;1(2):170-85
159. Sottoriva A, Kang H, Ma Z, Graham TA, Salomon MP, Zhao J, *et al.* A Big Bang model of human colorectal tumor growth. Nature genetics **2015**;47(3):209-16
160. Sun R, Hu Z, Curtis C. Big Bang Tumor Growth and Clonal Evolution. Cold Spring Harbor perspectives in medicine **2017**
161. Hiley C, de Bruin EC, McGranahan N, Swanton C. Deciphering intratumor heterogeneity and temporal acquisition of driver events to refine precision medicine. Genome biology **2014**;15(8):453
162. Cieslik M, Chugh R, Wu YM, Wu M, Brennan C, Lonigro R, *et al.* The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. Genome research **2015**;25(9):1372-81
163. Kakugawa S, Langton PF, Zebisch M, Howell S, Chang TH, Liu Y, *et al.* Notum deacylates Wnt proteins to suppress signalling activity. Nature **2015**;519(7542):187-92
164. Park BY, Hong CS, Sohail FA, Saint-Jeannet JP. Developmental expression and regulation of the chemokine CXCL14 in Xenopus. The International journal of developmental biology **2009**;53(4):535-40
165. Li CM, Kim CE, Margolin AA, Guo M, Zhu J, Mason JM, *et al.* CTNNB1 mutations and overexpression of Wnt/beta-catenin target genes in WT1-mutant Wilms' tumors. The American journal of pathology **2004**;165(6):1943-53

166. Pendas-Franco N, Garcia JM, Pena C, Valle N, Palmer HG, Heinaniemi M, *et al.* DICKKOPF-4 is induced by TCF/beta-catenin and upregulated in human colon cancer, promotes tumour cell invasion and angiogenesis and is repressed by 1alpha,25-dihydroxyvitamin D3. Oncogene **2008**;27(32):4467-77
167. Larraguibel J, Weiss AR, Pasula DJ, Dhaliwal RS, Kondra R, Van Raay TJ. Wnt ligand-dependent activation of the negative feedback regulator Nkd1. Molecular biology of the cell **2015**;26(12):2375-84
168. Pavlakis K, Messini I, Vrekoussis T, Panoskaltsis T, Chrissanthakis D, Yiannou P, *et al.* PTEN-loss and nuclear atypia of EIN in endometrial biopsies can predict the existence of a concurrent endometrial carcinoma. Gynecologic oncology **2010**;119(3):516-9
169. Slomovitz BM, Jiang Y, Yates MS, Soliman PT, Johnston T, Nowakowski M, *et al.* Phase II study of everolimus and letrozole in patients with recurrent endometrial carcinoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology **2015**;33(8):930-6
170. FitzGerald LM, Jung CH, Wong EM, Joo JE, Gould JA, Vasic V, *et al.* Obtaining high quality transcriptome data from formalin-fixed, paraffin-embedded diagnostic prostate tumor specimens. Laboratory investigation; a journal of technical methods and pathology **2018**;98(4):537-50
171. Ueda G, Yamasaki M, Inoue M, Tanaka Y, Inoue Y, Nishino T, *et al.* Immunohistochemical demonstration of amylase in endometrial carcinomas. International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists **1986**;5(1):47-51
172. Ueda G, Yamasaki M, Inoue M, Tanaka Y, Inoue Y, Nishino T, *et al.* Capacity for amylase production of endometrial carcinomas. Nihon Sanka Fujinka Gakkai zasshi **1985**;37(2):305-6
173. Lee YS, Raju GC. The expression and localization of amylase in normal and malignant glands of the endometrium and endocervix. The Journal of pathology **1988**;155(3):201-5
174. Haffner MC, Weier C, Xu MM, Vaghasia A, Gurel B, Gumuskaya B, *et al.* Molecular evidence that invasive adenocarcinoma can mimic prostatic intraepithelial neoplasia (PIN) and intraductal carcinoma through retrograde glandular colonization. The Journal of pathology **2016**;238(1):31-41
175. Espiritu SMG, Liu LY, Rubanova Y, Bhandari V, Holgersen EM, Szyca LM, *et al.* The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical Aggression. Cell **2018**;173(4):1003-13 e15
176. Hayes MP, Wang H, Espinal-Witter R, Douglas W, Solomon GJ, Baker SJ, *et al.* PIK3CA and PTEN mutations in uterine endometrioid carcinoma and complex atypical hyperplasia. Clinical cancer research : an official journal of the American Association for Cancer Research **2006**;12(20 Pt 1):5932-5
177. Berg A, Hoivik EA, Mjos S, Holst F, Werner HM, Tangen IL, *et al.* Molecular profiling of endometrial carcinoma precursor, primary and metastatic lesions suggests different targets for treatment in obese compared to non-obese patients. Oncotarget **2015**;6(2):1327-39
178. Silverberg SG. Problems in the differential diagnosis of endometrial hyperplasia and carcinoma. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc **2000**;13(3):309-27

179. Wang Y, Li L, Douville C, Cohen JD, Yen TT, Kinde I, *et al.* Evaluation of liquid from the Papanicolaou test and other liquid biopsies for the detection of endometrial and ovarian cancers. Science translational medicine **2018**;10(433)
180. Joshi A, Miller C, Jr., Baker SJ, Ellenson LH. Activated mutant p110alpha causes endometrial carcinoma in the setting of biallelic Pten deletion. The American journal of pathology **2015**;185(4):1104-13
181. Dotto GP, Rustgi AK. Squamous Cell Cancers: A Unified Perspective on Biology and Genetics. Cancer cell **2016**;29(5):622-37
182. Campbell JD, Yau C, Bowlby R, Liu Y, Brennan K, Fan H, *et al.* Genomic, Pathway Network, and Immunologic Features Distinguishing Squamous Carcinomas. Cell reports **2018**;23(1):194-212 e6
183. Pickering CR, Zhou JH, Lee JJ, Drummond JA, Peng SA, Saade RE, *et al.* Mutational landscape of aggressive cutaneous squamous cell carcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research **2014**;20(24):6582-92
184. Li YY, Hanna GJ, Laga AC, Haddad RI, Lorch JH, Hammerman PS. Genomic analysis of metastatic cutaneous squamous cell carcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research **2015**;21(6):1447-56
185. Inman GJ, Wang J, Nagano A, Alexandrov LB, Purdie KJ, Taylor RG, *et al.* The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine associated mutational signature. Nature communications **2018**;9(1):3667
186. Chitsazzadeh V, Coarfa C, Drummond JA, Nguyen T, Joseph A, Chilukuri S, *et al.* Cross-species identification of genomic drivers of squamous cell carcinoma development across preneoplastic intermediates. Nature communications **2016**;7:12601
187. Galor A, Karp CL, Sant D, Joag M, Shalabi N, Gustafson CB, *et al.* Whole Exome Profiling of Ocular Surface Squamous Neoplasia. Ophthalmology **2016**;123(1):216-7 e1
188. Rosen T, Lebwohl MG. Prevalence and awareness of actinic keratosis: barriers and opportunities. Journal of the American Academy of Dermatology **2013**;68(1 Suppl 1):S2-9
189. Rowert-Huber J, Patel MJ, Forschner T, Ulrich C, Eberle J, Kerl H, *et al.* Actinic keratosis is an early in situ squamous cell carcinoma: a proposal for reclassification. The British journal of dermatology **2007**;156 Suppl 3:8-12
190. Boukamp P. Non-melanoma skin cancer: what drives tumor development and progression? Carcinogenesis **2005**;26(10):1657-67
191. McDaniel AS, Hovelson DH, Cani AK, Liu CJ, Zhai Y, Zhang Y, *et al.* Genomic Profiling of Penile Squamous Cell Carcinoma Reveals New Opportunities for Targeted Therapy. Cancer research **2015**;75(24):5219-27
192. Napora C, Cohen EJ, Genvert GI, Presson AC, Arentsen JJ, Eagle RC, *et al.* Factors associated with conjunctival intraepithelial neoplasia: a case control study. Ophthalmic Surg **1990**;21(1):27-30
193. Karp CL, Scott IU, Chang TS, Pflugfelder SC. Conjunctival intraepithelial neoplasia. A possible marker for human immunodeficiency virus infection? Arch Ophthalmol **1996**;114(3):257-61
194. Sayed-Ahmed IO, Palioura S, Galor A, Karp CL. Diagnosis and Medical Management of Ocular Surface Squamous Neoplasia. Expert review of ophthalmology **2017**;12(1):11-9
195. Kiire CA, Srinivasan S, Karp CL. Ocular surface squamous neoplasia. Int Ophthalmol Clin **2010**;50(3):35-46

196. Newton R, Ferlay J, Reeves G, Beral V, Parkin DM. Effect of ambient solar ultraviolet radiation on incidence of squamous-cell carcinoma of the eye. Lancet **1996**;347(9013):1450-1
197. Ateenyi-Agaba C. Conjunctival squamous-cell carcinoma associated with HIV infection in Kampala, Uganda. Lancet **1995**;345(8951):695-6
198. Tabbara KF, Kersten R, Daouk N, Blodi FC. Metastatic squamous cell carcinoma of the conjunctiva. Ophthalmology **1988**;95(3):318-21
199. Karp CL, Galor A, Chhabra S, Barnes SD, Alfonso EC. Subconjunctival/perilesional recombinant interferon alpha2b for ocular surface squamous neoplasia: a 10-year review. Ophthalmology **2010**;117(12):2241-6
200. Tabin G, Levin S, Snibson G, Loughnan M, Taylor H. Late recurrences and the necessity for long-term follow-up in corneal and conjunctival intraepithelial neoplasia. Ophthalmology **1997**;104(3):485-92
201. Teixeira VH, Pipinikas CP, Pennycuick A, Lee-Six H, Chandrasekharan D, Beane J, *et al.* Deciphering the genomic, epigenomic, and transcriptomic landscapes of pre-invasive lung cancer lesions. Nature medicine **2019**
202. Bosic M, Kirchner M, Brasanac D, Leichsenring J, Lier A, Volckmar AL, *et al.* Targeted molecular profiling reveals genetic heterogeneity of poromas and porocarcinomas. Pathology **2018**;50(3):327-32
203. Harms PW, Hovelson DH, Cani AK, Omata K, Haller MJ, Wang ML, *et al.* Porocarcinomas harbor recurrent HRAS-activating mutations and tumor suppressor inactivating mutations. Human pathology **2016**;51:25-31
204. Rodriguez-Paredes M, Bormann F, Raddatz G, Gutekunst J, Lucena-Porcel C, Kohler F, *et al.* Methylation profiling identifies two subclasses of squamous cell carcinoma related to distinct cells of origin. Nature communications **2018**;9(1):577
205. Kanellou P, Zaravinos A, Zioga M, Stratigos A, Baritaki S, Soufla G, *et al.* Genomic instability, mutations and expression analysis of the tumour suppressor genes p14(ARF), p15(INK4b), p16(INK4a) and p53 in actinic keratosis. Cancer letters **2008**;264(1):145-61
206. Nelson MA, Einspahr JG, Alberts DS, Balfour CA, Wymer JA, Welch KL, *et al.* Analysis of the p53 gene in human precancerous actinic keratosis lesions and squamous cell cancers. Cancer letters **1994**;85(1):23-9
207. Rehman I, Takata M, Wu YY, Rees JL. Genetic change in actinic keratoses. Oncogene **1996**;12(12):2483-90
208. Jin Y, Jin C, Salemark L, Wennerberg J, Persson B, Jonsson N. Clonal chromosome abnormalities in premalignant lesions of the skin. Cancer genetics and cytogenetics **2002**;136(1):48-52
209. Garcia-Diez I, Hernandez-Munoz I, Hernandez-Ruiz E, Nonell L, Puigdecanet E, Bodalo-Torruella M, *et al.* Transcriptome and cytogenetic profiling analysis of matched in situ/invasive cutaneous squamous cell carcinomas from immunocompetent patients. Genes, chromosomes & cancer **2019**;58(3):164-74
210. Mortier L, Marchetti P, Delaporte E, Martin de Lassalle E, Thomas P, Piette F, *et al.* Progression of actinic keratosis to squamous cell carcinoma of the skin correlates with deletion of the 9p21 region encoding the p16(INK4a) tumor suppressor. Cancer letters **2002**;176(2):205-14

211. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science signaling **2013**;6(269):pl1
212. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer discovery **2012**;2(5):401-4
213. Lazo de la Vega L, Samaha MC, Hu K, Bick NR, Siddiqui J, Hovelson DH, *et al.* Multiclonality and Marked Branched Evolution of Low-Grade Endometrioid Endometrial Carcinoma. Mol Cancer Res **2019**;17(3):731-40
214. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science **2015**;348(6237):880-6
215. Durinck S, Ho C, Wang NJ, Liao W, Jakkula LR, Collisson EA, *et al.* Temporal dissection of tumorigenesis in primary cancers. Cancer discovery **2011**;1(2):137-43
216. Kim N, Song M, Kim S, Seo Y, Kim Y, Yoon S. Differential regulation and synthetic lethality of exclusive RB1 and CDKN2A mutations in lung cancer. International journal of oncology **2016**;48(1):367-75
217. Knudsen ES, Knudsen KE. Tailoring to RB: tumour suppressor status and therapeutic response. Nature reviews Cancer **2008**;8(9):714-24
218. Pinto AP, Miron A, Yassin Y, Monte N, Woo TY, Mehra KK, *et al.* Differentiated vulvar intraepithelial neoplasia contains Tp53 mutations and is genetically linked to vulvar squamous cell carcinoma. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc **2010**;23(3):404-12
219. Reeves MQ, Kandyba E, Harris S, Del Rosario R, Balmain A. Multicolour lineage tracing reveals clonal dynamics of squamous carcinoma evolution from initiation to metastasis. Nature cell biology **2018**;20(6):699-709
220. Toll A, Salgado R, Yebenes M, Martin-Ezquerra G, Gilaberte M, Baro T, *et al.* MYC gene numerical aberrations in actinic keratosis and cutaneous squamous cell carcinoma. The British journal of dermatology **2009**;161(5):1112-8
221. Toll A, Salgado R, Yebenes M, Martin-Ezquerra G, Gilaberte M, Baro T, *et al.* Epidermal growth factor receptor gene numerical aberrations are frequent events in actinic keratoses and invasive cutaneous squamous cell carcinomas. Experimental dermatology **2010**;19(2):151-3
222. South AP, Purdie KJ, Watt SA, Haldenby S, den Breems N, Dimon M, *et al.* NOTCH1 mutations occur early during cutaneous squamous cell carcinogenesis. The Journal of investigative dermatology **2014**;134(10):2630-8
223. Hosoda W, Chianchiano P, Griffin JF, Pittman ME, Brosens LA, Noe M, *et al.* Genetic analyses of isolated high-grade pancreatic intraepithelial neoplasia (HG-PanIN) reveal paucity of alterations in TP53 and SMAD4. The Journal of pathology **2017**;242(1):16-23
224. Baker SJ, Fearon ER, Nigro JM, Hamilton SR, Preisinger AC, Jessup JM, *et al.* Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. Science **1989**;244(4901):217-21
225. Lambert SR, Mladkova N, Gulati A, Hamoudi R, Purdie K, Cerio R, *et al.* Key differences identified between actinic keratosis and cutaneous squamous cell carcinoma by transcriptome profiling. British journal of cancer **2014**;110(2):520-9

226. Ateenyi-Agaba C, Dai M, Le Calvez F, Katongole-Mbidde E, Smet A, Tommasino M, *et al.* TP53 mutations in squamous-cell carcinomas of the conjunctiva: evidence for UV-induced mutagenesis. Mutagenesis **2004**;19(5):399-401
227. Seshacharyulu P, Ponnusamy MP, Haridas D, Jain M, Ganti AK, Batra SK. Targeting the EGFR signaling pathway in cancer therapy. Expert Opin Ther Targets **2012**;16(1):15-31
228. Gulati N, Beguelin W, Giulino-Roth L. Enhancer of zeste homolog 2 (EZH2) inhibitors. Leuk Lymphoma **2018**;59(7):1574-85
229. Chae YK, Ranganath K, Hammerman PS, Vaklavas C, Mohindra N, Kalyan A, *et al.* Inhibition of the fibroblast growth factor receptor (FGFR) pathway: the current landscape and barriers to clinical application. Oncotarget **2017**;8(9):16052-74
230. Mizrachi A, Shamay Y, Shah J, Brook S, Soong J, Rajasekhar VK, *et al.* Tumour-specific PI3K inhibition via nanoparticle-targeted delivery in head and neck squamous cell carcinoma. Nature communications **2017**;8:14292
231. C1 The first automated solution for single cell genomics, now capable of even more. <https://www.fluidigm.com/products/c1-system>. Accessed 2019.
232. Martelotto LG, Baslan T, Kendall J, Geyer FC, Burke KA, Spraggon L, *et al.* Whole-genome single-cell copy number profiling from formalin-fixed paraffin-embedded samples. Nature medicine **2017**;23(3):376-85
233. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, *et al.* IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. J Clin Invest **2017**;127(8):2930-40
234. Kluger HM, Zito CR, Turcu G, Baine MK, Zhang H, Adeniran A, *et al.* PD-L1 Studies Across Tumor Types, Its Differential Expression and Predictive Value in Patients Treated with Immune Checkpoint Inhibitors. Clinical cancer research : an official journal of the American Association for Cancer Research **2017**;23(15):4270-9
235. Davicioni E, Wai DH, Anderson MJ. Diagnostic and prognostic sarcoma signatures. Molecular diagnosis & therapy **2008**;12(6):359-74
236. Somers GR, Gupta AA, Doria AS, Ho M, Pereira C, Shago M, *et al.* Pediatric undifferentiated sarcoma of the soft tissues: a clinicopathologic study. Pediatric and developmental pathology : the official journal of the Society for Pediatric Pathology and the Paediatric Pathology Society **2006**;9(2):132-42
237. Qualman SJ, Coffin CM, Newton WA, Hojo H, Triche TJ, Parham DM, *et al.* Intergroup Rhabdomyosarcoma Study: update for pathologists. Pediatric and developmental pathology : the official journal of the Society for Pediatric Pathology and the Paediatric Pathology Society **1998**;1(6):550-61
238. Kreiger PA, Judkins AR, Russo PA, Biegel JA, Lestini BJ, Assanasen C, *et al.* Loss of INI1 expression defines a unique subset of pediatric undifferentiated soft tissue sarcomas. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc **2009**;22(1):142-50
239. Specht K, Sung YS, Zhang L, Richter GH, Fletcher CD, Antonescu CR. Distinct transcriptional signature and immunoprofile of CIC-DUX4 fusion-positive round cell tumors compared to EWSR1-rearranged Ewing sarcomas: further evidence toward distinct pathologic entities. Genes Chromosomes Cancer **2014**;53(7):622-33
240. Kawamura-Saito M, Yamazaki Y, Kaneko K, Kawaguchi N, Kanda H, Mukai H, *et al.* Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. Hum Mol Genet **2006**;15(13):2125-37

241. Rakheja D, Goldman S, Wilson KS, Lenarsky C, Weinthal J, Schultz RA. Translocation (4;19)(q35;q13.1)-associated primitive round cell sarcoma: report of a case and review of the literature. Pediatric and developmental pathology : the official journal of the Society for Pediatric Pathology and the Paediatric Pathology Society **2008**;11(3):239-44
242. Italiano A, Sung YS, Zhang L, Singer S, Maki RG, Coindre JM, *et al.* High prevalence of CIC fusion with double-homeobox (DUX4) transcription factors in EWSR1-negative undifferentiated small blue round cell sarcomas. Genes Chromosomes Cancer **2012**;51(3):207-18
243. Himeda CL, Debarnot C, Homma S, Beermann ML, Miller JB, Jones PL, *et al.* Myogenic enhancers regulate expression of the facioscapulohumeral muscular dystrophy-associated DUX4 gene. Molecular and cellular biology **2014**;34(11):1942-55
244. Machado I, Cruz J, Lavernia J, Rubio L, Campos J, Barrios M, *et al.* Superficial EWSR1-negative undifferentiated small round cell sarcoma with CIC/DUX4 gene fusion: a new variant of Ewing-like tumors with locoregional lymph node metastasis. Virchows Archiv : an international journal of pathology **2013**;463(6):837-42
245. Bielle F, Zanello M, Guillemot D, Gil-Delgado M, Bertrand A, Boch AL, *et al.* Unusual primary cerebral localization of a CIC-DUX4 translocation tumor of the Ewing sarcoma family. Acta neuropathologica **2014**;128(2):309-11
246. Graham C, Chilton-MacNeill S, Zielenska M, Somers GR. The CIC-DUX4 fusion transcript is present in a subgroup of pediatric primitive round cell sarcomas. Human pathology **2012**;43(2):180-9
247. Richkind KE, Romansky SG, Finklestein JZ. t(4;19)(q35;q13.1): a recurrent change in primitive mesenchymal tumors? Cancer genetics and cytogenetics **1996**;87(1):71-4
248. Panagopoulos I, Gorunova L, Bjerkehagen B, Heim S. The "grep" command but not FusionMap, FusionFinder or ChimeraScan captures the CIC-DUX4 fusion gene from whole transcriptome sequencing data on a small round cell tumor with t(4;19)(q35;q13). PloS one **2014**;9(6):e99439
249. Smith SC, Buehler D, Choi EY, McHugh JB, Rubin BP, Billings SD, *et al.* CIC-DUX sarcomas demonstrate frequent MYC amplification and ETS-family transcription factor expression. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc **2015**;28(1):57-68
250. Kadakia KC, Tomlins SA, Sanghvi SK, Cani AK, Omata K, Hovelson DH, *et al.* Comprehensive serial molecular profiling of an "N of 1" exceptional non-responder with metastatic prostate cancer progressing to small cell carcinoma on treatment. Journal of hematology & oncology **2015**;8(1):109
251. Chang X, Wang K. wANNOVAR: annotating genetic variants for personal genomes via the web. J Med Genet **2012**;49(7):433-6
252. Brohl AS, Solomon DA, Chang W, Wang J, Song Y, Sindiri S, *et al.* The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation. PLoS genetics **2014**;10(7):e1004475
253. Crompton BD, Stewart C, Taylor-Weiner A, Alexe G, Kurek KC, Calicchio ML, *et al.* The genomic landscape of pediatric Ewing sarcoma. Cancer discovery **2014**;4(11):1326-41
254. Tirode F, Surdez D, Ma X, Parker M, Le Deley MC, Bahrami A, *et al.* Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. Cancer Discov **2014**;4(11):1342-53

255. Jones S, Li M, Parsons DW, Zhang X, Wesseling J, Kristel P, *et al.* Somatic mutations in the chromatin remodeling gene ARID1A occur in several tumor types. Human mutation **2012**;33(1):100-3
256. Roberts P, Burchill SA, Brownhill S, Cullinane CJ, Johnston C, Griffiths MJ, *et al.* Ploidy and karyotype complexity are powerful prognostic indicators in the Ewing's sarcoma family of tumors: a study by the United Kingdom Cancer Cytogenetics and the Children's Cancer and Leukaemia Group. Genes Chromosomes Cancer **2008**;47(3):207-20
257. Neale G, Su X, Morton CL, Phelps D, Gorlick R, Lock RB, *et al.* Molecular characterization of the pediatric preclinical testing panel. Clinical cancer research : an official journal of the American Association for Cancer Research **2008**;14(14):4572-83
258. Jahromi MS, Jones KB, Schiffman JD. Copy Number Alterations and Methylation in Ewing's Sarcoma. Sarcoma **2011**;2011:362173
259. Hattinger CM, Rumpler S, Strehl S, Ambros IM, Zoubek A, Potschger U, *et al.* Prognostic impact of deletions at 1p36 and numerical aberrations in Ewing tumors. Genes Chromosomes Cancer **1999**;24(3):243-54
260. Bichakjian CK, Lowe L, Lao CD, Sandler HM, Bradford CR, Johnson TM, *et al.* Merkel cell carcinoma: critical review with guidelines for multidisciplinary management. Cancer **2007**;110(1):1-12
261. Iyer JG, Storer BE, Paulson KG, Lemos B, Phillips JL, Bichakjian CK, *et al.* Relationships among primary tumor size, number of involved nodes, and survival for 8044 cases of Merkel cell carcinoma. J Am Acad Dermatol **2014**;70(4):637-43
262. Schwartz JL, Griffith KA, Lowe L, Wong SL, McLean SA, Fullen DR, *et al.* Features predicting sentinel lymph node positivity in Merkel cell carcinoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology **2011**;29(8):1036-41
263. Smith FO, Yue B, Marzban SS, Walls BL, Carr M, Jackson RS, *et al.* Both tumor depth and diameter are predictive of sentinel lymph node status and survival in Merkel cell carcinoma. Cancer **2015**;121(18):3252-60
264. Harms PW, Vats P, Verhaegen ME, Robinson DR, Wu YM, Dhanasekaran SM, *et al.* The Distinctive Mutational Spectra of Polyomavirus-Negative Merkel Cell Carcinoma. Cancer research **2015**;75(18):3720-7
265. Wong SQ, Waldeck K, Vergara IA, Schroder J, Madore J, Wilmott JS, *et al.* UV-Associated Mutations Underlie the Etiology of MCV-Negative Merkel Cell Carcinomas. Cancer research **2015**;75(24):5228-34
266. Goh G, Walradt T, Markarov V, Blom A, Riaz N, Doumani R, *et al.* Mutational landscape of MCPyV-positive and MCPyV-negative Merkel cell carcinomas with implications for immunotherapy. Oncotarget **2016**;7(3):3403-15
267. Orlow I, Tommasi DV, Bloom B, Ostrovnaya I, Cotignola J, Mujumdar U, *et al.* Evaluation of the clonal origin of multiple primary melanomas using molecular profiling. The Journal of investigative dermatology **2009**;129(8):1972-82
268. Satter EK, Derienzo DP. Synchronous onset of multiple cutaneous neuroendocrine (Merkel cell) carcinomas localized to the scalp. Journal of cutaneous pathology **2008**;35(7):685-91
269. Blumenthal L, VandenBoom T, Melian E, Peterson A, Hutchens KA. Multiple Primary Merkel Cell Carcinomas Presenting as Pruritic, Painful Lower Leg Tumors. Case reports in dermatology **2015**;7(3):316-21

270. Schrama D, Thiemann A, Houben R, Kahler KC, Becker JC, Hauschild A. Distinction of 2 different primary Merkel cell carcinomas in 1 patient by Merkel cell polyomavirus genome analysis. Archives of dermatology **2010**;146(6):687-9
271. Thakur S, Chalioulias K, Hayes M, While A. Bilateral primary Merkel cell carcinoma of the upper lid misdiagnosed as Basal cell carcinoma. Orbit **2008**;27(2):139-41
272. Pollock J, Caranosos T, Polack EP. Metachronous merkel cell carcinoma: a case report. Case reports in dermatology **2011**;3(3):206-8
273. Kamiyama T, Ohshima N, Satoh H, Fukumoto H, Katano H, Imakado S. Metachronous Merkel cell carcinoma on both cheeks. Acta dermato-venereologica **2012**;92(1):54-6
274. Ahronowitz IZ, Daud AI, Leong SP, Shue EH, Bastian BC, McCalmont TH, *et al.* An isolated Merkel cell carcinoma metastasis at a distant cutaneous site presenting as a second 'primary' tumor. Journal of cutaneous pathology **2011**;38(10):801-7
275. Nagy J, Feher LZ, Sonkodi I, Lesznyak J, Ivanyi B, Puskas LG. A second field metachronous Merkel cell carcinoma of the lip and the palatine tonsil confirmed by microarray-based comparative genomic hybridisation. Virchows Archiv : an international journal of pathology **2005**;446(3):278-86
276. Harms PW, Collie AM, Hovelson DH, Cani AK, Verhaegen ME, Patel RM, *et al.* Next generation sequencing of Cytokeratin 20-negative Merkel cell carcinoma reveals ultraviolet-signature mutations and recurrent TP53 and RB1 inactivation. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc **2016**;29(3):240-8
277. Harms PW, Fullen DR, Patel RM, Chang D, Shalin SC, Ma L, *et al.* Cutaneous basal cell carcinosarcomas: evidence of clonality and recurrent chromosomal losses. Human pathology **2015**;46(5):690-7
278. Nemes S, Danielsson A, Parris TZ, Jonasson JM, Bulow E, Karlsson P, *et al.* A diagnostic algorithm to identify paired tumors with clonal origin. Genes Chromosomes Cancer **2013**;52(11):1007-16
279. Fisher CA, Harms PW, McHugh JB, Edwards PC, Siddiqui J, Palanisamy N, *et al.* Small cell carcinoma in the parotid harboring Merkel cell polyomavirus. Oral surgery, oral medicine, oral pathology and oral radiology **2014**;118(6):703-12
280. Rodig SJ, Cheng J, Wardzala J, DoRosario A, Scanlon JJ, Laga AC, *et al.* Improved detection suggests all Merkel cell carcinomas harbor Merkel polyomavirus. J Clin Invest **2012**;122(12):4645-53
281. Laude HC, Jonchere B, Maubec E, Carlotti A, Marinho E, Couturaud B, *et al.* Distinct merkel cell polyomavirus molecular features in tumour and non tumour specimens from patients with merkel cell carcinoma. PLoS pathogens **2010**;6(8):e1001076
282. Harms PW, Collie AM, Hovelson DH, Cani AK, Verhaegen ME, Patel RM, *et al.* Next generation sequencing of Cytokeratin 20-negative Merkel cell carcinoma reveals ultraviolet-signature mutations and recurrent TP53 and RB1 inactivation. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc **2016**
283. Cimino PJ, Robirds DH, Tripp SR, Pfeifer JD, Abel HJ, Duncavage EJ. Retinoblastoma gene mutations detected by whole exome sequencing of Merkel cell carcinoma. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc **2014**;27(8):1073-87

284. Ahronowitz I, Nghiem P, Yu S. Importance of genetic studies in patients with multiple merkel cell carcinomas. Acta dermato-venereologica **2012**;92(6):633; author reply p 4
285. Loukeri AA, Kampolis CF, Ntokou A, Tsoukalas G, Syrigos K. Metachronous and synchronous primary lung cancers: diagnostic aspects, surgical treatment, and prognosis. Clinical lung cancer **2015**;16(1):15-23
286. Gallagher KK, Spector ME, Pepper JP, McKean EL, Marentette LJ, McHugh JB. Esthesioneuroblastoma: updating histologic grading as it relates to prognosis. The Annals of otology, rhinology, and laryngology **2014**;123(5):353-8
287. Czapiewski P, Kunc M, Haybaeck J. Genetic and molecular alterations in olfactory neuroblastoma - implications for pathogenesis, prognosis and treatment. Oncotarget **2016**
288. Bockmuhl U, You X, Pacyna-Gengelbach M, Arps H, Draf W, Petersen I. CGH pattern of esthesioneuroblastoma and their metastases. Brain pathology **2004**;14(2):158-63
289. Guled M, Myllykangas S, Frierson HF, Jr., Mills SE, Knuutila S, Stelow EB. Array comparative genomic hybridization analysis of olfactory neuroblastoma. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc **2008**;21(6):770-8
290. Riazimand SH, Brieger J, Jacob R, Welkoborsky HJ, Mann WJ. Analysis of cytogenetic aberrations in esthesioneuroblastomas by comparative genomic hybridization. Cancer genetics and cytogenetics **2002**;136(1):53-7
291. Holland H, Koschny R, Krupp W, Meixensberger J, Bauer M, Kirsten H, *et al.* Comprehensive cytogenetic characterization of an esthesioneuroblastoma. Cancer genetics and cytogenetics **2007**;173(2):89-96
292. Szymas J, Wolf G, Kowalczyk D, Nowak S, Petersen I. Olfactory neuroblastoma: detection of genomic imbalances by comparative genomic hybridization. Acta neurochirurgica **1997**;139(9):839-44
293. Valli R, De Bernardi F, Frattini A, Volpi L, Bignami M, Facchetti F, *et al.* Comparative genomic hybridization on microarray (a-CGH) in olfactory neuroblastoma: Analysis of ten cases and review of the literature. Genes Chromosomes Cancer **2015**;54(12):771-5
294. Weiss GJ, Liang WS, Izatt T, Arora S, Cherni I, Raju RN, *et al.* Paired tumor and normal whole genome sequencing of metastatic olfactory neuroblastoma. PloS one **2012**;7(5):e37029
295. Wang L, Ding Y, Wei L, Zhao D, Wang R, Zhang Y, *et al.* Recurrent Olfactory Neuroblastoma Treated With Cetuximab and Sunitinib: A Case Report. Medicine **2016**;95(18):e3536
296. Cha S, Lee J, Shin JY, Kim JY, Sim SH, Keam B, *et al.* Clinical application of genomic profiling to find druggable targets for adolescent and young adult (AYA) cancer patients with metastasis. BMC cancer **2016**;16(1):170
297. Cani AK, Hovelson DH, McDaniel AS, Sadis S, Haller MJ, Yadati V, *et al.* Next-Gen Sequencing Exposes Frequent MED12 Mutations and Actionable Therapeutic Targets in Phyllodes Tumors. Molecular cancer research : MCR **2015**;13(4):613-9
298. Cani AK, Soliman M, Hovelson DH, Liu CJ, McDaniel AS, Haller MJ, *et al.* Comprehensive genomic profiling of orbital and ocular adnexal lymphomas identifies frequent alterations in MYD88 and chromatin modifiers: new routes to targeted therapies. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc **2016**

299. Robinson DR, Wu YM, Vats P, Su F, Lonigro RJ, Cao X, *et al.* Activating ESR1 mutations in hormone-resistant metastatic breast cancer. Nature genetics **2013**;45(12):1446-51
300. Choi W, Porten S, Kim S, Willis D, Plimack ER, Hoffman-Censits J, *et al.* Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. Cancer cell **2014**;25(2):152-65
301. Katoh M. Therapeutics Targeting FGF Signaling Network in Human Diseases. Trends Pharmacol Sci **2016**;37(12):1081-96
302. Papadaki H, Kounelis S, Kapadia SB, Bakker A, Swalsky PA, Finkelstein SD. Relationship of p53 gene alterations with tumor progression and recurrence in olfactory neuroblastoma. The American journal of surgical pathology **1996**;20(6):715-21
303. Shaw AT, Hsu PP, Awad MM, Engelman JA. Tyrosine kinase gene rearrangements in epithelial malignancies. Nat Rev Cancer **2013**;13(11):772-87
304. Gozgit JM, Wong MJ, Moran L, Wardwell S, Mohemmad QK, Narasimhan NI, *et al.* Ponatinib (AP24534), a multitargeted pan-FGFR inhibitor with activity in multiple FGFR-amplified or mutated cancer models. Mol Cancer Ther **2012**;11(3):690-9
305. Dienstmann R, Rodon J, Prat A, Perez-Garcia J, Adamo B, Felip E, *et al.* Genomic aberrations in the FGFR pathway: opportunities for targeted therapies in solid tumors. Annals of oncology : official journal of the European Society for Medical Oncology **2014**;25(3):552-63
306. Williams SV, Hurst CD, Knowles MA. Oncogenic FGFR3 gene fusions in bladder cancer. Hum Mol Genet **2013**;22(4):795-803
307. Heravi-Moussavi A, Anglesio MS, Cheng SW, Senz J, Yang W, Prentice L, *et al.* Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers. The New England journal of medicine **2012**;366(3):234-42
308. Harms PW, Harms KL, Moore PS, DeCaprio JA, Nghiem P, Wong MKK, *et al.* The biology and treatment of Merkel cell carcinoma: current understanding and research priorities. Nature reviews Clinical oncology **2018**;15(12):763-76
309. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. Nature genetics **2003**;33 Suppl:238-44
310. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. Nature **2016**;536(7616):285-91
311. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, *et al.* A global reference for human genetic variation. Nature **2015**;526(7571):68-74
312. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, *et al.* dbSNP: the NCBI database of genetic variation. Nucleic acids research **2001**;29(1):308-11
313. Rottenberg S, Jaspers JE, Kersbergen A, van der Burg E, Nygren AO, Zander SA, *et al.* High sensitivity of BRCA1-deficient mammary tumors to the PARP inhibitor AZD2281 alone and in combination with platinum drugs. Proceedings of the National Academy of Sciences of the United States of America **2008**;105(44):17079-84
314. Maxwell KN, Wubbenhorst B, Wenz BM, De Sloover D, Pluta J, Emery L, *et al.* BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. Nat Commun **2017**;8(1):319
315. Colditz GA, Wolin KY, Gehlert S. Applying what we know to accelerate cancer prevention. Science translational medicine **2012**;4(127):127rv4
316. Peirson L, Fitzpatrick-Lewis D, Ciliska D, Warren R. Screening for cervical cancer: a systematic review and meta-analysis. Systematic reviews **2013**;2:35

317. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA: a cancer journal for clinicians **2017**;67(1):7-30
318. Jacobs I, Gentry-Maharaj A, Burnell M, Manchanda R, Singh N, Sharma A, *et al.* Sensitivity of transvaginal ultrasound screening for endometrial cancer in postmenopausal women: a case-control study within the UKCTOCS cohort. The Lancet Oncology **2011**;12(1):38-48
319. van Nagell JR, Jr., Hoff JT. Transvaginal ultrasonography in ovarian cancer screening: current perspectives. International journal of women's health **2013**;6:25-33
320. Buys SS, Partridge E, Black A, Johnson CC, Lamerato L, Isaacs C, *et al.* Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. Jama **2011**;305(22):2295-303
321. Abrams J, Conley B, Mooney M, Zwiebel J, Chen A, Welch JJ, *et al.* National Cancer Institute's Precision Medicine Initiatives for the new National Clinical Trials Network. American Society of Clinical Oncology educational book American Society of Clinical Oncology Annual Meeting **2014**:71-6
322. Mangat PK, Halabi S, Bruinooge SS, Garrett-Mayer E, Alva A, Janeway KA, *et al.* Rationale and Design of the Targeted Agent and Profiling Utilization Registry (TAPUR) Study. JCO precision oncology **2018**;2018
323. 2018 Decision Memo for Next Generation Sequencing (NGS) for Medicare Beneficiaries with Advanced Cancer (CAG-00450N). <https://www.cms.gov/medicare-coverage-database/details/nca-decision-memo.aspx?NCAId=290&bc=AAAAAAAAAACAA&>. Accessed 2019.
324. Parsons DW, Roy A, Yang Y, Wang T, Scollon S, Bergstrom K, *et al.* Diagnostic Yield of Clinical Tumor and Germline Whole-Exome Sequencing for Children With Solid Tumors. JAMA oncology **2016**
325. Mody RJ, Wu YM, Lonigro RJ, Cao X, Roychowdhury S, Vats P, *et al.* Integrative Clinical Sequencing in the Management of Refractory or Relapsed Cancer in Youth. Jama **2015**;314(9):913-25
326. Harris MH, DuBois SG, Glade Bender JL, Kim A, Crompton BD, Parker E, *et al.* Multicenter Feasibility Study of Tumor Molecular Profiling to Inform Therapeutic Decisions in Advanced Pediatric Solid Tumors: The Individualized Cancer Therapy (iCat) Study. JAMA oncology **2016**
327. Worst BC, van Tilburg CM, Balasubramanian GP, Fiesel P, Witt R, Freitag A, *et al.* Next-generation personalised medicine for high-risk paediatric cancer patients - The INFORM pilot study. European journal of cancer **2016**;65:91-101
328. Ishida H, Iguchi A, Aoe M, Takahashi T, Tamefusa K, Kanamitsu K, *et al.* Panel-based next-generation sequencing identifies prognostic and actionable genes in childhood acute lymphoblastic leukemia and is suitable for clinical sequencing. Annals of hematology **2019**;98(3):657-68
329. Harttrampf AC, Lacroix L, Deloger M, Deschamps F, Puget S, Auger N, *et al.* Molecular Screening for Cancer Treatment Optimization (MOSCATO-01) in Pediatric Patients: A Single-Institutional Prospective Molecular Stratification Trial. Clinical cancer research : an official journal of the American Association for Cancer Research **2017**;23(20):6101-12
330. Mody RJ, Prensner JR, Everett J, Parsons DW, Chinnaiyan AM. Precision medicine in pediatric oncology: Lessons learned and next steps. Pediatric blood & cancer **2017**;64(3)

331. Izquierdo E, Yuan L, George S, Hubank M, Jones C, Proszek P, *et al.* Development of a targeted sequencing approach to identify prognostic, predictive and diagnostic markers in paediatric solid tumours. Oncotarget **2017**;8(67):112036-50
332. Roy S, Raskin L, Raymond VM, Thibodeau SN, Mody RJ, Gruber SB. Pediatric duodenal cancer and biallelic mismatch repair gene mutations. Pediatric blood & cancer **2009**;53(1):116-20
333. Maehara Y, Egashira A, Oki E, Kakeji Y, Tsuzuki T. DNA repair dysfunction in gastrointestinal tract cancers. Cancer science **2008**;99(3):451-8
334. Hiemenz MC, Ostrow DG, Busse TM, Buckley J, Maglinte DT, Bootwalla M, *et al.* OncoKids: A Comprehensive Next-Generation Sequencing Panel for Pediatric Malignancies. The Journal of molecular diagnostics : JMD **2018**;20(6):765-76
335. Northcott PA, Buchhalter I, Morrissy AS, Hovestadt V, Weischenfeldt J, Ehrenberger T, *et al.* The whole-genome landscape of medulloblastoma subtypes. Nature **2017**;547(7663):311-7

Appendix

Chapter 2

Previously published in *Molecular Cancer Research*, co-authored with Scott A. Tomlins, Cody S. Carter, Kathleen R. Cho, Andrew P. Sciallis, Kevin Hu, Mia C. Samaha, Nolan R. Bick, Daniel H. Hovelson, Javed Siddiqui, and Chia-Jen Liu.

Supplementary information and manuscript is available online at the following addresses:

https://doi.org/10.1158/1541-7786.MCR-18-1178

Contributions: Lorena Lazo de la Vega and Scott Tomlins conceived and designed the experiments. Samples were collected and diagnosed by trained pathologists Scott A. Tomlins, Cody S. Carter, Kathleen R. Cho, Andrew P. Sciallis. Lorena Lazo de la Vega with the assistance of Nolan R. Bick, Chia-Jen Liu, and Mia C. Samaha conducted the DNA and RNA sequencing and performed the data analysis. The transcriptome analysis and data visualization was conducted by Kevin Hu with the assistance of Daniel H. Hovelson. Javed Siddiqui helped with the IHC of the FFPE tissue samples. Lorena Lazo de la Vega wrote the final manuscript with the help of Kathleen R. Cho, Scott A. Tomlins, and Cody S. Carter.

Chapter 3

Contributions: Lorena Lazo de la Vega, Scott Tomlins, Paul Harms, and Rajesh Rao conceived and designed the experiments. Cutaneous squamous cell carcinoma samples were collected and classified by Scott A. Tomlins and Paul Harms. Ocular surface squamous neoplasia samples were collected and classified by Scott A. Tomlins, Kellogg Eye Center members Rajesh Rao, Alan Sugar, Hunson Kaz Soong, Shahzad I. Mian, and Hakan Demirci, as well as, Anthony B. Daniels from Vanderbilt University Medical Center, Charles G. Eberhart from Johns Hopkins University School of Medicine, and Camilla Duarte Silva, Suzana Matayoshi from the University of Sao Paulo. Lorena Lazo de la Vega with the assistance of Samantha E. Rahrig, Nolan R. Bick, Chia-Jen Liu, and Mia C. Samaha conducted the sequencing and performed the data analysis. The

transcriptome analysis and heatmap was done by Kevin Hu. Xiaoming Wang conducted all of the HPV testing. Lorena Lazo de la Vega wrote the final manuscript with the help of Paul Harms, Rajesh Rao, and Scott A. Tomlins.

Chapter 4.1

Previously published in *Human Pathology*, co-authored with Daniel H. Hovelson, Andi K. Cani, Chia-Jen Liu, Jonathan B. McHugh, David R. Lucas, Dafydd G. Thomas, Rajiv M. Patel, and Scott A. Tomlins.

Supplementary information and manuscript is available online at the following addresses: https://doi.org/10.1016/j.humpath.2016.09.004

Contributions: UM pathologists David R. Lucas, Dafydd G. Thomas, Jonathan B. McHugh and Rajiv M. Patel obtained and evaluated the *CIC-DUX4* soft tissue sarcoma cohort. Lorena Lazo de la Vega, Scott A. Tomlins, and Rajiv M. Patel conceived and designed the experiments. Lorena Lazo de la Vega with technical assistance from Chia-Jen Liu and Andi Cani completed the scraping of the tissue samples through the sequencing and analysis of the data, as well as the qPCR used to assess *ARID1A* copy number loss. Daniel H. Hovelson assisted in generating the copy number heatmaps. Lorena Lazo de la Vega and Scott A. Tomlins wrote the final manuscript.

Chapter 4.2

Previously published in *JAMA Dermatology*, co-authored with Kelly L. Harms, Daniel H. Hovelson, Samantha Rahrig, Andi K. Cani, Douglas R. Fullen, Chia-Jen Liu, Min Wang, Aleodor A. Andea, Christopher K. Bichakjian, Timothy M. Johnson, Paul W. Harms, and Scott A. Tomlins.

Supplementary information and manuscript is available online at the following addresses: https://doi.org/10.1001/jamadermatol.2017.0507

Contributions: UM physicians and pathologists Kelly Harms, Paul W. Harms, Aleodor A. Andea, Douglas R. Fullen, Christopher K. Bichakjian, Timothy M. Johnson, and Scott Tomlins obtained and evaluated the Merkel cell carcinoma cases. Lorena Lazo de la Vega, Kelly Harms, Paul Harms, and Scott Tomlins conceived and designed the experiments. Lorena Lazo de la Vega with technical assistance from Chia-Jen Liu, Samantha Rahrig, and Andi Cani completed the scraping of the tissue samples through the sequencing and analysis of all variant data reported in

manuscript. Min Wang helped identify MCPyV status of the samples though PCR-Sanger. Daniel H. Hovelson assisted in generating the copy number heatmaps and worked with Paul Harms to generate a similarity index for copy number alterations and mutations. Lorena Lazo de la Vega, Kelly Harms, and Paul W. Harms wrote the final manuscript.

Chapter 4.3

Previously published in *Molecular Cancer Research*, co-authored with Jonathan B. McHugh, Andi K. Cani, Komal Kunder, Frances M. Walocko, Chia-Jen Liu, Daniel H. Hovelson, Dan Robinson. Arul M. Chinnaiyan, Scott A. Tomlins, and Paul W. Harms.
Supplementary information and manuscript is available online at the following addresses:
https://doi.org/10.1158/1541-7786.MCR-17-0135

Contributions: Lorena Lazo de la Vega, Scott A. Tomlins, and Paul W. Harms conceived and designed the experiments. Samples were collected and diagnosed by trained pathologists Paul W. Harms and Jonathan B. McHugh with assistance from Frances M. Walocko. Arul M. Chinnaiyan and Dan Robinson assisted with providing us samples/ analysis from additional samples sequenced through the Mi-Oncoseq program at University of Michigan. Lorena Lazo de la Vega with technical assistance from Chia-Jen Liu and Andi Cani conducted DNA targeted sequencing while Komal Kunder assisted with targeted RNA-seq. Lorena Lazo de la Vega and Scott Tomlins analyzed the RNA-seq. Lastly, Daniel H. Hovelson conducted copy number analysis. Scott A. Tomlins, Lorena Lazo de la Vega, and Paul W. Harms wrote the final manuscript.