Investigation of Intratumoral Heterogeneity in Sarcoma Progression and Metastasis: From Tumor Propagating Cell Dynamics to Clonal Evolution

By

Yuning Jackie Tang

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy in the Department of Laboratory Medicine and Pathobiology, University of Toronto, Canada

2019

© Copyright by Yuning Jackie Tang
Investigation of Intratumoral Heterogeneity in Sarcoma Progression and Metastasis: From Tumor Propagating Cell Dynamics to Clonal Evolution

Yuning Jackie Tang

Doctor of Philosophy

Department of Laboratory Medicine and Pathobiology

University of Toronto

2019

ABSTRACT

Sarcoma is composed of heterogeneous cancer cells characterized by molecular and phenotypic differences. It has been hypothesized that intratumoral heterogeneity can derive from a subpopulation of tumor cells with stem-cell like properties known as tumor propagating cells (TPCs), and/or from different tumor clones progressing in response to evolutionary forces. However, the functional consequences of cellular heterogeneity in sarcoma and the molecular alterations associated with the different sarcoma cell phenotypes is poorly understood. To identify a cell surface marker of sarcoma TPCs, I showed that CD146 is enriched on the surface of side population (SP) cells and are enhanced for tumorigenicity and self-renewal, suggesting they are enriched for TPCs. Furthermore, they represent a distinct but overlapping population of TPCs from SP cells. Gene expression profiling identified activation of Notch and TGF-β pathways are common to both SP and CD146+ TPCs. Inhibition of Notch signaling inhibited tumor self-renewal. To test the stability of TPCs, I performed a series of lineage tracing experiments by labeling different tumor cells in an autochthonous mouse model of undifferentiated pleomorphic sarcoma
driven by $Kras^{G12D}$ and $p53$ mutations. I showed that Non-SP cells that are not enriched for TPCs can give rise to SP cells enriched for TPCs de novo in transplanted tumors.

*In vivo* lineage tracing of tumor clones by multiple fluorescent reporters and DNA barcoding demonstrated that competition and selection induce substantial changes in tumor architecture over time. Specifically, primary tumor growth is driven by a loss of clonal heterogeneity, and local recurrence following therapy is driven by multiple clones. In contrast, a single metastatic clone (MC) is responsible for advanced metastasis and arise through biological selection. Gene expression profiling of MCs compared to Non-MCs found suppression of functional genes in the metastatic cells of the primary tumor. The suppression of these putative metastasis suppressor genes may represent early changes that contribute to metastatic capacity.

In summary, findings described in this thesis revealed that the heterogeneous sarcoma cells consist of CD146+ TPCs, a potentially dynamic TPC state; and as the tumor progresses, the clonal architecture of the tumor alters in response to biological competition and selection.
ACKNOWLEDGEMENT

My pursuit of graduate study was made possible by the many brilliant, inspirational and helpful individuals that I encountered along the way. First and foremost, I would like to express my deep gratitude for my Ph.D. mentor Dr. Benjamin Alman. He has mentored me patiently and tirelessly from the moment I stepped into his lab. With his guidance, I learned to think critically, to be creative, to lead, and to overcome challenges. Dr. Alman gave me every resource and opportunity that I needed to succeed in my graduate studies. It was because of him that I traveled the world for my research, collaborated with renowned scientists, and worked with the most cutting-edge technologies. He afforded me with enormous intellectual freedom, inspired me to develop new ideas, and encouraged me to grow as an independent scientist. Overall, my training with Dr. Alman transformed me from a wide-eyed novice in both science and the world into a critical thinker that may stand on my own.

I was extremely fortunate to be taught and supported by many established scientists during my graduate career. I want to thank my committee members: Drs. Michael Taylor, Gary Bader, and John McPherson for their invaluable feedback on my work, guidance, and willingness to impart me with their expertise. A very special thanks to Dr. David Kirsch, who has co-supervised a critical piece of this thesis work. He generously opened his lab to me and mentored me as if I was his student. I want to thank Dr. Courtney Karner, who I steadily looked to for enthusiasm, scientific insights, and advice while traversing the trenches of science.

I want to thank all the present and past members of the Alman lab, and members of the scientific community at the University of Toronto (Toronto, ON) and Duke University (Durham, NC) for their support. In particular, I want to thank Drs. Makoto Hirata, Shingo Sato, and Hidetoshi Tsushima for their technical assistance and encouragements. I want to thank Dr. Jianguo
Huang for his experimental expertise and helpful discussions. To my peers, Krsitianne Orisian, Hongyuan (Hazel) Zhang, Yihan Liao, Linda Vi, Deepika Sharma, and Hunter Newman, I thank them for creating a collegial environment in which I thrived. And I like to acknowledge Dr. Michael Cook, Dr. Yasuhiro Yahara, Yilin Yu, Anthony Mirando, Vijitha Puvinndran, Puvi Nadesan, and Raymond Poon for technical support.

Last but not least, I want to express my deepest gratitude to my wife and my parents for their unconditional love, understanding and support. They are the source of my strength.
# TABLE OF CONTENTS

ABSTRACT .......................................................................................................................... ii

ACKNOWLEDGEMENT ....................................................................................................... iv

TABLE OF CONTENTS ....................................................................................................... vi

LIST OF FIGURES ............................................................................................................... x

LIST OF TABLES ................................................................................................................. xii

LIST OF ABBREVIATIONS .................................................................................................. xiii

CHAPTER 1 INTRODUCTION .............................................................................................. 1

1.1 INTRATUMORAL HETEROGENEITY AND ITS IMPLICATIONS ...................................... 1

  1.1.1 SOURCES OF HETEROGENEITY IN CANCER ...................................................... 1

  1.1.2 INTRATUMORAL HETEROGENEITY AND METASTASIS ................................. 3

  1.1.3 CLINICAL IMPLICATIONS OF INTRATUMORAL HETEROGENEITY ............... 5

1.2 CONCEPTUAL MODELS OF INTRATUMORAL HETEROGENEITY ............................ 8

  1.2.1 CLONAL EVOLUTION ......................................................................................... 8

  1.2.2 CANCER STEM CELLS AND CELLULAR PLASTICITY .................................... 12

  1.2.3 TUMOR PROPAGATING CELLS AND CLONAL EVOLUTION: THE TWO SIDES OF
       A COIN ...................................................................................................................... 16

1.3 INTRATUMORAL HETEROGENEITY IN SARCOMA ...................................................... 19

1.4 BIOLOGY AND TREATMENTS OF SARCOMA .............................................................. 20

  1.4.1 GENETIC MECHANISMS OF SARCOMAGENESIS ........................................ 20

  1.4.2 CELL SIGNALING MECHANISMS OF SARCOMAGENESIS .............................. 23

  1.4.3 SARCOMA METASTASIS .................................................................................... 25

  1.4.4 MOUSE MODELS OF SARCOMA ...................................................................... 27
1.4.5 TRANSLATIONAL THERAPY FOR SARCOMA…………………………………31

1.5 HYPOTHESIS AND SPECIFIC AIMS…………………………………………………………33

CHAPTER 2 IDENTIFICATION AND CHARACTERIZATION OF CD146+ TUMOR PROPAGATING CELLS IN SARCOMA…………………………………………………………35

2.1 PUBLISHED MATERIALS AND AUTHOR CONTRIBUTION…………………………………35

2.2 INTRODUCTION ........................................................................................................36

2.3 RESULTS....................................................................................................................37

2.3.1 FLOW CYTOMETRY SCREEN OF CELL SURFACE ANTIGEN IN SP CELLS……..37

2.3.2 AN ANTIBODY TO CD146 IDENTIFIED A POPULATION OF SARCOMA CELLS ENRICHED IN SP CELLS..............................................................39

2.2.3 CD146+ CELLS SHOW INCREASED TUMORIGENICITY........................................42

2.3.4 CD146+ CELLS POSSESS SELF-RENEWAL AND GENERATE TUMORS THAT RECAPITULATE PARENT TUMOR..............................................................44

2.3.5 CD146+ ENRICHES FOR A DISTINCT TPC POPULATION FROM SP CELLS……..46

2.3.6 COMMON SIGNALING PATHWAYS ARE ACTIVATED IN CD146+ AND SP POPULATIONS .................................................................46

2.3.7 INHIBITION OF NOTCH SIGNALING REDUCES TUMOR GROWTH AND SELF-RENEWAL.........................................................................................50

2.4 DISCUSSION.............................................................................................................51

2.5 METHODS...............................................................................................................54

CHAPTER 3 TUMOR PROPAGATING CELLS IN SARCOMA IS A DYNAMIC POPULATION..............................................................................................................61

3.1 AUTHORS CONTRIBUTION.....................................................................................61
4.5 Methods

Chapter 5 Conclusions and Future Directions

5.1 Conclusions

5.2 Future Directions

5.2.1 The Epigenome of TPCs and Metastatic Clones

5.2.2 Identification of Druggable Targets in Sarcoma Metastasis

5.2.3 The Role of the Tumor Microenvironment in Intratumor Heterogeneity

5.2.4 Tracing Clonal Heterogeneity at Single Cell Resolution

5.3 Concluding Remarks

References
# LIST OF FIGURES

**Figure 1.1** Intratumoral heterogeneity and its implications ........................................... 1

**Figure 1.2** Models of intratumoral heterogeneity ............................................................. 11

**Figure 2.1** CD146 expression is enriched on the surface of SP cells in human UPS and osteosarcoma ................................................................. 41

**Figure 2.2** Localization of CD146+ cells in human osteosarcoma and UPS tumors ...... 42

**Figure 2.3** CD146 enriches for sarcoma TPCs ................................................................. 45

**Figure 2.4** Notch and TGF-β signaling pathways are significantly upregulated in SP and CD146+ cells ................................................................. 48

**Figure 2.5** Pathway analyses of SP versus NSP and CD146+ versus CD146- cells .... 49

**Figure 2.6** Pharmacological inhibition of Notch signaling significantly reduces osteosarcoma growth and self-renewal ................................................................. 50

**Figure 3.1** KPCC model of sarcoma .................................................................................. 64

**Figure 3.2** SP cells are polyclonal in KPCC sarcomas ....................................................... 65

**Figure 3.3** SP cells are enriched for TPCs in Foxn1NUNU nude mice .............................. 66

**Figure 3.4** SP cells are enriched for TPCs in NSG mice .......................................................... 67

**Figure 3.5** NSP cells can give rise to SP cells de novo ........................................................ 69

**Figure 3.6** NSP cells can compensate for the ablation of SP cells ................................. 72

**Figure 4.1** KPCC and K-sgP53-Cas9 models for tracing clonal evolution ................. 82

**Figure 4.2** Primary tumor progression is associated with a loss of clonal heterogeneity ................................................................. 85

**Figure 4.3** Loss of clonal heterogeneity during primary tumor progression ............. 86

**Figure 4.4** Deep sequencing of DNA barcodes in primary tumors of K-sgP53-Cas9 mice 87
Figure 4.5 Intravital microscopy of tumor progression........................................87
Figure 4.6 Local recurrence in sarcoma is driven by multiple clones...................88
Figure 4.7 Lung metastases is driven by a single clone.........................................93
Figure 4.8 Clonality of metastasis to other organs..................................................95
Figure 4.9 Deep sequencing of DNA barcodes in primary tumors and metastases of K-
sgP53-Cas9 Mice......................................................................................................96
Figure 4.10 Monoclonal origin of metastases is driven by selection......................99
Figure 4.11 Metastasis in KPCC tumors is driven by selection..............................100
Figure 4.12 Metastatic clones (MCs) can derive from minor or dominant clone in the
primary tumor...........................................................................................................101
Figure 4.13 Analysis of CTCs and early monoclonal metastatic seeding..............104
Figure 4.14 Clonality of the early metastatic process..............................................105
Figure 4.15 Transcriptome analysis of tumor clone from primary KPCC sarcomas...106
Figure 4.16 Transcriptome analysis of Non-MCs, MCs, and Lung-Mets..................109
Figure 4.17 Identification of in vivo metastasis suppressors ...................................111
Figure 4.18 Functional validation of suppressors of metastasis............................113
LIST OF TABLES

Table 2.1 Identification of candidate markers enriched on the surface of SP cells.....38

Table 2.2 Abundance of CD146+ cells in human UPS........................................40

Table 2.3 Abundance of CD146+ cells in human osteosarcoma..............................40

Table 2.4. Serial transplantation of CD146+ cells in human UPS tumors...............43

Table 2.5. Serial transplantation of CD146+ cells in human osteosarcoma tumors.....43

Table 2.6. Serial transplantation of different cell fractions in human UPS............47
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKT</td>
<td>Serine/Threonine Kinase 1</td>
</tr>
<tr>
<td>ALDH</td>
<td>Aldehyde Dehydrogenase</td>
</tr>
<tr>
<td>ALDH1A2</td>
<td>Aldehyde Dehydrogenase family 1, subfamily A2</td>
</tr>
<tr>
<td>AML</td>
<td>Acute Myeloid Leukemia</td>
</tr>
<tr>
<td>APC</td>
<td>Adenomatosis Polyposis Coli</td>
</tr>
<tr>
<td>B-ALL</td>
<td>B-Cell Acute Lymphoblastic Leukemia</td>
</tr>
<tr>
<td>BMP</td>
<td>Bone Morphogenetic Protein</td>
</tr>
<tr>
<td>BRCA</td>
<td>Breast Cancer, DNA Repair Associated</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>c-JUN</td>
<td>Jun Proto-oncogene, AP-1 transcription factor subunit</td>
</tr>
<tr>
<td>CAF</td>
<td>Cancer Associated Fibroblast</td>
</tr>
<tr>
<td>CD</td>
<td>Cell Differentiation</td>
</tr>
<tr>
<td>CDK</td>
<td>Cyclin-dependent Kinases</td>
</tr>
<tr>
<td>CDKN2A/B</td>
<td>Cyclin-dependent Kinase Inhibitor 2A/B</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CFP</td>
<td>Cyan Fluorescent Protein</td>
</tr>
<tr>
<td>CHAD</td>
<td>Chondroadherin</td>
</tr>
<tr>
<td>ChIP-Seq</td>
<td>chromatin immunoprecipitation sequencing</td>
</tr>
<tr>
<td>CINSARC</td>
<td>Complexity Index in Sarcomas</td>
</tr>
<tr>
<td>CK5</td>
<td>Cytokeratin 5</td>
</tr>
<tr>
<td>CNV</td>
<td>Copy Number Variations</td>
</tr>
<tr>
<td>CRISPR-Cas9</td>
<td>Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated 9</td>
</tr>
<tr>
<td>CSC</td>
<td>Cancer Stem Cell</td>
</tr>
<tr>
<td>CTC</td>
<td>Circulating Tumor Cell</td>
</tr>
<tr>
<td>CTGF</td>
<td>Connective Tissue Growth Factor</td>
</tr>
<tr>
<td>DAPI</td>
<td>4',6-diamidino-2-phenylindole</td>
</tr>
<tr>
<td>DAPT</td>
<td>N-[N-(3,5-Difluorophenacyl)-L-alanyl]-S-phenylglycine t-butyl ester</td>
</tr>
<tr>
<td>DMNT1</td>
<td>DNA Methyltransferase 1</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DT</td>
<td>Diphtheria Toxin</td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular Matrix</td>
</tr>
<tr>
<td>EGF</td>
<td>Epidermal Growth Factor</td>
</tr>
<tr>
<td>EGFR</td>
<td>Epidermal Growth Factor Receptor</td>
</tr>
<tr>
<td>ELDA</td>
<td>Extreme Limiting Dilution Assay</td>
</tr>
<tr>
<td>EMT</td>
<td>Epithelial-to-Mesenchymal Transition</td>
</tr>
</tbody>
</table>
ETS  Erythroblast transformation-specific
EWSR1 EWS RNA binding protein 1
FACS Fluorescence-Activated Cell Sorting
FC Flow Cytometry
FITC Fluorescein isothiocyanate
FLI1 Fli-1 proto-oncogene, ETS transcription factor
FOXN1 Forkhead box N1
GDP Guanosine Diphosphate
GESTALT Genome Editing of Synthetic Target Arrays for Lineage Tracing
GFP Green Fluorescent Protein
GSEA Gene Set Enrichment Analysis
GTP Guanosine Triphosphate
Gy Gray (unit)
H&E Hematoxylin and Eosin
H3F3A H3 histone family member 3A
HEY1 Hes Related Family BHLH Transcription Factor With YRPW Motif 1
HEY2 Hes Related Family BHLH Transcription Factor With YRPW Motif 2
HRAS HRas Proto-Oncogene, GTPase
IDH Isocitrate Dehydrogenase (NADP(+)) 1, Cytosolic
IL Interleukin
INDELs Insertions/Deletions
INK4A Cyclin Dependent Kinase Inhibitor 2A
ITH Intratumoral Heterogeneity
JNK Mitogen-Activated Protein Kinase 8
KEGG Kyoto Encyclopedia of Genes and Genomes
KMT2A/MLL Lysine Methyltransferase 2A/Mixed-Lineage Leukemia
KP-Cas9 LSL-Kras:p53f/f; LSL-Cas9
KP-DTR LSL-Kras:p53f/f;LSL-DTR
KPCC LSL-Kras:p53f/f;R26R-Confetti/Confetti
KRAS KRAS Proto-Oncogene, GTPase
LGR5 Leucine Rich Repeat Containing G Protein-Coupled Receptor 5
MAPK/ERK Mitogen-Activated Protein Kinase 1/Extracellular Regulated Kinases
MARCl Mouse for Actively Recording Cells!
MCAM/MUC18 Melanoma Cell Adhesion Molecule
MDM2 MDM2 Proto-Oncogene
MEK Mitogen-Activated Protein Kinase Kinase 1
MET Metastases
MC Metastatic Clone
STS  Soft Tissue Tumor
TAM  Tumor Associated Macrophages
TAZ  transcriptional coactivator with a PDZ-binding domain
Tgf-β Transforming Growth Factor Beta 1
TIC  Tumor Initiating Cell
TME  Tumor Microenvironment
TNF-α Tumor Necrosis Factor
TP53/TRP53 Tumor Transformation Related Protein P53
TPC  Tumor Propagating Cell
UPS  Undifferentiated Pleomorphic Sarcoma
WGS  Whole Genome Sequencing
WNT  Wingless-related integration site
YAP  Yes Associated Protein 1
YFP  Yellow Flourescent Protein
ZEB1 Zinc Finger E-Box Binding Homeobox 1
CHAPTER 1
INTRODUCTION

1.1 Intratumoral Heterogeneity and Its Implications

1.1.1 Sources of Heterogeneity in Cancer

Cancers are diseases of uncontrolled clonal growth and invasion (Figure 1.1A). Since the earliest days of cancer pathology in the 19th century, it was recognized that cells within a tumor exhibited distinct morphology from each other. The advent of molecular biology further revealed that the diversity among cancer cells extended into virtually every aspect of cell biology, including features such as proliferative potential, self-renewal, metabolism, and invasiveness (Figure 1.1B) (Almendro, Marusyk, & Polyak, 2013). Research in the past decade suggests that intratumoral heterogeneity may be driven by genomic instability, Darwinian selective pressures, and epigenomic alterations, leading to variable phenotypes among cancer clones.

Figure 1.1. Intratumoral heterogeneity and its implications. A. Common heterogeneous features among tumor cells. B. Intratumoral heterogeneity is a major factor driving therapeutic failure. Current cancer therapy can act as a potent, selective agent on the heterogeneous cancer
cells, leading to the survival and proliferation of more resistant cells after therapy that drives cancer progression.

The most well-established evidence of genetic heterogeneity is from deep sequencing data of different cancer genomes. Genomes sequenced from spatially distinct tumor biopsies of kidney, lung, breast, prostate and esophageal cancers all revealed substantial genetic heterogeneity within the same tumor tissue (Boutros et al., 2015; de Bruin et al., 2014; Gerlinger et al., 2012; Hao et al., 2016; Yates et al., 2015). Other studies using single-cell based sequencing and copy number analysis demonstrated an even higher level of genetic diversity between cancer cells (Navin et al., 2011; Y. Wang et al., 2014; Zack et al., 2013). This diversification of genome may occur through several routes. First, progression towards malignancy requires cells to overcome the telomere crisis; a period of chromosomal instability caused by the shortening of telomeres (Maser & DePinho, 2002). This process leads to aneuploidy, translocations, and dramatic chromosome rearrangements known as chromothripsis (Forment, Kaidi, & Jackson, 2012). In addition, stochastic genetic diversification can arise from replication errors, as DNA repair mechanisms are frequently lost in cancer cells (Dietlein, Thelen, & Reinhardt, 2014). The large number of proliferation cycles for tumor formation and increases in mutation rates allows for diversification of the cancer genome. Furthermore, chemotherapy and irradiation are both potent mutagens, and can increase the mutation loads in cells that survive treatments (Szikriszt et al., 2016). From profiling these mutations, the evolutionary path from tumor initiation to subsequent stepwise development of malignant subclones can be retrospectively constructed. However, phenotypic penetrance of genetic heterogeneity can vary and may be context dependent. Most of the mutations accrued during genome diversification are likely to be passenger mutations that do
not impact phenotype. Indeed, mutations that confer metastatic capacity or therapy resistance are usually present early during tumorigenesis (Hata et al., 2016; Makohon-Moore et al., 2017).

Although progressive genetic mutations and selections can create complex and unpredictable clonal architecture, increasing evidence indicates that epigenetic changes are another crucial driver of intratumoral heterogeneity. DNA methylation and histone modifications have been shown to be important in driving functional heterogeneities such as stem-like cell states and variable therapy responses. Conditional deletion of DNA methyltransferase Dmnt1 blocked the development of leukemia by impairing self-renewal of the “leukemia stem cell” population (Trowbridge et al., 2012). In genetically homogenous non-small cell lung cancer cell lines, a small subpopulation of cells persists upon EGFR inhibition (Sharma et al., 2010). These persisters show altered chromatin landscape and would return to a drug-sensitive state upon therapy withdrawal (Sharma et al., 2010). Combined treatment with EGFR and histone demethylase inhibitors reduced the emergence of persister population (Sharma et al., 2010).

Genetic and epigenetic alterations can work together to influence cancer heterogeneity. Clonal phylogeny inferred from epigenetic analyses shows substantial similarity compared to those inferred from genomic analyses, suggesting co-dependency of both mechanisms in tumor evolution (Brocks et al., 2014; Mazor et al., 2015). Furthermore, studies comparing tumor samples with or without particular somatic mutations identified associations between mutation and DNA methylation profiles. For example, in glioma, IDH1 mutations alter DNA methylation status, and altered DNA methylation promotes the acquisition of mutations in the BRAF gene of colorectal cancers (Turcan et al., 2012; Weisenberger et al., 2006).

1.1.2 Intratumoral Heterogeneity and Metastasis
Metastasis is the colonization of tumor cells to distant organs and is responsible for the vast majority of cancer-related mortality. The process from the primary tumor to metastatic colonization follows a series of orderly biological steps known as the invasion-metastatic cascade (Lambert, Pattabiraman, & Weinberg, 2017). First, the cancer cells invade into the surrounding tissue and intravasate into the circulation. This process often involves the acquisition of biological programs that enhance motility and the ability to degrade the extracellular matrix. In the circulatory system, the circulating tumor cells may travel as single cells or clusters of cells. They must suppress anoikis and evade immune detection to survive (Douma et al., 2004). After reaching the site of colonization, the circulating cells extravasate into the recipient tissue and may undergo a period of latency that can last for years (Malladi et al., 2016). During this time, the metastasis initiating cells must achieve long-term survival in an unfamiliar environment and resist host defenses. Emergence from latency represents the final step of metastasis, in which cancer cells begin overt outgrowth and overtake the tissue.

Whether all or a subset of cancer cells are capable of metastases is unclear. This is in part due to the limited availability of matched primary tumors and metastases samples, and the lack of robust in vivo systems to model autochthonous metastasis. Nonetheless, emerging evidence is beginning to shed light on this question, and the results appear to depend largely on the types of cancer. In breast cancer, lentiviral barcoding of stable human cancer cell lines and multi-lineage tracing techniques in mice showed that multiple clones can initiate overt metastases (Cheung et al., 2016; Wagenblast et al., 2015). Studies in pancreatic cancer yielded similar results. Lineage tracing in an animal model of pancreatic adenocarcinoma and deep sequencing of matched human metastases both point to polyclonal initiation of metastases (Maddipati & Stanger, 2015; Makohon-Moore et al., 2017). However, in prostate cancer, copy number and whole-genome
sequencing comparison of matched human metastases and primary tumors implicate a monoclonal origin (Gundem et al., 2015; W. Liu et al., 2009). Future studies to dissect the clonal relationship between metastases and primary tumors are necessary to predict disease course and to identify effective therapy that targets metastases.

1.1.3 Clinical Implications of Intratumoral Heterogeneity

Despite the substantial advances in the development of novel cancer therapy, most advanced cancers remain incurable. Intratumoral heterogeneity is a key factor leading to therapeutic failure and lethal patient outcome (Figure 1.1B). It is well-established in a wide range of cancer types that increase in cellular heterogeneity in a tumor closely associated with poor prognosis. In chronic lymphocytic leukemia, chemotherapy select for minor clones with driver mutations and is a significant risk factor for rapid disease progression (Landau et al., 2013). In head-and-neck cancer, patients with higher clonal diversity, measured by increased mutant allele burden, correlate with shorter disease-free survival (Mroz & Rocco, 2013). Similar associations are demonstrated in non-small cell lung cancer, glioblastoma, clear cell renal carcinoma and multiple myeloma (Gulati et al., 2014; Keats et al., 2012; Patel et al., 2014; J. Zhang et al., 2014). Patients with a broader assortment of tumor clones, either at the genetic or transcriptional level, have a worse prognosis compared to patients with less number of tumor clones.

The link between patient outcome and intratumoral heterogeneity indicate the need to measure the extent of intratumoral heterogeneity at diagnosis. Current diagnosis is usually based on samples obtained from a needle biopsy or surgical excision. However, this approach is unable to capture the entirety of the disease, because of the widespread spatial heterogeneity in solid tumors, and dynamic changes in clonal composition during therapy. The study by Gerlinger et al.
shows that there are extensive genetic differences between biopsies from different tumor regions of renal cell carcinoma patients (Gerlinger et al., 2012). Specifically, 63–69% of somatic mutations are not detected in all tumor regions. Many tumor suppressor genes have several distinct and spatially separated inactivating mutations in the same tumor, and gene expression signatures of good and poor prognosis are in different regions of a single tumor. The presence of spatial heterogeneity are consistently observed in other cancer types such as breast, brain, and prostate cancers (Boutros et al., 2015; Cooper et al., 2015; J. K. Lee et al., 2017; Nik-Zainal, Van Loo, et al., 2012; Yates et al., 2015). Taken together, these data suggest that representative tissue sampling and genomic assessment of tumor clonality may help inform clinically valuable tumor biology.

Current cancer treatment modalities involve a combination of surgery, chemotherapy, radiotherapy, and targeted therapy. Initially, most cancers will respond to the treatments but ultimately relapses with the outgrowth of cancer cells that are no longer sensitive to treatment. The development of resistant cancer cells is a direct result of intratumoral heterogeneity. The diversity of tumor cells provides a repertoire of differentially drug-sensitive clones, in which intrinsically drug-resistant clones were present before therapy. Equally important is the genomic instability and heterogeneity in the tumor microenvironment that allows the acquisition and selection of resistant phenotypes through evolutionary mechanisms. A well-characterized example of intrinsic resistant clones causing relapse is in hematopoietic malignancies, due to the relative ease of longitudinal sampling in patients. In a study examining the clonal architecture of 30 acute lymphoblastic leukemia patients, five cases were matched pre- and post-treatment samples. This study uses multiplexing fluorescence in situ hybridization to show that clonal architecture at relapse is distinct to that observed at diagnosis, and the relapsed disease can derive
from either major or minor subclones (Anderson et al., 2011). In multiple myeloma, serial sampling of patients undergoing therapy shows that existing cancer clones display different therapeutic responses, and clonal composition change as a result of the selective advantage of the more resistant clones (Bolli et al., 2014). Evidence of pre-existing resistant clones in solid cancers is starting to emerge as well. The study by Wang et al, comparing the genomic and transcriptomic data of 114 glioblastoma patients before and after therapy find that relapse-associated clones typically exist many years before diagnosis (J. Wang et al., 2016).

Incorporating intratumoral heterogeneity and the evolutionary dynamics of heterogeneous cancer clones is essential to overcome therapy failure. However, most current drug development programs do not consider the clonal frequencies of a driver mutation. Given that many actionable targets, such as PTEN, PIK3CA, and MTOR mutations can present as either dominant clonal events or in rare subclones, identifying the relative frequency of distinct cancer clones may significantly improve the efficacy of targeted-therapy (Gerlinger et al., 2012; Voss et al., 2014). Targeting the dominant, truncal mutations are likely more effective than simply considering whether the actionable alterations are present or absent. Furthermore, it is essential to consider the emergence of resistant clones from competitive release or de novo acquisition of mutations during therapy. For example, the EGFRT790M mutation that confers resistance to EGFR inhibition in non-small cell lung cancer can dominate the recurrent tumor either from the expansion of rare pre-existing clones or from newly acquired mutations (Hata et al., 2016). Accurate longitudinal sampling is necessary to overcome the changes in clonal composition during therapy. Serial sampling of tumor genomes from circulating tumor cells can reveal somatic mutations acquired at resistance following conventional therapy. Moreover, exploiting the evolutionary and genetic mechanisms that give rise to resistant clones may be a promising strategy against cancer.
Enriquez-Navas et al. find that decreasing the dose of paclitaxel based on the tumor response can prevent the evolution of resistant clones, thereby improving survival (Enriquez-Navascues et al., 2016). Conceivably, this is accomplished by preventing the release of resistant subclones to undergo rapid growth in the absence of drug-sensitive clones. Drugs that target genomic stability, the mechanism driving clonal divergence can further arrest tumor evolution. Using poly (ADP-ribose) polymerase (PARP) inhibitors on BRCA mutant tumors to increase genomic instability to lethal levels significantly improve patient outcome (J. M. Lee, Ledermann, & Kohn, 2014; Tutt et al., 2010). Taken together, understanding and leveraging intratumoral heterogeneity is an area of priority to advance cancer management.

1.2 Conceptual Models of Intratumoral Heterogeneity

1.2.1 Clonal Evolution

While the biological and clinical importance of epigenetic and genetic heterogeneity in cancer unravel, the central mechanisms that drive the diversification of cancer cells remain uncertain. A guiding principle in understanding tumorigenesis is that cancer is a genetic disease that results from the progressive accumulation of genetic mutations, leading to clonal expansion (Fearon & Vogelstein, 1990; Nowell, 1976). This view is strongly supported by early studies that found increasingly more genetic mutations in later stages of cancer, and by more recent studies modeling tumor initiation via sequential mutagenesis (Drost et al., 2015; Fearon & Vogelstein, 1990; Vogelstein et al., 1988). These studies demonstrate that heritable genetic alterations drive cancer phenotype and laid the ground for viewing cancer as a Darwinian evolutionary process.

The basic principle of Darwinian evolution is the selection of heritable traits with different survival advantages or fitness (Darwin, 1861; van Doorn, Edelaar, & Weissing, 2009; Williams,
During cancer progression, the genomic instability from intrinsic factors such as loss of DNA repair genes and extrinsic factors such as insults from carcinogens or chemotherapy generates a wide array of mutations (Negrini, Gorgoulis, & Halazonetis, 2010). These mutations can be selectively advantageous "driver" mutations, neutral passenger mutations, or deleterious mutations. The cells with advantageous mutations will flourish to produce a clonal population. Over time, new mutations and selective pressures can arise, generating unique new clones. These new mutations can either enhance the fitness of certain clones or diminish their survival. Furthermore, Darwinian selection is context-dependent, and the effect of mutations on survival is subject to change. Ongoing alterations in the microenvironment can generate mutations that are advantageous at one stage of tumorigenesis but become evolutionary dead-ends in later stages (Marusyk & Polyak, 2010). Overall, this process of clonal evolution creates multiple phylogenetic lineages of genetically and phenotypically distinct populations, akin to the branching evolution that Darwin described that led to the diversification of species (Darwin, 1861).

Evidence supporting the clonal evolution model of intratumoral heterogeneity is mainly from high-throughput sequencing of tumor genomes. Several important principles emerged from these studies. First, sequence and copy number heterogeneity at the clonal level are widespread across most malignancies. The mutational catalog of a tumor can be used to decipher the tumor's natural history (Boutros et al., 2015; de Bruin et al., 2014). Alterations identified in every cancer cell of a tumor are likely early events, while mutations that are present in a subset of cells likely arose later in subclonal populations. The phylogenetic trees constructed from such data, especially in the context of strong selective events such as metastasis or therapy, provide support for the role of evolutionary principles in cancer progression. Second, the mutational burden is highly variable between different tumor types. For example, leukemia has a relatively low number of mutations
compared to solid tumors (Landau et al., 2013). Third, few driver mutations are required to initiate tumorigenesis, and they are not restricted to tumor types. Most tumors have 2 to 6 known driver mutations, and there are considerable differences in driver mutations for the same cancer type (Kandoth et al., 2013). Importantly, the same driver mutations can occur in different tumor types, implicating that the same molecular processes are required for tumorigenesis (Alexandrov, Nik-Zainal, Wedge, Campbell, & Stratton, 2013).

Although the evolutionary model gained from deep sequencing efforts prove to be powerful in understanding cancer progression, many challenges remain unresolved. One of these challenges is the separation of passenger and driver mutations among a large number of genetic variants. The thousands of variants identified from whole-genome sequencing makes it difficult to determine their phenotypic impacts without functional tests. Recent studies suggest that most of the mutations identified in a tumor may be fitness-neutral. Sottoriva et al. show that early mutations, occurring when tumors are relatively small, are the most relevant in driving tumor progression (Sottoriva et al., 2015). The majority of mutations accumulated later during clonal expansion have little influence on the tumor phenotype. Other issues in deep sequencing studies reside in sample processing and sequencing technique. The ability to discern clonality is dependent on the number of tumor regions sequenced, the depth and purity of the sequence, and whether single-cell sequencing is implemented (McGranahan & Swanton, 2017). In addition, the conclusions drawn from sequencing data are constructed from mathematical algorithms. Most studies of clonal evolution and tumor phylogenetics adapted algorithms that are developed for species phylogenetics (Schwartz & Schaffer, 2017). Depending on the algorithm used, different conclusions can be derived on important topics such as the role of selection versus stochastic events in driving tumor phylogeny. To illustrate, when looking at single nucleotide variations
(SNVs) and copy number variations (CNVs), some studies conclude that there is little selection in some colorectal tumors, but selection via evolutionary mechanisms may be apparent when examining karyotypes or methylation patterns (Ling et al., 2015; Schwartz & Schaffer, 2017). With reduced cost of deep sequencing technologies and the increasing availability of different algorithm for constructing clonal relationships, the implementation of clonal models is becoming integral for many studies of intratumoral heterogeneity. Rigorous functional test to identify driver mutations, careful design of sequencing methods and samples, and applying multiple phylogenetic algorithm to ensure robustness are crucial factors to consider in future studies of the evolutionary model of carcinogenesis.

Figure 1.2. Models of intratumoral heterogeneity. A. Clonal evolution model of intratumoral heterogeneity. Tumor clones diversify through stochastic genetic alterations in response to evolutionary forces such as competition and selection. B. CSC/TPC model of intratumoral heterogeneity. CSC or TPC at the apex of a differentiation hierarchy are capable of self-renewal and giving rise to other cells of the tumor, such as transit-amplifying cells and differentiated tumor
cells. Transit amplifying cells have limited ability to self-renew and propagate the tumor. Differentiated tumor cells lack the ability to propagate the tumor in transplant experiments.

1.2.2 Cancer Stem Cells and Cellular Plasticity

Most adult tissues are maintained by stem cells that are capable of long term self-renewal and the capacity to differentiate into one or multiple lineages (Wagers & Weissman, 2004). Upon asymmetric cell division, stem cells can give rise to itself and a transit-amplifying cell that will terminally differentiate, and become lost from the tissues (Snippert et al., 2010). Through self-renewal and differentiation, stem cells are responsible for tissue homeostasis and repair. For example, the intestinal stem cells continuously give rise to enterocytes and goblet, Paneth, and enteroendocrine cells throughout life to maintain the intestinal epithelium. Following injury, the intestinal stem cells actively expand and repair the damaged tissue (Snippert et al., 2010). Thus, normal tissues are hierarchically organized with stem cells at the apex, generating the various tissue-specific cell populations. Given all cells in normal tissues are genetically identical, the stem cell hierarchies are maintained by a combination of intrinsic mechanisms such as expression of specific genes, and extrinsic mechanisms such as the microenvironment.

Cancer cells with stem cell like properties, namely the capacity to self-renewal and differentiate, is identified in many types of leukemia and solid tumors. These cells are termed cancer stem cells (CSCs), tumor initiating cells (TICs) or tumor propagating cells (TPCs). The nomenclature can confuse due to the nuanced meaning for each term, and thus need to be clarified. The term CSCs should be restricted to describe cases where self-renewing cancer cells can be prospectively isolated (Kreso & Dick, 2014). TICs, coined based on the ability of these cells to initiate tumors when transplanted, may be ambiguous with the concept of cancer cell-of-origin.
As such, the term TPCs is most applicable to most cancers, where subpopulations of cancer cells show enhanced ability to propagate the tumor continuously.

The method to identify and isolate TPCs are similar regardless of cancer types. It involves dissociation of the primary tumor into single cells, fluorescence-activated cell sorting (FACS) to isolate cells expressing certain markers and transplanting the marker positive and negative cells into immunodeficient mice. Using this approach, TPCs are functionally defined by 3 criteria: 1) the ability to generate xenograft that is representative of the primary tumor; 2) the ability to self-renew as demonstrated by serial passage in transplants; and 3) the ability to give rise to daughter cells (non-TPCs) that are limited in establishing tumors in serial transplantations (O'Brien, Kreso, & Jamieson, 2010). Since the first isolation of TPCs in acute myeloid leukemia (AML), TPCs are identified in malignancies of various tissues such as the lung, colon, prostate, breast, brain, bone and pancreas (Al-Hajj, Wicha, Benito-Hernandez, Morrison, & Clarke, 2003; Bonnet & Dick, 1997; Eramo et al., 2008; Li et al., 2007; Singh et al., 2004; Wu et al., 2007). In all cases, non-TPCs are generated following transplantation of TPCs, providing evidence for a hierarchical organization of tumor tissues. Therefore, TPCs appear to be a common feature across different cancer types. Lineage tracing studies of murine cancer models give further credence to the notion that cancer growth is driven by TPCs. For instance, using the multi-color Cre-reporter R26R-Confetti, Schepers et al demonstrated that Lgr5+ cells in intestinal adenomas fueled tumor growth by generating additional Lgr5+ cells and other cell types in the tumor (Schepers et al., 2012). The transplantation and lineage tracing studies collectively illustrate that the ability to propagate tumors vary between different populations within tumors and this variation may be due to a hierarchical relationship between TPCs and non-TPCs. At the apex of this hierarchy is TPCs,
which through self-renewal and differentiation, give rise to the heterogeneous cancer cells of a tumor.

Despite the evidence supporting TPCs, considerable controversy remains to which tumor types follow the hierarchical model and the methodology used to define TPCs. The gold-standard assay of serially transplanting sorted tumor cells into immunodeficient mice presents several important limitations. The degree of immunodeficiency in the recipient mice can significantly bias the frequency of TPCs. In lung squamous cell carcinoma, transplantation of cancer cells into NOD/SCID/IL2RG/- (NSG) mice increases the TPC frequency by several orders of magnitude compared to transplanting into less immunocompromised NOD/SCID mice (Shultz et al., 2014). In melanoma, transplanting 1 in 4 cells with Matrigel, irrespective of their marker expression, can initiate tumors (Quintana et al., 2008). This suggests that cellular hierarchy does not exist in melanoma under more permissive experimental conditions. Moreover, the transplantation model is inherently unable to recapitulate key aspects of the native tumor microenvironment. Even for orthotopic transplants, a lack of cancer-associated fibroblasts (CAFs), or blood vessels, can potentially lead to a lack of appropriate growth factors for the tumor cells. Indeed, for benign skin tumor cells, the addition of CAFs or endothelial cells is required for successful engraftment (Marsh, Pietras, & McAllister, 2013).

Beyond the technical limitations in characterizing the TPC model, biological limitations must be considered as well. The transplantation assay provides information about what tumor cells are capable of under particular experimental conditions but do not tell us what the cells do during native tumor progression. It is unclear whether the cancer cell population with enhanced tumor propagation capacity in serial transplantation is the same population of cells that contributes to tumor growth, metastasis, and resistance to therapy in their native environment. Furthermore,
transplantation assays examine different cancer cell populations in isolation, therefore unable to study the dynamics of different cancer cell populations and do not consider clonal cooperation or competition that is observed in primary tumors (Cleary, Leonard, Gestl, & Gunther, 2014). In addition, the TPC population itself may be heterogeneous. The validity of one marker in a given group of patients may not be applicable to another group of patients with the same tumor. Indeed, CD133 enriches for TPCs in some glioblastoma patients, whereas Nestin+CD133- cells enriches for TPCs in other patients (Beier & Beier, 2011). This heterogeneity can be extended to within tumors, where distinct TPC populations can be isolated with different markers (Tang, 2012).

The reversible state of stem cells in normal tissues under specific conditions raise the question of whether TPCs are plastic or stable. The seminal study by Yamanaka et al., show that overexpression of 4 transcription factors can convert differentiated fibroblasts into cells that resemble embryonic stem cells (Takahashi & Yamanaka, 2006). Recently, multiple studies have demonstrated that differentiated cells or progenitor cells committed to terminal differentiation in normal tissues can reacquire stem cell features following injury (K. R. Jessen, Mirsky, & Arthur-Farraj, 2015; Tetteh et al., 2016). In the airway epithelium, ablation of CK5+ basal stem cells using inducible diphtheria toxin promote committed luminal secretory cells to dedifferentiate into stable CK5+ basal stem cells (Tata et al., 2013).

Investigating the plasticity between TPCs and non-TPCs have significant biological and therapeutic implications. A plastic TPC population may render the hierarchical organization of tumors inconsequential, and the targeting of TPCs ineffective in treating cancer. Studies demonstrating plastic TPCs are emerging. In melanoma, Jarid1b, a histone demethylase, marks a subpopulation of slow-cycling cells required for tumor growth (Roesch et al., 2010). Nevertheless, single Jarid1b- melanoma cells can give rise to Jarid1b+ cells, highlighting the dynamic nature of
the TPC population. Studies in breast cancer show that factors used to induce epithelial-to-mesenchymal transition (EMT) can generate TPCs from non-TPCs (Mani et al., 2008). The mechanism that underlies this conversion may be epigenetic. The tumor microenvironment can induce the conversion of non-TPCs to TPCs. Secreted factors from myofibroblasts, such as hepatocyte growth factor, can activate Wnt signals in colon cancer cells, thereby inducing a TPC state in more differentiated cancer cells in vivo (Colak & Medema, 2016). The mechanisms that underlie the TPC plasticity likely involve epigenetic modifications. For example, the non-TPCs in breast cancer maintained the Zeb1 promoter in a bivalent state, containing both active and repressive epigenetic marks, enabling these cells to respond to environmental signals such as Tgf-β, which lead to the conversion of cell states (Chaffer et al., 2013).

Given the importance of cellular plasticity in TPCs, it is imperative to determine the extent of plasticity in primary tumor tissues, as opposed to cell lines, from which much of the current literature are built. It is also essential to study cell state conversion using homogenous clonal populations. This will help delineate whether if all clones of non-TPCs or only a subset of non-TPCs can convert to TPCs. Furthermore, the probability of conversion from non-TPCs to TPCs should be investigated. Some studies show that the conversion may be a rare event, and thus the hierarchy model may still be applicable (Gupta et al., 2011). Finally, even for tumors where TPC plasticity is common, stem-cell related phenotypes may still play important roles in therapy resistance and metastasis.

1.2.3 Tumor Propagating Cells and Clonal Evolution: The Two Sides of a Coin

The clonal evolution model and the TPC model of cancer progression both explain the phenotypic heterogeneity observed in cancer. However, each model alone is insufficient to fully
elucidate the diversity of tumor cells. The clonal evolution model focuses on the genetic determinants of cellular phenotypes, often overlooking the non-genetic mechanisms. Similarly, the major limitation of the hierarchical model is that it’s a static, top-down differentiation program in a dynamic system of cells where genetic and environmental changes are frequent. The recent literature demonstrating TPCs can be derived from non-tumorigenic cancer cells reflects that at least certain cancers, do not follow a hierarchy (Cabrera, Hollingsworth, & Hurt, 2015).

To reconcile the two models, it is important to recognize the interplay between epigenetic and genetic determinants of cell behavior and potentially move away from a static hierarchy. For example, genetically identical cells from the same clonal population can possess different capacity to proliferate and self-renew as a result of epigenetic modifications. Since changes in the epigenome are reversible, the less tumorigenic cells in a clone can gain tumorigenic capacity from interactions with the microenvironment or with other cancer cells. This is reflected in the literature of inducible TPCs (Chaffer et al., 2013; Colak & Medema, 2016). Furthermore, TPCs do not necessarily arise from one clonal lineage. Studies in transplanting leukemia cells from the same human sample into multiple mice found genetically distinct clones that are capable of maintaining long-term tumor propagation. In addition to the epigenome, genetic evolution can influence the TPC phenotype as well. In B cell acute lymphoblastic leukemia (B-ALL), TPC frequency in cells with CDKN2A/B mutation is on average 1,000-fold higher than that of samples without these mutations (Mullighan, Williams, Downing, & Sherr, 2008). For solid tumors, TPCs in lung cancer initiated by different genetic mutations express different cell surface markers. In lung tumors initiated by Kras and p53 mutations, Sca-1+ cells are enriched for TPCs (Curtis et al., 2010). However, in lung adenocarcinomas expressing a mutant human EGFR transgene, only Sca-1-negative cells harbor TPC activity (Curtis et al., 2010). These studies highlight the importance of
genetic background on TPCs and may be the reason for distinct TPCs populations present in the same tumor.

Collectively, the epigenetic and genetic influences of tumor propagating potential point to a dynamic model of TPCs that can be integrated with clonal evolution. As the tumor cells evolve genetically and epigenetically, the TPC state can be acquired. In early stage tumors, perhaps a minority of cells have TPC properties. As one or more tumor clones accumulate advantageous mutations, the differentiation process is perturbed further, increasing self-renewal such that the TPC population expands, resulting in their increased frequency (Meacham & Morrison, 2013). In certain cancers, even the progenies of the TPCs can acquire the TPC state with the appropriate combination of external signals. When tumors reach the late stage, the frequency of TPCs may be so high that the tumor clone essentially becomes functionally homogeneous without evidence of a hierarchy (Quintana et al., 2008).

Future studies that combine genetic analyses and TPC assays is needed to determine the validity of the reconciled model of intratumoral heterogeneity. Unbiased genetic analyses of clonal TPCs can reveal their evolutionary history, providing insights into the genetic alterations that contribute to TPC behavior. Additionally, TPC markers should be tested on different stages of a tumor type to test if the hierarchical organization change with disease severity. Studies that combine genetic ablation of TPCs and lineage tracing should be performed to test the ability of non-TPCs to maintain tumor growth. This will help us determine the extent of cell state conversion and their functional relevance in maintaining intratumoral heterogeneity. Integrating the dynamic model of TPCs within the context of clonal evolution adds a new layer of complexity to the mechanisms that drive cancer cell heterogeneity. Understanding the genetic from non-genetic
drivers of TPCs will enable us to gain a unified view of tumor progression and develop better ways to target the diverse cell populations in cancer.

1.3 Intratumoral Heterogeneity (ITH) in Sarcoma

Similar to most cancers, sarcomas are composed of heterogeneous cell populations with distinct biological properties. Studies investigating ITH in sarcoma commonly focuses on identifying and characterizing of TPCs. Cells with tumor-propagating capacity are identified in several sarcoma subtypes. These cells display enhanced self-renewal, asymmetric cell division that gives rise to non-TPC cancer cells and is usually more chemoresistant. In osteosarcoma, CD117, and Stro-1 double positive cells are more tumorigenic in serial transplants models and are more resistant to doxorubicin (Adhikari et al., 2010). In Ewing’s sarcoma and rhabdomyosarcoma, CD133 enriches for cells with TPC characteristics. Cells expressing high levels of CD133 formed more spheres in vitro and are more tumorigenic in vivo (Jiang et al., 2010; Stratford et al., 2011; Terry & Nielsen, 2010; Tirino et al., 2011; Walter et al., 2011). Aldehyde dehydrogenase (ALDH), an enzyme that catalyzes the oxidation of intracellular aldehydes and biogenic amines, is another common marker used to identify TPCs in sarcoma. ALDHhi cells enrich for TPCs in osteosarcoma, Ewing sarcoma and liposarcoma cell lines (Awad et al., 2010; Stratford et al., 2011; L. Wang, Park, Zhang, La Marca, & Lin, 2011).

Even though the presence of TPCs and cellular heterogeneity has been demonstrated in multiple sarcoma types, many critical questions are unanswered. First, the vast majorities of TPC study in sarcoma are using established cell lines. The utility of the markers identified and the characteristic of the marker positive cells in human primary sarcomas are unclear. So far, only the side population (SP) assay, which is based on the ability of cells to efflux the Hoechst dye has shown
to enrich for TPCs in human primary sarcomas (Wu et al., 2007). Second, whether the wide variety of markers and assays used to identify TPCs represent the same or distinct populations is unknown. Lastly, whether TPCs in sarcoma is a stable or dynamic population require investigation. The clinical importance of identifying and characterizing TPCs relies on the goal to target and eradicate these cells. Hence, a dynamic or transient TPC population would render targeting these cells ineffective in treating sarcomas.

1.4 Biology and Treatments of Sarcoma

1.4.1 Genetic Mechanisms of Sarcomagenesis

Sarcomas are a group of diverse tumors that arise in mesenchymal tissues such as bone, muscle, and cartilage. There are over 50 histological subtypes of sarcoma and can be broadly categorized into primary bone sarcoma and soft tissue sarcoma. Osteosarcoma is the most common bone sarcoma, and undifferentiated pleomorphic sarcoma (UPS) is the most common soft tissue sarcoma in adults. In addition to histological distinctions, sarcomas can be categorized based on their genetic features into simple and complex karyotype sarcomas. Simple karyotype sarcomas include Ewing’s sarcoma, alveolar rhabdomyosarcoma, and synovial sarcoma. These tumors typically harbor disease-specific chromosome translocations, most of which involve transcription factors. For example, Ewing’s sarcoma is characterized by fusion between the EWSR1 gene and genes in the ETS family, usually FLI-1 (Grunewald et al., 2015). The fusion mutations are central to the pathogenesis of simple karyotype sarcomas and are incorporated into their diagnosis. The majority of sarcoma subtypes belong to the complex karyotype groups, which include UPS, liposarcoma, osteosarcoma, chondrosarcoma, and leiomyosarcoma (Taylor et al., 2011). These tumors show substantial copy number alteration, complex chromosomal
rearrangements, and the reciprocal translocations characteristic of simple karyotype sarcomas are absent. Large sequencing studies have shown that there are relatively few recurrent gene mutations associated with complex karyotype sarcomas, namely TP53, RB1, CDKN2A and ATRX, all of which are known to have tumor suppressive functions (Cancer Genome Atlas Research Network. Electronic address & Cancer Genome Atlas Research, 2017; X. Chen et al., 2014; Chudasama et al., 2018; Taylor et al., 2011). While the prevalence of mutations in TP53, RB1, CDKN2A and ATRX suggest they may be driver genes in sarcoma initiation or progression, it is important recognize that some patients with complex karyotype sarcomas do not carry mutations in these genes, implicating other genetic events can lead to sarcoma formation (Burningham, Hashibe, Spector, & Schiffman, 2012; Cancer Genome Atlas Research Network. Electronic address & Cancer Genome Atlas Research, 2017; Perry et al., 2014).

Among the recurrent mutations, TP53 is the most commonly mutated gene across a wide-spectrum of sarcoma subtypes. For example, TP53 is mutated in 80% of osteosarcoma, 65% of UPS, 20% of chondrosarcoma, and 15% of rhabdomyosarcoma (Thoenen, Curl, & Iwakuma, 2019). TP53 is a transcription factor that is involved in a myriad of cellular functions including cell cycle regulation, apoptosis, DNA damage response, and metabolism. Mutations in TP53 occur with a range of patterns, co-mutations and allelic configurations that produce diverse functional consequences. SNVs of TP53 are mostly missense mutations, which frequently occur at specific nucleotide residue hotspots such as at R175, G245, R248, R249, R273 and R282 (Petitjean et al., 2007; Shirole et al., 2016). Approximately 25% of TP53 mutations are non-sense or frameshift mutations that predicted to result in loss-of-function, while the impact of other mutations at splice site variants and in-frame indels are unknown. Loss of the second TP53 alleles usually occurs through loss-of-heterozygosity, where allelic patterns of whole chromosome loss, sub arm loss,
uniparental disomy and mutation in \textit{trans} are observed (Kastenhuber & Lowe, 2017). Copy number alterations of TP53 may be particularly important in sarcomas. In UPS, up to 50\% of mutations copy number alterations of TP53 are associated with copy number changes (Steele et al., 2019). Recently, increasingly sophisticated sequencing studies revealed mutations in TP53 co-occurs with mutations in other genes. For example, in lung cancer, TP53 mutation co-occur with mutant Kras and RB1, which can influence tumor cell fitness (Arbour et al., 2018; Rogers et al., 2018).

Although TP53 is a tumor-suppressor, in which loss-of-function mutations drive tumor transformation, certain TP53 mutations can have gain-of-function properties that produce oncogenic phenotypes. For instance, TP53 mutants R175H and R273H induce up-regulation of PDGFR\(\beta\) to promote cancer invasion and metastasis (Weissmueller et al., 2014). Furthermore, mice expressing R172H mutant TP53 spontaneously develop metastatic osteosarcoma and may be mediated through ONZIN-CXCL5-MAPK signaling or binding with TP63, TP73 and ETS2 (Thoenen et al., 2019). Collectively, these data suggest that the spectrum of TP53 mutations affect the tumorigenic process beyond the simple loss-of-function, producing diverse biological consequences dependent on the mutation.

Genomic copy number and structural alterations are critical mechanisms driving sarcomagenesis, and may be more important than SNV/indel mutations, particularly for complex karyotype sarcomas. Indeed, in soft tissue sarcomas, the majority of genetic mutations are copy number alterations and rearrangements, and the overall SNVs/indel mutation burden is relatively low compared to other cancer types (Cancer Genome Atlas Research Network. Electronic address & Cancer Genome Atlas Research, 2017; Steele et al., 2019). In liposarcoma and osteosarcoma, amplifications of 12q15 is found in >90\% and 20\%, respectively (Heidenblad et al., 2006 (Cancer
The co-amplified oncogenes in this region, CDK4, a cell-cycle regulator and MDM2, a negative regulator of TP53 can serve as confirmatory diagnostic markers and potential therapeutic targets in liposarcoma (Weaver et al., 2008). Amplification of 5p is often found in UPS and other high-grade soft tissue sarcomas (Barretina et al., 2010). This amplification is usually accompanied by large deletions affecting well-known tumor suppressors, including TP53, PTEN, and RB1. These tumor suppressor genes play vital roles in regulating genomic stability, and their mutations may represent early events that lead to the copy number changes. Other notable amplifications are JUN at 1p32 and 6q23-25, which spans genes involved in P38 and JNK signaling, tend to occur in liposarcoma and UPS (Taylor et al., 2011 (Cancer Genome Atlas Research Network. Electronic address & Cancer Genome Atlas Research, 2017). In leiomyosarcoma, copy number loss is more frequent than gains. Specifically, loss of 10q21.3 and 13q14.2-q14.3 are found in 75% of cases (Meza-Zepeda et al., 2006).

1.4.2 Cell Signaling Mechanisms of Sarcomagenesis

Dysregulation in cell signaling is widely present in sarcoma and may drive its development. The most well-established transcriptional abnormality that contributes to sarcomagenesis is disruption of the TP53 tumor suppressor pathway. The TP53 gene encodes a transcription factor required for the activation of DNA damage response, cell cycle regulation, and apoptosis (Bieging, Mello, & Attardi, 2014; Muller & Vousden, 2013). Germline loss of TP53 function in inherited diseases such as Li-Fraumeni Syndrome is a leading risk factor for sarcoma development (Hisada, Garber, Fung, Fraumeni, & Li, 1998). In mouse models, loss of TP53 alone is sufficient to induce multiple sarcoma subtypes in the appropriate cell lineage (Sato et al., 2016). Interestingly, while
the phenotypic loss of TP53 tumor suppressor pathway is common, many sarcoma patients retain wildtype TP53. These findings suggest that changes in other components of the TP53 pathway, such as amplification of MDM2, a negative regulator of the TP53 pathway, or suppression of TP53 targets genes may be involved. Indeed, elevated levels of MDM2 induced by single nucleotide polymorphism SNP309 in both human and mice can accelerate the development of sarcoma (Bond et al., 2006).

The retinoblastoma (RB) pathway represents a second major tumor suppressor pathway dysregulated in sarcomas. Similar to Li-Fraumeni patients, inherited germline RB1 mutations significantly increase the risk for sarcoma development. Inactivation of the RB1 is especially frequent in osteosarcoma and rhabdomyosarcoma (Kleinerman, Schonfeld, & Tucker, 2012). In patients with wildtype RB1, suppression of RB pathway often occurs through the loss of p16INK4A, a negative regulator of the CDK-cyclin complexes that phosphorylate and activate RB (Gonin-Laurent et al., 2007; Y. Liu et al., 2013).

Activation of oncogenic signaling also drives sarcoma development. RAS and PI3K/AKT/MTOR pathways are commonly activated in sarcoma. RAS family of proteins consist of three canonical members HRAS, NRAS, and KRAS (Pylayeva-Gupta, Grabocka, & Bar-Sagi, 2011). They belong to a larger family of low-molecular weight GTP binding proteins. Like other G proteins, RAS cycles between the GDP-bound inactive form and the GTP-bound active form. In the active form, RAS act as a molecular switch to turn on a vast network of signaling pathways, such as the MAPK/ERK signaling, and stimulates cell growth, proliferation, and migration (Pylayeva-Gupta et al., 2011). While somatic mutations in the RAS family of genes are rare in UPS, the activation of RAS pathways is observed in over 50% of patient cases, and RAS family mutations are present in 93% of patients with rhabdomyosarcoma (Pylayeva-Gupta et al., 2011;
Furthermore, deletions of NF1, a negative regulator of RAS signaling is observed in sarcoma patients as well (Barretina et al., 2010; Cancer Genome Atlas Research Network. Electronic address & Cancer Genome Atlas Research, 2017). In mouse models, activation of KRAS or deletion of NF1 in TP53 deficient mice is sufficient to generate high grade UPS and rhabdomyosarcoma (Blum et al., 2013; Dodd et al., 2013). Activation of PI3K/AKT is found primarily in leiomyosarcoma and myxoid liposarcomas (Barretina et al., 2010). The PI3K/AKT is stimulated by growth factor receptor tyrosine kinases such as EGFR to regulate cellular processes such as metabolism, inflammation, cell survival, and motility (Martini, De Santis, Braccini, Gulluni, & Hirsch, 2014). Recently, study shows that activating mutation in PIK3CA gene, which encode the catalytic subunit of PI3K is one of the most common mutations in liposarcoma and is associated with poor survival in these patients (Barretina et al., 2010). Increased activation of PI3K pathway activity is highly prevalent in leiomyosarcoma (Cancer Genome Atlas Research Network. Electronic address & Cancer Genome Atlas Research, 2017). Inhibition of PI3K/AKT signaling can slow tumor progression in vivo in leiomyosarcoma and osteosarcoma (Fang, Li, Liu, Lee, & Aaronson, 2001; Zhao et al., 2013). Therefore, activation PI3K/AKT signaling may contribute to tumor progression in multiple types of sarcomas and may be a potential therapeutic target.

1.4.3 Sarcoma Metastasis

Metastatic disease in sarcoma primarily occur in the lungs, and is the leading cause of patient deaths. For bone sarcoma, which includes osteosarcoma and Ewing’s sarcoma, 15-25% of patients present with metastases and 30%-40% of patients with high-grade STS develop metastases (Frezza, Stacchiotti, & Gronchi, 2017; M. Huang & Lucas, 2011; Lindsey, Markel, & Kleinerman,
Treatments for metastatic sarcoma is challenging, with a 5-year survival rates of less than 20% for bone sarcoma and STS (Frezza et al., 2017; M. Huang & Lucas, 2011; Lindsey et al., 2017; Meyer & Seetharam, 2019). The standard of care for osteosarcoma involves neoadjuvant chemotherapy, surgical resection of the tumor, and post-operative adjuvant chemotherapy. Chemotherapy for osteosarcoma are usually used in combination. The four drugs that shown consistent activity include: cisplatin, doxorubicin, high-dose methotrexate with leucovorin rescue, and isofosfamide (Lindsey et al., 2017). The use of chemotherapy drastically improved the overall survival of osteosarcoma patients, from 20% to over 65%. Most of the benefits are derived from patients that did not present with metastases, suggesting the presence of micrometastases prior to surgery that responded to systemic chemotherapy (Misaghi, Goldin, Awad, & Kulidjian, 2018). In contrast, the use of chemotherapy to treat metastatic STS is controversial (Frezza et al., 2017). The most commonly used single-agent chemotherapy is doxorubicin and isofosfamide. Most trials show response rates of 10% to 25%, and trials using combination chemotherapy do not show significant survival benefits compared to single agents (Meyer & Seetharam, 2019). The pooling of different histological subtypes in many of these trials may have limited the study outcomes, especially considering the etiological heterogeneity that exists within STS subtypes (Frezza et al., 2017). Nonetheless, the relatively limited response rate and potential long-term toxicities from chemotherapy remain as major hurdles for adopting chemotherapy to treat STS metastases.

The genetic and molecular mechanism that drive human sarcoma metastasis is not well understood, and may depend on the specific subtype of sarcoma. Currently, emerging evidence suggest that dysregulation of gene expression, particularly genes associated with extracellular matrix interaction and epigenetic modification may play crucial roles. Microarray analysis of 64 human primary myxofibrosarcoma identified over-expression of ITGA10, which encodes for
integrin-α10 is associated with increased distant metastasis (Okada et al., 2016). Functional experiments in cell lines and xenograft models show that ITGA10 regulates TRIO and RICTOR, which are co-amplified on 5p and overexpressed in half of analyzed myxofibrosarcomas to promote tumor survival and metastasis (Okada et al., 2016). Somatic SNV in ADAM17, a gene involved cell-cell and cell-matrix interaction has been shown to contribute to metastasis in human synovial sarcoma (Xing et al., 2018). In a genome-wide association study of 935 osteosarcoma, and concurrent mouse sleeping beauty transposon mutagenesis identified germline single-nucleotide polymorphism NFIB rs7034162 is positively associated with metastasis at diagnosis (Mirabello et al., 2015). A recent study that examined primary metastatic sarcomas from 43 patients of varying subtypes found that large-scale structural variation may be the primary mechanism of mutagenesis, and these genetic profiles may inform clinical decisions to improve patient response (Feng et al., 2019). Specifically, patients with null mutations in TCS2 and mutational signature of homologous recombination deficiency may benefit from rapamycin inhibitors and certain DNA-damaging agents (Feng et al., 2019). In addition genetic mutations, epigenetic modification can impact metastasis in human sarcoma. Profiling of histone modifications H3K4me1 and H3K27ac by ChIP-seq revealed substantial differences in enhancer landscape between primary and metastatic osteosarcoma (Morrow et al., 2018). Metastatic activity can be diminished by targeting genes such as F3 that are associated with the variable enhancer loci in the metastatic cells. Importantly, disruption of the gained metastatic variant enhancer loci is sufficient to mitigate lung metastasis in osteosarcoma (Morrow et al., 2018).

1.4.4 Mouse Models of Sarcoma
Modeling sarcomagenesis and progression in mice relies on two approaches, namely transplanting sarcoma cell lines into immunocompromised animals or inducing tumor formation in genetically engineered mouse models (GEMMs). The key advantages of using the transplant model are that the experiments can be performed quicker compared to GEMMs; multiple experiments can be tested on the same cancer cell line; and in the case of human patient derived xenografts (PDX), the tumors may better represent the genetic abnormalities found in human cancers. Transplantation of sarcoma cell lines, either murine (allograft) or human (xenograft) can be performed subcutaneously or at the orthotopic site to initiate local tumor formation. Subcutaneous injection are technically easy to perform, and the tumors are easily detectable and accessible during experiments. However, the subcutaneous microenvironment may be different from the native tumor microenvironment. The subcutaneous transplants are able to maintain the histopathological and genetic features of sarcomas (Cornillie et al., 2019; Mayordomo et al., 2010; Sampson, Kamara, & Kolb, 2013; C. Y. Wang et al., 2012). Studies in osteosarcoma and UPS has shown that the certain aspects of gene expression patterns and copy number alterations characteristic of sarcoma are maintained in the subcutaneous transplants, and therefore are valuable tools for molecular studies and drug testing (Cornillie et al., 2019; Mayordomo et al., 2010; C. Y. Wang et al., 2012). For example, mouse osteosarcoma xenografts has been used to optimize effects of standard chemotherapeutic drugs and investigation of potential molecular drivers of tumor formation (Bruheim, Bruland, Breistol, Maelandsmo, & Fodstad, 2004; Sayles et al., 2019). However, tumors formed subcutaneously in mice do not fully recapitulate the human tumor microenvironment, and the complex interplay between stroma and tumor may be lost (Balkwill, Capasso, & Hagemann, 2012; Cornillie et al., 2019; Fidler, 2003; Koga & Ochiai, 2019). Orthotopic transplants such as implanting osteosarcoma at the femur through intraosseous
injections may better recapitulate the tumor microenvironment compared to the subcutaneous model, but the replacement of human stromal cells that developed with the tumor throughout the process of tumor formation with mouse cells may still impact the tumor biology (Jacques et al., 2018). For instance, mouse cytokines and growth factors may have different affinity for the engrafted human tumor. Equally important, transplants in immunocompromised animals largely removes the role of the host immune system, particularly the adaptive immune system on the tumor. Another potential limitation of the transplant mouse model is that each successive transplant may select for tumor cells that grow better in transplantation, thereby biasing the cellular architecture of the tumor over time (Eirew et al., 2015).

GEMMs of cancer are generated using genetic knock-in or knock-out strategies to introduce genetic modifications that produce autochthonous tumors. The genetic modifications are often under the control of site-specific recombinase systems such as Cre-LoxP or Flp-FRT to spatially and temporally control the formation of tumors. GEMM models of sarcoma has been developed for osteosarcoma, UPS, rhabdomyosarcoma, chondrosarcoma, synovial sarcoma, liposarcoma and gastrointestinal stromal tumor (GIST) (Dodd, Mito, & Kirsch, 2010; Walkley et al., 2008). Compared to the transplant model, tumors from GEMMs are developed de novo in a fully immunocompetent host environment. The use of GEMMs have provided important insights to the genetic determinants of tumorigenesis, including sarcoma (Dodd et al., 2010). Nonetheless, GEMMs have several important differences from human tumors. First, mutations in human tumors are accumulated in a step-wise manner over time, while mutations in GEMMs such as Kras\textsuperscript{G12D} mutation and p53 deletions are usually induced simultaneously, and tumor latency is likely shorter compared to human tumors, which may take decades to form. Importantly, for mouse UPS, which frequently uses Kras\textsuperscript{G12D} mutation, rarely occurs in human patients, despite frequent activation of
RAS/MAPK signaling pathway (Cancer Genome Atlas Research Network. Electronic address & Cancer Genome Atlas Research, 2017; Li et al., 2015). Second, mutations in human tumors are accumulated over long periods, often with large numbers of passenger mutations, whereas in GEMMs, the mutational burden is relatively low. For example, sequencing of mouse model of UPS showed fewer mutations compared to human tumors (Cancer Genome Atlas Research Network. Electronic address & Cancer Genome Atlas Research, 2017; J. Huang et al., 2017). Third, human tumors are clonal, whereas multiple cells in GEMMs are induced with the founder mutations at simultaneously, creating multiple tumor clones.

Modeling sarcoma metastasis primarily uses direct injection of tumor cells into the circulation of immunocompromised mice, usually through the tail-vein. The tumor cells are carried to the lungs through the blood, and establishes metastatic lesions. The lungs are the major site of metastasis for different sarcoma subtypes, including osteosarcoma, chondrosarcoma, synovial sarcoma, and UPS (Burningham, Hashibe, Spector, & Schiffman, 2012; Lindsey et al., 2017; Singhi, Moore, & Muslimani, 2018). While the injection model can deliver sarcoma cells to the most common location of metastasis, it bypasses the local invasion and intravasation steps of the metastatic cascade (Gomez-Cuadrado, Tracey, Ma, Qian, & Brunton, 2017). For the mouse UPS model, the primary autochthons tumor can be surgically removed, and 40% of mice develops metastasis to the lungs, and a small proportion presents with liver metastases in addition to lung lesions (Kirsch et al., 2007; Sachdeva et al., 2014). While this model captures the major steps in metastasis, from tumor growth, invasion, intravsation, extravasation and distant tissue colonization, it may differ with human cancers with a shorter latency of 2 months post amputation compared to potentially years in humans (Sachdeva et al., 2014). This difference may result in fewer genetic divergence between mouse metastases relative to the primary tumor.
1.4.5 Translational Therapy for Sarcoma

Conventional therapy for sarcomas mainly utilizes surgery with adjuvant radiotherapy, and is often curative for localized disease. The local recurrence rates for sarcomas of the extremity is between 10-15%, and the association between local recurrence and metastatic spread remains controversial (Clark, Fisher, Judson, & Thomas, 2005; Skubitz & D'Adamo, 2007). Cytotoxic chemotherapy is generally ineffective for most sarcomas, except for osteosarcoma, rhabdomyosarcomas and Ewing's sarcomas. A meta-analysis of multiple trials using doxorubicin or combination chemotherapy showed no overall survival benefits ("Adjuvant chemotherapy for localised resectable soft-tissue sarcoma of adults: meta-analysis of individual data. Sarcoma Meta-analysis Collaboration," 1997). As a result, the principle aim for chemotherapy in sarcoma is palliation. The overall survival rates for sarcoma patients under the current standard of care is 65%, and given the limited utility of chemotherapy in treating systemic disease, the 5-year survival for metastatic sarcoma is less than 20% (Clark et al., 2005).

Research into the basic biology of sarcoma is extending targeted therapies and immunotherapies to improve patient care. A substantial amount of work on targeted sarcoma therapy is on kinase inhibition. For example, genetic and chemical inhibition of CDK4, a serine/threonine kinase overexpressed in multiple sarcoma subtypes, can significantly slow tumor growth in preclinical models of fibrosarcoma, liposarcoma, and leiomyosarcoma (Perez, Munoz-Galvan, Jimenez-Garcia, Marin, & Carnero, 2015). Combination of doxorubicin with CDK inhibitor flavopiridol significantly improved tumor response in a clinical trial, suggesting a synergy between the two drugs. Mitogen-activated protein kinase (MAPK), particularly mitogen-activated protein kinase kinase (MEK) is another promising target for sarcoma therapy. Allosteric inhibition of MEK can significantly reduce tumor growth in vitro and patient-derived xenograft
models of neurofibromas, UPS, and osteosarcoma (Dodd et al., 2013; Jessen et al., 2013; Yu, Luk, Yang, & Walsh, 2011). A different area of interest for targeted therapy is to inhibit stem-cell pathways such as Wnt, Notch, and Hedgehog signaling. These pathways regulate the self-renewal and differentiation of mesenchymal progenitor cells and may specifically target TPCs in sarcoma. Inhibition of Notch and Hedgehog pathway significantly reduces tumor growth and self-renewal in osteosarcoma and UPS (C. Y. Wang et al., 2012; Wei et al., 2015).

Significant strides in understanding immunobiology related to cancer led to several promising immunotherapy for epithelial cancers such as melanoma and breast cancer (Khalil, Smith, Brentjens, & Wolchok, 2016). These successes prompted efforts to develop immunotherapy for sarcomas. One example is adoptive T-cell therapy directed against NY-ESO-1, a tumor-associated antigen expressed in myxoid liposarcoma, leiomyosarcoma, and 80% of synovial sarcomas (Robbins et al., 2011). In a clinical trial of synovial sarcoma patients that were refractory to standard therapy, treatments with autologous T cells transduced with a T-cell receptor targeting NY-ESO-1 resulted in objective clinical response in 4 out of 6 patients (Robbins et al., 2011). In addition, the use of monoclonal antibodies targeting programmed death 1 (PD-1) receptor is being pursued to treat different subtypes of soft tissue sarcoma. PD-1 is a cell surface receptor expressed on activated and exhausted T cells, and function to inhibit T cell activity upon binding with its ligands PD-L1 or PD-L2 (Chen & Han, 2015). Tumor-associated PD-L1 expression has been reported in up to 65% of different sarcomas subtypes and is associated with poor survival (D'Angelo et al., 2015; Kim et al., 2013). The use of anti PD-1 monoclonal antibodies such as pembrolizumab and nivolumab is reported to significantly reduce tumor burden in some patients with UPS and liposarcoma, but not leiomyosarcoma (Ben-Ami et al., 2017; Mitsis, Francescutti, & Skitzki, 2016; Paoluzzi et al., 2016; Tawbi et al., 2017).
1.5 Hypothesis and Specific Aims

Since the identification of tumor propagating cells (TPCs) in sarcoma, it is unclear whether TPCs is a homogeneous or heterogeneous population, and whether the TPC phenotype are a stable or dynamic. Furthermore, the biological significance of different sarcoma clones contributing to tumor progression is not known. I hypothesized that sarcomas consist of heterogeneous cell populations with dynamic cell states and the clonal architecture of the tumor alters in response to biological selection during different stages of tumor progression. To test this hypothesis, I addressed the following questions:

Aim 1: Can sarcoma TPCs be enriched with a cell surface marker, and does the marker positive TPCs represent the same population as previously identified SP cells?

I performed a high throughput cell surface antigen screen on SP cells from primary human sarcoma to identify markers that can enrich for sarcoma TPCs. The marker positive cells are tested for in vivo tumor initiation potential and self-renewal using serial transplant models. The gene expression profile of marker positive population is characterized by microarray. The inter-relationship between SP and marker positive cells are investigated by FACS sorting different fractions of cells that are SP and marker positive or expressing either markers alone and testing their tumor initiating potential is compared.

Aim 2: Are TPCs in sarcoma a stable or dynamic population of tumor cells?

I labeled tumor subpopulations in a spatially and temporally restricted mouse model of sarcoma using R26R-Confetti allele. The stability of SP in this model is determined by co-transplanting SP and Non-SP (NSP) cells expressing different stable reporters to evaluate whether NSP cells
gave rise to SP cells *de novo*. I also genetically ablated SP cells expressing diphtheria toxin receptor and ablated these cells in a transplant model to determine its impact on tumor growth and self-renewal. The dynamic nature of SP cells is analyzed using lineage trajectory projections from single-cell RNaseq experiments.

**Aim 3: What are the biological differences between distinct tumor clones, and which clones are responsible for tumor progression and metastasis?**

By labeling tumor clones in an autochthnous mouse model of undifferentiated pleomorphic sarcoma using R26R-Confetti and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) barcoding, I performed lineage tracing of sarcoma clones throughout the major stages of tumor progression. The contribution of tumor clones to primary tumor growth, local recurrence following therapy, and distant metastasis are elucidated. A single clone responsible for advanced metastasis was identified, and their transcriptomic profiles were characterized.
CHAPTER 2
IDENTIFICATION AND CHARACTERIZATION OF CD146+ TUMOR PROPAGATING CELLS IN SARCOMA

2.1 Published Materials and Author Contribution


Q. Wei and I performed all the experiments. I performed all the data analyses, and V. Voisin assisted with the bioinformatic analysis. All experiments and data analyses were conducted in the laboratory of Dr. Benjamin Alman, who helped conceive and supervise the project.
2.2 Introduction

Sarcomas are a diverse group of tumors that arise in the mesenchymal tissues. Osteosarcoma is the most common form of primary bone sarcoma and undifferentiated pleomorphic sarcoma (UPS), previously known as malignant fibrous histiocytoma, is a common form of soft tissue sarcoma. Similar to many human cancers, osteosarcoma and UPS both display substantial intratumoral heterogeneity. Previous studies found that many sarcomas contain a small subpopulation of cells called tumor-propagating cells (TPCs), characterized by enhanced tumorigenicity and self-renewal capacity (Dela Cruz, 2013; C. Y. Wang et al., 2012; Wu et al., 2007). TPCs have been hypothesized to drive tumor initiation and progression (Kreso & Dick, 2014; Kreso et al., 2014). Therefore, selective targeting of these cells may provide more effective treatment than conventional therapies of cancer.

The identification of TPCs in many tumor types relies on flow cytometry analysis of specific stem cell surface markers. For sarcomas with complex karyotypes, the lack of clearly defined mesenchymal stem cell markers has limited such efforts. Previous studies relied on markers that were known to enrich for TPCs in other cancers, such as CD133 and ALDH (Awad et al., 2010; Tirino et al., 2011). However, these studies were performed in cell lines, rather than primary tumors, and did not demonstrate robust in vivo serial transplantation capacity (Dela Cruz, 2013; Jiang et al., 2010; Tirino et al., 2011). Another approach is to use functional assays to enrich for sarcoma TPCs, such as using the side population (SP) assay (Wu & Alman, 2008; Wu et al., 2007). This assay is based on the ability of stem-like and progenitor cells to efflux Hoechst dye. Cells that can exclude the dye from their nucleus are termed SP cells and have been shown to have both increased tumorigenicity and self-renewal ability compared to non-side population (NSP) cells. However, dye efflux is a dynamic process, and the lack of specific criteria and guidelines
for delineating the SP fraction can lead to large variability between studies (Challen & Little, 2006). As such, a cell surface marker would be of important utility for sarcoma TPC research.

Self-renewal, a defining characteristic of TPCs, is associated with tumor recurrence (Kreso et al., 2014; Zhu et al., 2014). Expressions of genes that regulate self-renewal of normal stem cells are significant predictors of disease relapse (Ben-Porath et al., 2008; R. Liu et al., 2007; Merlos-Suarez et al., 2011), and the clinical outcome of patients with recurrent or metastatic sarcoma remains poor (Bielack et al., 2009). Thus, the inhibition of self-renewal in sarcoma TPCs may offer valuable targets of therapy.

Here, I used a flow cytometry screen to identify cell surface markers enriched on SP cells compared to bulk tumor cells. I found CD146 (also known as MCAM or MUC18), can reliability isolate TPCs in osteosarcoma and UPS. Importantly, I show that CD146+ and SP cells are independently tumorigenic and represent overlapping and distinct populations of sarcoma TPCs. Furthermore, pathway analysis reveals that Notch signaling pathway is activated in both of these two TPC populations in osteosarcoma. Treatment with a γ-secretase inhibitor significantly reduced the tumor growth and self-renewal capacity of human osteosarcoma in vivo.

2.3 Results

2.3.1 Flow cytometry screen of cell surface antigen in SP cells

SP cells are significantly enriched for TPCs in sarcomas (Murase et al., 2009; C. Y. Wang et al., 2012; Wu & Alman, 2008; Wu et al., 2007; Yang, Yan, Zhang, Li, & Luo, 2011). To identify cell surface marker(s) that might identify TPCs, I performed a flow cytometry screen to examine cell surface proteins expression enriched on the SP cells. Two primary human UPS samples and 1 primary bone sarcoma were obtained from the initial biopsy and processed into a single cell
The cells were stained with Hoechst33342 dye to sort for the SP cells. I screened 235 cell surface antibodies and found five markers that were enriched by greater than 4-fold in the SP population compared to the NSP cells. Specifically, on average 25.13% (±13.64% SEM) of the SP population expressed CD31, 29.51% (±15.01% SEM) expressed CD66, 11.02% (±3.46% SEM) expressed CD104, 36.34% (±24.27% SEM) expressed CD144 and 16.60% expressed CD146 (±8.10% SEM). In the NSP population, 1.37% (±0.52 SEM) expressed CD31, 0.75% (±0.24% SEM) expressed CD66, 2.47% (±1.63% SEM) expressed CD104, 0.95% (±0.62%) expressed CD144, and 4.62% (±1.47%) expressed CD146 (Table 2.1).

Table 2.1 Identification of candidate markers enriched on the surface of SP cells

<table>
<thead>
<tr>
<th>Marker</th>
<th>OS1 Marker+ in Non-SP cells (%)</th>
<th>OS1 Total SP (%)</th>
<th>OS1 Marker+ in SP (%)</th>
<th>UPS1 Marker+ in Non-SP cells (%)</th>
<th>UPS1 Total SP (%)</th>
<th>UPS1 Marker+ in SP (%)</th>
<th>UPS2 Marker+ in Non-SP cells (%)</th>
<th>UPS2 Total SP (%)</th>
<th>UPS2 Marker+ in SP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD31</td>
<td>1.1</td>
<td>0.44</td>
<td>20</td>
<td>2.38</td>
<td>1.63</td>
<td>50.9</td>
<td>0.64</td>
<td>0.56</td>
<td>4.5</td>
</tr>
<tr>
<td>CD66</td>
<td>0.96</td>
<td>0.88</td>
<td>27.4</td>
<td>1.03</td>
<td>1.61</td>
<td>56.5</td>
<td>0.26</td>
<td>0.75</td>
<td>4.64</td>
</tr>
<tr>
<td>CD104</td>
<td>4.1</td>
<td>0.9</td>
<td>7.56</td>
<td>11.2</td>
<td>2.24</td>
<td>14.8</td>
<td>0.83</td>
<td>0.48</td>
<td>2.37</td>
</tr>
<tr>
<td>CD144</td>
<td>0.43</td>
<td>0.47</td>
<td>8.7</td>
<td>2.19</td>
<td>2.52</td>
<td>84.8</td>
<td>0.23</td>
<td>0.36</td>
<td>5.83</td>
</tr>
<tr>
<td>CD146</td>
<td>3.15</td>
<td>0.49</td>
<td>5.84</td>
<td>6.09</td>
<td>2.08</td>
<td>24.1</td>
<td>32.6</td>
<td>0.39</td>
<td>0.11</td>
</tr>
</tbody>
</table>

OS: Osteosarcoma
2.3.2 An antibody to CD146 identified a population of sarcoma cells enriched in SP cells

I tested the ability of each of the 5 markers to enrich for a subpopulation of sarcoma cells with enhanced tumor-initiating capacity in vivo. The marker-positive cells were FACS sorted from a UPS and an osteosarcoma sample, and subcutaneously injected at dilutions ranging from 100, to 10,000 cells into NOD-scid IL2rnull (NSG) mice. After 20 weeks, the mice were sacrificed, and the tumors that formed were weighed and examined by histologic examination. CD31+, CD66+, CD104+ and CD144+ cells did not show higher tumor initiating ability compared to their respective marker negative populations or bulk tumor cells (data not show). In contrast, CD146+ cells enriched for TPCs close to 50-folds higher than CD146- cells.

I then analyzed the expression of CD146 using flow cytometry in an independent cohort of 10 human UPS samples and 5 human osteosarcoma samples. The mean percentage of SP and CD146 cells in UPS is 0.70% (±0.16%SEM) and 3.63% (±0.95%SEM) respectively, per tumor. The expression of CD146 was significantly enriched in the SP population compared to the NSP cells (P<0.001), with 53.2% (±9.51% SEM) of SP cells expressing CD146, and 2.98% (±0.90% SEM) of NSP cells expressing CD146 (Figure 2.1A, B, Table 2.2). I observed 1 UPS sample (UPS106) with higher percentage of CD146+ cells in the NSP populations than the SP population. This was likely due to the heterogeneity among different patient tumor biopsies. In osteosarcoma, the mean percentage SP and CD146+ cells are 0.68% (± 0.28 SEM) and 4.92% (±0.90 SEM) respectively. Similar to UPS, 49.37% (±15.48% SEM) of osteosarcoma SP cells express CD146, as compared to 4.73% (±0.87% SEM) of NSP (P<0.05, Figure 1B, Table 2.3). Overall, the enrichment of CD146+ cells in SP suggests that there is an overlapping population of CD146+ cells and SP cells.
Table 2.2 Abundance of CD146+ cells in human UPS

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Total SP (%)</th>
<th>Total CD146+ (%)</th>
<th>CD146+ in SP (%)</th>
<th>CD146+ in NSP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPS100</td>
<td>0.54</td>
<td>0.64</td>
<td>55.6</td>
<td>0.34</td>
</tr>
<tr>
<td>UPS101</td>
<td>0.9</td>
<td>0.57</td>
<td>33.3</td>
<td>0.27</td>
</tr>
<tr>
<td>UPS102</td>
<td>1.41</td>
<td>1.41</td>
<td>49.6</td>
<td>0.72</td>
</tr>
<tr>
<td>UPS103</td>
<td>1.34</td>
<td>8.32</td>
<td>74.6</td>
<td>7.42</td>
</tr>
<tr>
<td>UPS104</td>
<td>0.01</td>
<td>0.58</td>
<td>80</td>
<td>0.5</td>
</tr>
<tr>
<td>UPS105</td>
<td>0.046</td>
<td>4.4</td>
<td>65.2</td>
<td>4.37</td>
</tr>
<tr>
<td>UPS106</td>
<td>1.13</td>
<td>1.53</td>
<td>0.61</td>
<td>1.55</td>
</tr>
<tr>
<td>UPS107</td>
<td>0.66</td>
<td>7.03</td>
<td>75.8</td>
<td>6.57</td>
</tr>
<tr>
<td>UPS108</td>
<td>1.43</td>
<td>2.79</td>
<td>0.72</td>
<td>3.15</td>
</tr>
</tbody>
</table>

Table 2.3 Abundance of CD146+ cells in human osteosarcoma

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Total SP (%)</th>
<th>Total CD146+ (%)</th>
<th>CD146+ in SP (%)</th>
<th>CD146+ in NSP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS100</td>
<td>0.38</td>
<td>3.86</td>
<td>78.2</td>
<td>3.56</td>
</tr>
<tr>
<td>OS101</td>
<td>0.28</td>
<td>6.13</td>
<td>64.2</td>
<td>5.97</td>
</tr>
<tr>
<td>OS102</td>
<td>0.4</td>
<td>7.86</td>
<td>80</td>
<td>7.57</td>
</tr>
<tr>
<td>OS103</td>
<td>1.8</td>
<td>3.55</td>
<td>7.74</td>
<td>3.45</td>
</tr>
<tr>
<td>OS104</td>
<td>0.54</td>
<td>3.16</td>
<td>16.7</td>
<td>3.11</td>
</tr>
</tbody>
</table>

The location of CD146+ cells in UPS and osteosarcoma was visualized using immunofluorescence. Since CD146 is also a marker of pericytes (Crisan et al., 2008), I stained
frozen primary patient tumor sections with CD31 and CD146 to distinguish between blood vessels and tumor cells. I found that CD146+ cells were present both near blood vessels and within the tumors away from the vasculature, indicating that a population of tumor cells, unrelated to the vasculature are expressing CD146 (Figure 2.2).

Figure. 2.1 CD146 expression is enriched on the surface of SP cells in human UPS and osteosarcoma. A. Representative flow cytometry analysis of SP, NSP, and enrichment of CD146 on SP cells in human sarcoma. The NSP is labeled with a box in the upper right quadrant, and SP is in the lower left quadrant. Treating the cells with verapamil inhibits Hoechst dye exclusion and is used as a negative control for SP analysis. Expression of CD146 is gated on the SP and NSP.
cells. B. Analysis of CD146 expression on SP and NSP cells in 10 primary human UPS samples and 5 primary human osteosarcoma samples, showing CD146 is significantly enriched on the sarcoma SP cells. *, P<0.05; ** <0.01.

![Image](image_url)

**Figure 2.2 Localization of CD146+ cells in human osteosarcoma and UPS tumors.** Fresh frozen patient biopsies are stained for CD31 (green), CD146 (red) and DAPI (blue).

### 2.3.3 CD146+ cells show increased tumorigenicity

The ability of CD146+ cells to initiate tumors was tested in 5 additional primary human UPS and 5 primary osteosarcoma samples using limiting dilution xenograft assay. As few as 10 CD146+ cells in UPS formed tumors at high frequency. In contrast, only a few mice developed tumors when injected with 1,000 cells from the CD146- fraction (Table 2.4). Similarly, the CD146+ cells in osteosarcoma exhibited substantially higher tumor-forming capacity compared to the CD146- cells (Table 2.5).
Table 2.4. Serial transplantation of CD146+ cells in human UPS tumors

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Cell dose</th>
<th>No. of sample tested</th>
<th>No. primary mice with tumors/total no. injected</th>
<th>No. secondary mice with tumors/total no. injected</th>
<th>No. tertiary mice with tumors/total no. injected</th>
<th>Total no. of mice with tumors(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD146+</td>
<td>1x10</td>
<td>5</td>
<td>4/19</td>
<td>2/12</td>
<td>1/8</td>
<td>7/39 (17.9)</td>
</tr>
<tr>
<td></td>
<td>1x10²</td>
<td>5</td>
<td>8/30</td>
<td>3/12</td>
<td>2/12</td>
<td>13/54 (24.1)</td>
</tr>
<tr>
<td></td>
<td>1x10³</td>
<td>5</td>
<td>13/24</td>
<td>7/10</td>
<td>3/12</td>
<td>23/46 (50)</td>
</tr>
<tr>
<td></td>
<td>1x10⁴</td>
<td>5</td>
<td>14/25</td>
<td>10/10</td>
<td>8/8</td>
<td>32/43 (74.4)</td>
</tr>
<tr>
<td>CD146−</td>
<td>1x10</td>
<td>5</td>
<td>0/22</td>
<td>0/14</td>
<td>0/8</td>
<td>0/34 (0)</td>
</tr>
<tr>
<td></td>
<td>1x10²</td>
<td>5</td>
<td>0/30</td>
<td>0/12</td>
<td>0/12</td>
<td>0/54 (0)</td>
</tr>
<tr>
<td></td>
<td>1x10³</td>
<td>5</td>
<td>2/24</td>
<td>3/10</td>
<td>0/12</td>
<td>5/46 (10.9)</td>
</tr>
<tr>
<td></td>
<td>1x10⁴</td>
<td>5</td>
<td>13/26</td>
<td>7/10</td>
<td>6/8</td>
<td>26/44 (59.1)</td>
</tr>
</tbody>
</table>

Table 2.5. Serial transplantation of CD146+ cells in human osteosarcoma tumors

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Cell dose</th>
<th>No. of sample tested</th>
<th>No. primary mice with tumors/total no. injected</th>
<th>No. secondary mice with tumors/total no. injected</th>
<th>No. tertiary mice with tumors/total no. injected</th>
<th>Total no. of mice with tumors(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD146+</td>
<td>1x10</td>
<td>5</td>
<td>2/30</td>
<td>2/40</td>
<td>1/12</td>
<td>5/82 (6.0)</td>
</tr>
<tr>
<td></td>
<td>1x10²</td>
<td>5</td>
<td>6/28</td>
<td>10/46</td>
<td>2/12</td>
<td>18/86 (20.9)</td>
</tr>
<tr>
<td></td>
<td>1x10³</td>
<td>5</td>
<td>13/30</td>
<td>21/46</td>
<td>3/10</td>
<td>37/86 (43.0)</td>
</tr>
<tr>
<td></td>
<td>1x10⁴</td>
<td>5</td>
<td>18/22</td>
<td>20/26</td>
<td>8/10</td>
<td>46/58 (79.3)</td>
</tr>
<tr>
<td>CD146−</td>
<td>1x10</td>
<td>5</td>
<td>0/30</td>
<td>0/44</td>
<td>0/12</td>
<td>0/86 (0)</td>
</tr>
<tr>
<td></td>
<td>1x10²</td>
<td>5</td>
<td>0/30</td>
<td>0/52</td>
<td>0/12</td>
<td>0/94 (0)</td>
</tr>
<tr>
<td></td>
<td>1x10³</td>
<td>5</td>
<td>1/28</td>
<td>1/43</td>
<td>0/10</td>
<td>2/81 (2.5)</td>
</tr>
<tr>
<td></td>
<td>1x10⁴</td>
<td>5</td>
<td>15/18</td>
<td>8/37</td>
<td>6/10</td>
<td>29/65 (44.6)</td>
</tr>
</tbody>
</table>
2.3.4 CD146\(^+\) cells possess self-renewal and generate tumors that recapitulate parent tumor

A fundamental characteristic of TPCs is the ability to self-renew and to recapitulate the histological characteristics of the parent tumor (Iso, Kedes, & Hamamori, 2003). As such, I serially injected CD146\(^+\) and CD146\(^-\) cells from primary xenografts into fresh NSG mice. In secondary and tertiary transplants, CD146\(^+\) cells continued to exhibit enhanced tumor-forming ability compared to CD146\(^-\) cells (Table 2.4 and 2.5). Furthermore, hematoxylin and eosin staining of tumors xenograft derived from CD146\(^+\) cells resembled the patient tumor in both UPS and osteosarcoma (Figure 2.3A).

To determine the frequency of TPC enrichment in CD146\(^+\) fraction, I used the Extreme Limiting Dilution Analysis (ELDA) algorithm on data from the secondary transplants (Hu & Smyth, 2009). For UPS, the frequency of TPCs in the CD146\(^+\) fraction is 1/555 cells, compared to 1/17002 cells in CD146\(^-\) fraction, indicating a 30.6-fold enrichment (P=3.60e-36, Figure 2.3B). For osteosarcoma, the TPC frequency is 1/2830 cells and 1/68375 cells in CD146\(^-\) fraction, indicating a 24.2-fold enrichment (P=5.13e-29, Figure 2.3B).

Taken together, these data show that CD146 reliably identifies a population of cells in sarcoma that are enriched for TPCs, which are characterized by enhanced tumorigenicity, the ability to self-renewal, and by the ability to initiate tumors that resembles the parent tumor.
Figure 2.3 CD146 enriches for sarcoma TPCs. A. Original patient sarcoma samples obtained from biopsy, primary and secondary xenograft tumors derived from the CD146+ cells are formalin-fixed, paraffin-embedded and stained with hematoxylin and eosin (H&E). The grafted tumors are identical histologically compared to the original patient tumor. B. Graphical representation of in vivo limiting dilution assay, which compares TPC frequency in CD146+ cells and CD146− cells. Tumor initiation data from secondary injection of CD146+ and CD146− cells into NSG mice are used to determine the enrichment of TPCs in each population. CD146+ cells significantly enrich for TPCs in UPS and osteosarcoma compared to CD146+ cells.
2.3.5 \textbf{CD146}+ enriches for a distinct TPC population from SP cells

Since CD146 was identified based on its enrichment in the SP population, I examined whether if the tumorigenic and self-renewal capacity of CD146+ cells are dependent on the SP population contained within the CD146 fraction. I isolated the non-SP (NSP) population following Hoechst dye staining and sorted for CD146+ and CD146− fractions. The sorted cells were injected into NSG mice at different dilutions. I found that the NSP\textit{CD146}+ population was significantly more tumorigenic than the NSP\textit{CD146}− population over multiple transplants (Table 2.6, \(P = 9.58 \times 10^{-13}\)). Therefore, despite the overlap between SP and CD146+ population, CD146+ cells depleted of SP remain enriched for tumor propagating capacity, suggesting that CD146 enrich for a subpopulation of TPCs in sarcoma that both overlap with, but also distinct from SP cells.

2.3.6 \textbf{Common signaling pathways are activated in CD146}+ and SP populations

To further examine similarities and differences between the cell populations, gene expression profiling was undertaken to compare the various cell populations. This data was also used to identify cell-signaling pathways that are differentially regulated in the same way in both CD146+ and SP cells. Since previous studies examined pathways differentially regulated in TPC populations in UPS, I focused this work on osteosarcomas (C. Y. Wang et al., 2012). Four separate populations of CD146+, CD146−, SP and NSP were sorted from 3 osteosarcoma samples, and expression profiles were analyzed. Using a 1.5-fold change and a \(P<0.05\) as thresholds, I identified 3763 differentially expressed genes between CD146+ versus CD146− cells and 757 differentially expressed genes between SP versus NSP cell. Using Gene Set Enrichment Analyses (GSEA), I performed pathways analysis for CD146+ versus CD146− cells and SP versus NSP cells. This analysis identified multiple targetable pathways that are similarly regulated in the SP and CD146+
populations (Figure 2.5). In particular, TGF-β and Notch signaling pathways, both known to regulate stem cell behavior and contribute to tumorigenesis (Tao et al., 2014; C. Y. Wang et al., 2012; Watabe & Miyazono, 2009) were upregulated (Figure 2.4A and 4B). To validate this, the expression of known TGF-β and Notch signaling target genes were measured by qPCR in SP and CD146+ relative to their negative counterparts (Figure 2.4C). Comparison between CD146+ and CD146- cells showed significant upregulation of CTGF, c-JUN, PAI-1, HEY1, and HEY2. With the exception of CTGF, these target genes were also significantly upregulated in the SP population (Iso et al., 2003; Ranganathan et al., 2007).

Table 2.6. Serial transplantation of different cell fractions in human UPS

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Cell dose</th>
<th>No. of primary mice with tumors/total no. injected</th>
<th>No. of secondary mice with tumors/total no. injected</th>
<th>Total number of mice with tumors</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSP CD146+</td>
<td>1x10</td>
<td>0/12</td>
<td>3/6</td>
<td>3/18 (16.7)</td>
</tr>
<tr>
<td></td>
<td>1x10²</td>
<td>9/12</td>
<td>6/6</td>
<td>15/18 (83.3)</td>
</tr>
<tr>
<td></td>
<td>1x10³</td>
<td>7/7</td>
<td>4/4</td>
<td>11/11 (100)</td>
</tr>
<tr>
<td></td>
<td>1x10⁴</td>
<td>3/3</td>
<td>4/4</td>
<td>7/7 (100)</td>
</tr>
<tr>
<td>NSP CD146-</td>
<td>1x10</td>
<td>0/12</td>
<td>0/6</td>
<td>0/18 (0)</td>
</tr>
<tr>
<td></td>
<td>1x10²</td>
<td>2/12</td>
<td>1/6</td>
<td>3/18 (16.7)</td>
</tr>
<tr>
<td></td>
<td>1x10³</td>
<td>6/10</td>
<td>2/4</td>
<td>8/14 (57.1)</td>
</tr>
<tr>
<td></td>
<td>1x10⁴</td>
<td>8/8</td>
<td>4/4</td>
<td>12/12 (100)</td>
</tr>
</tbody>
</table>

This analysis also identified pathways that were differentially regulated between SP and CD146+ cells. Specifically, FGFR, calmodulin, CREB and phospholipase C associated signaling were enriched in SP cells but not CD146+ cells (Figure 2.5). On the other hand, extracellular matrix remodeling, cell migration, hypoxia response and angiogenesis associated pathways were enriched only in the CD146+ cells (Figure 2.5). These differences are consistent with the CD146+
and SP having distinct populations.

Figure 2.4 Notch and TGF-β signaling pathways are significantly upregulated in SP and CD146+ cells. GSEA analysis comparing SP versus NSP cells (A) and CD146+ versus CD146- (B) cells for enrichment of Notch and TGF-β signaling pathway gene sets. C. Quantification of Notch and TGF-β signaling pathway target genes in SP versus NSP cells and CD146+ versus CD146- cells by qPCR. *, P<0.05; ** <0.01.
Figure 2.5 Pathway analyses of SP versus NSP and CD146$^+$ versus CD146$^-$ cells. GSEA results visualized by cytoscape showing pathways enriched in A. SP versus NSP cells and B. CD146$^+$
versus CD146\(^{-}\) cells. Nodes represent significantly enriched pathways and green lines connect overlapping pathways.

**Figure 2.6. Pharmacological inhibition of Notch signaling significantly reduces osteosarcoma growth and self-renewal.**

**A.** Tumor weight of mice bearing human osteosarcoma following 3 weeks of DAPT compared to vehicle treatment. **B.** Relative gene expression of Notch target genes in DAPT treated mice compared to vehicle treated mice. **C.** Tumors treated with DAPT or vehicle are digested and injected into fresh NSG mice at known concentrations to evaluate the effect of DAPT on self-renewal. The number of secondary tumors is counted after 16 weeks of observation.

### 2.3.7 Inhibition of Notch signaling reduces tumor growth and self-renewal

The self-renewal capacity of TPCs has been hypothesized to contribute to tumor recurrence (Kreso et al., 2014). To determine if pathways differentially regulated in both SP and CD146 positive cells would also drive self-renewal, I examined one such pathway, Notch signaling using
a pharmacological inhibitor. This signaling pathway is a regulator of mesenchymal stromal cells (MSCs) differentiation and has previously been shown to drive self-renewal in UPS (Hilton et al., 2008; Osathanon, Subbalekha, Sastravaha, & Pavasant, 2012; C. Y. Wang et al., 2012). Since Notch was recently studied for self-renewal capacity in UPS, I examined its function in osteosarcoma. NSG mice bearing primary human osteosarcoma cells were treated with DAPT, a γ-secretase inhibitor that targets Notch signaling. After 3 weeks of treatment, mice that received DAPT showed significantly smaller tumor sizes compared to the vehicle treated mice, and the expression of Notch target genes were reduced in the treated tumors (Figure 2.6A, B). Next, I compared the self-renewal potential of these cells by examining the tumor-initiating potential of DAPT and vehicle treated cells in fresh NSG mice. DAPT treated cells showed significantly reduced ability to initiate tumors (Figure 2.6 C). Thus, Notch signaling drives self-renewal in osteosarcoma, and its inhibition, using agents such as DAPT may serve as a potential therapy against sarcoma self-renewal.

2.4 Discussion

In this study, flow cytometry screen of cell surface markers on sarcoma SP cells and functional validation by serial transplantations identified CD146 as a cell surface marker that enriches for tumor propagating capacity in UPS and osteosarcoma. This is the first cell surface marker identified using primary sarcoma samples with complex karyotype. Previously identified cell markers in sarcomas have not been tested in human patient samples using the gold standard of in vivo serial transplantation assay for evaluating self-renewal. Since CD146 can be readily detected with various approaches using antibodies, it has significant advantages compared to the SP analysis, a method that is technically challenging, and as such can yield variable cell
CD146 is a cell-adhesion molecule that belongs to the immunoglobulin superfamily (Lehmann et al., 1987). It was first identified as a specific antigen human malignant melanoma, and its expression was subsequently shown to associate with poor survival, tumor progression, and metastasis in other cancers such as breast, lung, and prostate (Oka, Uramoto, Chikaishi, & Tanaka, 2012; Shih, Elder, Speicher, Johnson, & Herlyn, 1994; Zabouo et al., 2009). Certain MSC populations, such as pericytes, also express CD146 (Crisan et al., 2008). I identified CD146 based on its expression in the SP cells, a population that has been previously shown by multiple studies to be highly enriched for TPCs (Murase et al., 2009; C. Y. Wang et al., 2012; Wu et al., 2007; Yang et al., 2011; Yang et al., 2010). In our study, CD146+ cells in the tumors are more abundant with a mean 3.61% in UPS and 4.91% in osteosarcoma compared to 0.68% and 0.71% of SP cells, respectively. The ability of CD146+ to initiate tumors over multiple transplantations suggests it indeed enriches for TPCs with self-renewal ability. However, in many human cancers where TPCs has been identified, the percentage of these cells is generally less than 0.04%(Ishizawa et al., 2010). In theory, a single TPC is sufficient to generate tumors. The reduced ability of CD146+ cells to form tumors at the 10-cell dose suggest that while this population is enriched for TPCs but may also contain other cells such as transit amplifying cells.

Our serial transplantation data indicate that CD146+ cells still possess TPC properties in the absent of SP cells. This suggests that the TPC population in UPS and osteosarcoma may be heterogeneous, and NSP/CD146+ cells are a distinct TPC population with some overlap in the SP population. The heterogeneity of TPCs also suggests that these populations may respond differently to targeted therapy. Indeed, Kreso et al found that different clonal populations with TPC properties in colorectal cancer vary in their proliferative potential, self-renewal ability and
response to therapy (Kreso et al., 2013). As such, identification of common pathways in the heterogeneous TPC populations may offer one approach to identify effective targets of therapy.

Using gene expression profiling and pathway analysis, I show that Notch signaling is activated in both SP and CD146+ cells. Treatment of patient derived xenograft in mice with DAPT, a known inhibitor of Notch signaling, significantly reduced primary tumor growth. More importantly, DAPT inhibited the tumors from growing back in secondary transplants, even with treatment withdrawal, demonstrating that this treatment targets self-renewal. The importance of Notch signaling in osteosarcoma initiation is further substantiated by recent work from Tao et al., which showed that over-expression of Notch-1 Intracellular Domain in MSCs and osteoblast progenitor cells drives osteosarcoma formation in mice (Tao et al., 2014). In addition, I found other self-renewal pathways activated in SP and CD146+ cells. Hedgehog and YAP signaling pathways may be activated in CD146+ cells. TGF-β signaling, which participates in stem-cell maintenance is upregulated in both SP and CD146+ cells (Watabe & Miyazono, 2009). Wnt activation is observed in SP cells but was not statistically significant in our data. This suggests that self-renewal pathways may be heterogeneous in different tumor samples. Furthermore, our pathway analysis showed that metabolic pathways, especially, glucose metabolism are upregulated in the SP and CD146+ cells. Other studies have shown that TPCs often exhibit increased glucose uptake compared to bulk tumor cells, contributing to the survival advantage of these cells (Dong et al., 2013; Shen, Wang, Hsieh, Chen, & Wei, 2015; Vlashi et al., 2011). Future investigations to target differentially regulated pathways in TPCs may have therapeutic value.

The tumor microenvironment is essential in regulating the properties of TPCs(Sneddon & Werb, 2007). Emerging evidence suggest that stromal cells, non-TPC tumor cells, and the extracellular matrix provide important signaling molecules that supports TPCs growth, self-
renewal, and protects TPCs from immunosurveillance (Kreso & Dick, 2014; Plaks, Kong, & Werb, 2015). A potential limitation of the xenograft model is that certain aspects of the tumor microenvironment may not be completely recapitulated in the recipient animal. Developments of more immunodeficient mice, and humanized mice that expresses signaling molecules found in human tumors may allow more accurate assessment of TPCs. Alternatively, visualization of TPCs in transgenic mouse models of sarcomas may allow us to further characterize the role of TPCs in the native environment.

Our study identifies CD146 as a cell surface marker for a tumor population enriched in TPC properties in primary UPS and osteosarcoma. Using in vivo model, I demonstrate that CD146+ tumor cells show increased tumorigenicity, self-renewal ability, and can initiate tumors that resemble the primary patient tumor. Furthermore, the tumorigenic potential of CD146+ cells is also independent of the SP fraction, a known TPC enriched population, suggesting that there are heterogeneous populations of sarcomas TPCs. Using gene expression profiling and pathway analysis of both CD146+ and SP cells, I identified Notch signaling is a potential target for reducing osteosarcoma self-renewal. Targeting pathways activated in TPC subpopulations may be therapeutic for sarcoma.

2.5 Methods

Primary Tumor Samples

Human undifferentiated pleomorphic sarcoma (UPS) and osteosarcoma tumor tissues were obtained at the time of initial biopsy, prior to any therapy. A musculoskeletal pathologist verified the diagnosis of the tumors. To generate single-cell suspensions, the samples were mechanically dissociated into small pieces, and enzymatically digested with a mixture of 10mg/mL of
collagenase IV (Worthington), 2.4U/ml of Dispase (Becton Dickinson), and 0.05% trypsin (Wisent) for 45-60 minutes at 37°C, followed by filtering through a 70μm strainer as previously reported (C. Y. Wang et al., 2012; Wu et al., 2007). Hematopoietic cells were excluded with whole blood lysis buffer (Life Technologies). Afterwards, the single cell suspensions were stained with anti-human CD45-PE-Cy7 antibody (1:200, Becton Dickinson) to deplete the immune cells via flow cytometry activated cell sorting (FACS).

Flow Cytometry

Side population (SP) cells were collected as previously described (Wu et al., 2007). Briefly, single-cell suspensions were treated with 2.5 mg/mL of Hoechst 33342 dye (Sigma) alone, or in combination with 50 mmol/L of verapamil (Sigma) as a negative control, for 90 minutes at 37°C. SP, cells were identified using dual wavelength analysis (blue, 424–444nm; red, 675 nm) after excitation with 350 nm UV light (MoFlowXDP). Staining with CD45 antibody (BD Pharmingen) was used to eliminate hematopoietic cells. To sort for CD146+ cells, processed tumor cells were stained with anti-human CD146-PE conjugated antibody at 1:100 dilution for 30 minutes at 4°C (BD Pharmingen). Murine cells, which makes up the tumor stroma in the xenografts were excluded from staining with a biotin conjugated anti-mouse H-2k^d antibody (BD Pharmingen) streptavidin PE-Cy7 conjugate (Invitrogen) at 1: 1,000 for 30 minutes at 4°C. H-2k^d+ cells were removed from analysis during FACS. Since the stromal cells are likely derived from the mouse, removing the H-2k^+ cells allow us to deplete the stromal cells from the tumor (C. Y. Wang et al., 2012; Wu et al., 2007). In all flow cytometry experiments, cells were counterstained with 1 mg/mL of propidium iodide (PI; Molecular Probes) and the dead cells were removed from the analysis.
Flow-Cytometry Cell Surface Antigen Screen

The SP and NSP suspensions were incubated on ice in flow cytometry buffer (FC buffer: Hanks balanced salt solution with 1% BSA, 2mM EDTA). The cells were aliquoted into round-bottom 96-well plates containing 235 fluorochrome-conjugated cell-surface targeted antibodies (i.e. each of 235 wells contained a different antibody, Table 2.6). Antibodies were labeled with PE, FITC or APC. Wells containing buffer only were included as negative controls. Cells were suspended at a concentration to achieve a minimum cell number of 50,000 cells per well in a final volume of 200 ul per well. The final antibody dilution was 1:50 for all antibodies. Once cells were added to the wells, plates were incubated on ice in the dark for 20 minutes. Plates were centrifuged to pellet cells, buffer aspirated, and pellets were washed twice with 200 ul of FC buffer. Finally, pellets were suspended in 100 ul of FC buffer containing 1 ug/ml of propidium iodide to allow exclusion of dead cells. Cells were then analyzed on a BD LSRII equipped with a high throughput sampler. A minimum of 10,000 events per well were collected on FACSDiva software and resulting FCS 3.0 files exported to FlowJo version 9.3 for analysis. Dead cells, debris, cell doublets and CD45+ immune cells were excluded from the analysis, after which side-population and bulk tumor cells were analyzed for expression of each of the 235 markers. Enrichment of a cell surface marker was determined by comparing the percent of marker-positive cells in the SP population to the percent of marker-positive cells in bulk tumor cells. Markers with an enrichment of ≥ 4-fold in two of the three samples analyzed were selected for testing in vivo.

Limiting Dilution Assay

Individual cells sorted from flow cytometry according to various markers including CD146 were suspended in PBS. Murine cells from xenografts were excluded during FACS by staining with a biotin-conjugated anti-mouse H-2k antibody (BD Pharmingen). The sorted cell counts from
flow cytometry were manually confirmed with a hemocytometer and then serially diluted in PBS to achieve cell range between $10^{-1}$ to $10^5$ cells/100μl. For each injection, the cells were mixed with an equal volume of ice-cold Matrigel (Becton Dickinson) and subcutaneously injected into 6- to 8-week-old NOD-scid IL2γnull (NSG) mice. After injection, the mice were observed for 24 weeks and the tumors were dissected. The tumors were weighed using an analytical balance and were examined histologically. Tumor-propagating cell frequency was calculated based on extreme limiting dilution data from the secondary tumor transplantations as described in (Hu & Smyth, 2009). The model follows standard general linearized models to compare the frequency of TPCs in one or more populations and allows for one-sided confidence intervals if 0% or 100% positive responses are observed (Hu & Smyth, 2009; Nelder & Wedderburn, 1972). The model can be accessed online at http://bioinf.wehi.edu.au/software/elda/ from Walter and Eliza-Hall Institute (Hu & Smyth, 2009).

**Histology and immunofluorescence**

Tumor biopsies from patients and mice xenografts were formalin fixed and paraffin embedded. Tumor sections were cut at 8μm and stained with hematoxylin and eosin following standard procedure and observed in a blinded manner. At least 5 sections from each tumor sample were analyzed. For immunofluorescence, fresh tumors biopsies from patients were embedded in Tissue Tek O.C.T compound (Fisher Scientific) and snap frozen in dry ice. Sections of 8μm were cut and blocked with 10% donkey serum, 2%BSA, in 1X PBST for 1 hour at room temperature. Tissue slides were incubated with CD31 antibody (Abcam) at 1:50 dilution for 1 hour. This was followed by secondary antibody labeling with Alexa488 goat anti-rabbit (Life Technologies) at 1:1000 dilutions for 30 minutes. CD146 (BD Pharmagen) were diluted at 1: 5000 and incubated
on the tissue slides for 30 minutes, followed by Alexa594 donkey anti-mouse antibody at 1:1000 dilutions for 30 minutes. The slides were mounted with mounting media containing DAPI (Vectashield, Vector Laboratories Inc) and imaged.

**Gene expression profiling analyses**

Three independent human osteosarcoma xenografts were sorted for SP, NSP, CD146+ and CD146− fractions. The stromal cells were removed by staining with anti-mouse H-2k antibody (BD Pharmingen). RNA samples were extracted and converted to cDNA using Ovation RNA Amplification System V2 (Nugen) following manufacture’s protocol. The cDNA was analyzed using Illumina HT-12 v4 platform following standard protocol. Results were analyzed using R (version 3.2.0) with the LIMMA (linear models for microarray data) package (Smyth, 2004; Zhang, 2007). I examined the raw data with a Normexp by Control (neqc) algorithm pre-processing strategy described by Shi et al using R(Shi, Oshlack, & Smyth, 2010). This approach includes background correction, quantile normalization and log2 transformation to the raw data(Bolstad, Irizarry, Astrand, & Speed, 2003; Shi et al., 2010). Differentially expressed genes were identified using LIMMA as described in (M. E. Ritchie et al., 2015). Genes with a fold-change >1.5 and p<0.05 from the LIMMA analyses were considered significant. The microarray data was submitted to the NCBI GEO database with the identifier GSE63390.

**Gene Set Enrichment Analysis (GSEA)**

Pathway analyses were performed using Gene Set Enrichment Analysis (GSEA) with parameters set to 2,000 gene set permutations and gene sets size between 8 and 500 (Subramanian et al., 2005). Gene sets were obtained from KEGG, MsigDB-c2, NCI, Biocarta, IOB, Netpath, Human Cyc, Reactome and the Gene Ontology (GO) databases (Kanehisa & Goto, 2000; Merico,
Isserlin, Stueker, Emili, & Bader, 2010). An enrichment map was generated using Cytoscape with parameters set for a nominal P value of <0.005, FDR <0.25, and the Jaccard coefficient equal to 0.5 (Saito et al., 2012).

**Quantitative real-time reverse transcription PCR (qPCR)**

Analysis of gene expression using qPCR was performed as previously described (Rugg-Gunn et al., 2012). The reactions were performed on ViiA Real-Time PCR System (Applied Biosystems) with Taqman Fast Universal PCR Master Mix (Life Technologies). The primer sets for all the genes was purchased from Life Technologies. The expression level of the genes was calculated using standard ΔΔCt method, with GAPDH as internal control. All reactions were performed with at least 3 replicates.

**In vivo DAPT treatment**

Eight to 10 weeks old NSG mice were subcutaneously injected on the left flank at 200,000 cells dissolved in Matrigel. After 6 weeks the animals were treated with N-[(3,5-Difluorophenyl)acetyl]-L-alanyl-2-phenyl glycine-1,1-dimethylethyl ester (DAPT) dissolved in corn oil with 5% DMSO 100mg/kg/day (Selleck Chemicals, Texas) or vehicle via intraperitoneal injection for 22 days. The mice were euthanized and tumors were harvested. Each tumor was dissected free from surrounding tissues and weighed. The tumors were then digested into single cell suspension, and injected into fresh NSG mice at limiting dilutions, as described above. Secondary transplants were observed for up to 16 weeks.

Statistical analyses

Statistically analyses were performed using GraphPad Prism v6. All results are representative
of at least \( n=3 \). The data were shown as mean ± standard error of mean (SEM), unless otherwise stated, and was calculated using Prism. Statistical significance was calculated using two-tailed, unpaired Student’s t-test. A P value <0.05 is considered significant.
CHAPTER 3

TUMOR PROPAGATING CELLS IN SARCOMA IS A DYNAMIC POPULATION

3.1 Authors Contribution

I performed the experiments and the data analyses with the help of lab members and collaborators. All experiments and data analyses were conducted in the laboratory of Dr. Benjamin Alman, who helped conceive and supervise the project.
3.2 Introduction

Cancer is composed of multiple cell populations that differ in genetic and functional properties. In most cancers, a subpopulation of tumor-propagating cells (TPCs) with stem-cell like characteristics have been identified. These cells are defined by enhanced tumorigenic capacity, self-renewal, and the ability to give rise to other cancer cells that make up a tumor. Functionally, TPCs are identified by their ability to continuously propagate tumors following serial transplantation into animal hosts (O'Brien, Kreso, & Jamieson, 2010). TPCs has also been shown to display increased therapy resistance as well (Awad et al., 2010; Zhu et al., 2014). As such, these cells are believed to drive tumor progression and disease relapse; and hence eradication of tumors may depend on the selective elimination of TPCs. Indeed, many studies have attempted to characterize and target TPCs to impede tumor progression (Basu-Roy et al., 2015; Perry et al., 2014; Sato et al., 2016; C. Y. Wang et al., 2012; Zhu et al., 2014). However, implicit in these works is the unidirectional relationship between TPCs and non-TPCs. Interestingly, recent studies in several normal tissues suggest that stem cells may exist as a dynamic population, where their differentiated progenies can give rise to stem cells. For instance, following chemical injury in the liver, differentiated hepatocytes can convert to bipotential progenitors (Yanger et al., 2013). Similarly, in the lung epithelium, loss of basal stem cells can result in dedifferentiation of luminal secretory cells to basal stem cells (Tata et al., 2013). While cellular dedifferentiation and lineage plasticity are well-recognized phenomenon during tumor initiation, whether TPCs are a dynamic or stable state remains to be fully explored.

Sarcomas are a diverse group of tumors from the mesenchymal lineage. They can be broadly categorized into the bone and soft tissue sarcomas. Undifferentiated pleomorphic sarcoma (UPS) is the most common soft tissue sarcoma in adults. This tumor type is composed of a
heterogeneous population of cells with distinct cytologic features, epigenetic changes, and biologic behaviors (Francis et al., 2005; Orndal, Rydholm, Willen, Mitelman, & Mandahl, 1994; Reya, Morrison, Clarke, & Weissman, 2001). Previously, others and I have demonstrated that TPCs in human sarcomas can be isolated by the side population (SP) assay (Murase et al., 2009; Wu et al., 2007; Yang et al., 2010). In this study, I sought to investigate whether SP cells is a stable or dynamic population of sarcoma cells using an autochthonous mouse model of UPS. I demonstrate that the SP population is a dynamic or transient state and can be compensated by Non-SP (NSP). Depletion of this population has limited impact on tumor growth and self-renewal.

3.3 Results

3.3.1 Side Population Identifies Polyclonal TPCs

I generated a mouse model of sarcoma in which tumor cells are differentially labeled with homozygous R26R-confetti reporter (hereafter referred to as KPCC). Tumor formation in the KPCC model is driven by conditional activation of Kras$^{G12D}$ mutation and homozygous deletion of $p53$ (Kirsch et al., 2007). Upon injection of adenovirus expressing Cre into the gastrocnemius muscle, temporally and spatially restricted tumors resembling undifferentiated pleomorphic sarcoma (UPS) form at the site of the injection, and the different tumor clones are randomly labeled by one of 8 fluorescent reporter combinations (Figure 3.1A-B). Multiple studies have shown that fluorescent labeling by R26R-confetti reporters are permanent and stable and is suitable for tracing distinct cell populations over time. (Baggiolini et al., 2015; Maddipati & Stanger, 2015; Snippert et al., 2010). When KPCC tumor cells are orthotopically transplanted into immunocompromised hosts, multiple clones are able to engraft and grow, suggesting that the differently labeled cells have tumorigenic capacity (Figure 3.1C).
Figure 3.1. KPCC model of sarcoma. A. Schematic of KPCC mouse model. B. H&E and confocal image of a representative KPCC sarcoma. Scale 500μm, inset scale 100μm. C. H&E
and confocal image of a representative KPCC sarcoma orthotopically transplanted into Foxn1\textsuperscript{nu/nu} nude mouse. Scale 1mm, inset scale 100μm.

Figure 3.2. SP cells are polyclonal in KPCC sarcomas. A. FACS gating of SP and NSP cells in KPCC sarcoma. B. Expression R26R-Confetti florescent reporters in SP and NSP cells. C. Mean percentage of SP cells in KPCC sarcoma (n=11).

Since side population (SP) cells that can exclude the Hoechst 33342 dye and are known to enrich for TPCs in several types of human sarcoma (Wei et al., 2015; Wu & Alman, 2008), I thought to test the TPC potential of SP cells in the KPCC model. Primary KPCC tumor cells are FACS analyzed to identify SP and NSP cells. Overall, the SP population represented 10.33% (±2.36%SEM) of total tumor cells (Figure 3.2A-B). Furthermore, both SP cells and NSP cells consisted of cells expressing different fluorescent reporters, suggesting that both populations are labeled (Figure 3.2B). Since the degree of immunodeficiency in animal hosts has been shown to influence the ability of tumor cells to engraft (Quintana et al., 2008), I orthotopically implanted the SP and NSP cells into two types of immunocompromised mice: Foxn1\textsuperscript{nu/nu} and NOD-SCID IL2rγ\textsuperscript{null} (NSG) mice (Figure 3.3, Figure 3.4). Serial transplants at limiting dilutions and analysis of the results using Extreme Limiting Dilution Analysis (ELDA) algorithms showed that the SP cells are significantly more tumorigenic than NSP cells in both models (Figure 3.3B, 3.4B) (Hu &
Fewer cells are required to form tumors in the NSG mice, which is consistent with the weaker immune system of these mice. Nonetheless, the difference in TPC enrichment between SP and NSP populations is maintained (Figure 3.4). Taken together, this suggests that SP cells are enriched for sarcoma TPCs that are capable of tumor initiation and self-renewal.

<table>
<thead>
<tr>
<th>Number of Primary Tumors</th>
<th>Cell Dose</th>
<th>First Transplant Tumors/Mice (SP)</th>
<th>First Transplant Tumors/Mice (NSP)</th>
<th>Second Transplant Tumors/Mice (SP)</th>
<th>Second Transplant Tumors/Mice (NSP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>50</td>
<td>3/10</td>
<td>0/10</td>
<td>3/6</td>
<td>1/6</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>7/15</td>
<td>0/10</td>
<td>8/10</td>
<td>4/10</td>
</tr>
<tr>
<td>4</td>
<td>200</td>
<td>10/17</td>
<td>2/19</td>
<td>8/10</td>
<td>4/10</td>
</tr>
<tr>
<td>4</td>
<td>500</td>
<td>12/15</td>
<td>3/15</td>
<td>8/9</td>
<td>5/10</td>
</tr>
<tr>
<td>2</td>
<td>2,000</td>
<td>10/10</td>
<td>4/10</td>
<td>8/8</td>
<td>5/8</td>
</tr>
</tbody>
</table>

**Figure 3.3. SP cells are enriched for TPCs in Foxn1<sup>nu/nu</sup> nude mice.**

A. Number of tumors per total number animals orthotopically transplanted with SP and NSP cells at different concentrations.

B. TPC enrichment in SP and NSP cells by ELDA.
Figure 3.4. SP cells are enriched for TPCs in NSG mice. A. Number of tumors per total number animals orthotopically transplanted with SP and NSP cells at different concentrations. B. TPC enrichment in SP and NSP cells by ELDA.

### 3.3.2. SP Cells are a Dynamic Population in Vivo

To investigate the stability of TPCs, I sorted SP cells expressing YFP and NSP cells expressing RFP from KPCC tumors and co-transplanted them orthotopically into Foxn1<sup>nu/nu</sup> mice (n=5) at an approximate ratio of 1 SP : 3 NSP (Figure 3.5A-B). I hypothesized that if TPCs is a dynamic cell state, cells in the NSP state would be able to convert into the SP state, resulting in the presence of RFP expressing SP cells. When the transplanted tumors reached 500mm<sup>3</sup>, the tumors were digested and analyzed by FACS. Overall, both YFP and RFP expressing cells were present in the transplanted tumor. However, there is a wide range of RFP and YFP expressing cells found in mice transplanted with different tumors likely reflecting heterogeneity across individual tumors (Figure 3.5C). On average 27.21% (±14.05% SEM) of cells express RFP and
33.77% (±20.53% SEM) express YFP suggesting both injected SP and NSP cells survived and proliferated in the initial transplanted tumor. Importantly, in the SP compartment, on average 8.02% (±4.01% SEM) were RFP+ and 61.45 % ( ± 21.1% SEM) were YFP+ (Figure 3.5D). In the NSP compartment, 13.41% (±7.29% SEM) were RFP+ and 44.92% (±16.17% SEM) were YFP+. This suggests that NSP cells were able to convert to an SP phenotype de novo during tumor growth. To test whether the NSP derived SP cells (D-SP) have TPC properties, I performed secondary transplants using SP-RFP+ cells and NSP cells expressing either YFP or RFP. Compared to NSP cells, the SP-RFP+ cells are significantly more tumorigenic and did not contain any YFP+ cells (Figure 3.5E-F). Collectively, these data indicate that TPCs in sarcoma may be a dynamic population, which can be derived from non-TPCs as tumor progresses. I term the SP cells derived from NSP cells as Derived SP cells (D-SP).
**Figure 3.5. NSP cells can give rise to SP cells de novo.** A. Experimental schematic to determine the stability of SP cells. B. FACS gating strategy to isolate SP cells expressing YFP and NSP cells expressing RFP from the same tumor. C. Percent of YFP and RFP cells in the co-transplanted tumors. Percent of YFP and RFP cells within the SP and NSP fractions of the co-transplanted tumor. D. Transplantation at limiting dilution of Derived-SP cells (D-SP) and NSP cells in nude mice. *P<0.05 ELDA. E. Flow cytometry analysis and confocal image of the D-SP derived tumor indicating the tumor is derived from D-SP cells. Scale 100μm.

### 3.3.3. Genetic Ablation of SPs

TPCs are thought to be valuable therapeutic targets because the ability of these cells to continuously self-renew and propagate the tumor. To test the effect of ablating TPCs during tumor growth, I generated *LSL-KrasG12D; Tp53f/f; Rosa26-DTR* animals (KP-DTR). Injection of adenovirus expressing Cre recombinase will activate the oncogenic mutations for sarcoma formation, and the expression of diphtheria toxin receptor (DTR) in the tumor cells. I isolated the CD45-SP cells from KP-DTR tumors and mixed them with NSP cells expressing RFP from KPCC animals at a ratio of 1:2.5 (Figure 3.6A-B). The mixed cells were orthotopically transplanted into the gastrocnemius muscle of nude mice (*n*=10). Depletion of SP cells from the KP-DTR can be determined by changes in the number of CD45-SP cells that do not express any fluorescent reporters. Two weeks following transplant, animals with transplanted tumors were randomly divided to receive either diphtheria toxin or its vehicle PBS for a total of 6-8 doses. FACS analysis of the DT treated tumors showed a significant increase of total overall CD45-SP cells compared to PBS treated tumors, with a corresponding decrease in CD45-NSP in the overall number of CD45-NSP cells (Figure 3.6C-D). In addition, DT treatment significantly depleted the CD45-SP cells
that do not express any fluorescent reporters (SP-colorless), indicating effective ablation of SP cells from the KP-DTR animals (Figure 3.6C, E). Within the SP population, the percent of SP cells expressing RFP increased significantly in the DT treated tumors compared to the control tumors (Figure 3.6E). Taken together, these data suggest that DT treatment to the KPCC-DTR mixed tumors can reduce SP cells, but the lost SP cells are compensated by NSP cells.

Next, I performed transplantation assays at limiting dilutions to compare the tumor-initiating capacity of DT treated tumors against the control. Equal numbers of cells derived from KPCC-DTR mixture tumors treated with DT or vehicle were injected into the gastrocnemius muscles of nude mice. The tumor initiating capacity is not significantly different between tumor cells that were treated with DT or vehicle (Figure 3.6F). This is consistent with the compensation of SP cells from NSP cells and suggests that genetically ablating SP cells alone is not sufficient to inhibit tumor self-renewal.
A. Mix and Transplant

NSP-RFP From KPCC, SP from KP-DTR

↓

2.5 weeks

DT Treatment → Harvest

↓

Flow Cytometry

B. NSP-RFP From KPC

SP from KP-DTR

↑

Hoechst Blue

Hoechst Red

200k

100k

0

100k

250k

100k

250k

10^2

10^1

10^0

RFP

CD45

YFP

C. PBS

DT

↑

Hoechst Blue

Hoechst Red

200k

100k

0

100k

250k

100k

250k

10^2

10^1

10^0

CD45

RFP

D. %CD45- Cells

E. % CD45- SP Cells

F. Cell Number vs. Tumors/Second Transplant Mice

<table>
<thead>
<tr>
<th>Cell Number</th>
<th>Tumors/Second Transplant Mice (PBS)</th>
<th>Tumors/Second Transplant Mice (DT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0/4</td>
<td>2/5</td>
</tr>
<tr>
<td>100</td>
<td>3/5</td>
<td>4/5</td>
</tr>
<tr>
<td>500</td>
<td>4/5</td>
<td>3/5</td>
</tr>
<tr>
<td>800</td>
<td>4/4</td>
<td>5/5</td>
</tr>
<tr>
<td>Total</td>
<td>11/18</td>
<td>14/20</td>
</tr>
</tbody>
</table>
**Figure 3.6. NSP cells can compensate for the ablation of SP cells.** A. Experimental schematic for ablation of SP cells. B. FACS gates to isolate SP cells from KP-DTR animals (SP-DTR) and NSP cells expressing RFP from KPCC animals (NSP-RFP). C. Gating strategy to analyze the percent of Derived-SP (D-SP) and non-fluorescent CD45-SP colorless (C-SP) cells in the co-transplanted tumors treated with PBS or DT. D. Total percent of CD45-SP and CD45-NSP cells in co-transplanted tumors treated with PBS or DT. E. Percent of Derived-SP cells and CD45-SP cells in co-transplanted tumors treated with PBS or DT. F. Tumorigenic capacity determined by transplantation at limiting dilution of tumors treated with DT and PBS.

### 3.4 DISCUSSION

Morphological and phenotypic heterogeneity of cancer cells within individual tumors are key features that drive therapeutic failure. Among the diverse cancer cells, TPCs characterized by enhanced self-renewal and tumorigenicity has been identified in most cancer types (Matchett & Lappin, 2014). From a therapeutic perspective, elimination of TPCs is thought to be critical in disabling the growth and self-renewal potential of tumors. Nonetheless, recent evidence in several normal tissues suggests that stem cells may exist as a dynamic population where differentiated cells can acquire stem cell fates. It is unclear whether this compensatory phenomenon is present in sarcomas.

In this study, I used a multicolor reporter model of UPS to trace the stability of SP cells. I first confirmed that SPs are enhanced for tumorigenicity and self-renewal potential compared to their NSP counterparts. By co-transplanting SP and NSP cells expressing different fluorescent reporters into nude mice, I showed that NSP cells could give rise to SP cells *de novo*. To test the impact of eliminating SP cells on tumor growth and self-renewal, I co-transplanted SP cells from
KP-DTR mice with NSP cells from KPCC mice. Ablation of SP cells by DT injection showed compensation of the SP fraction by NSP cells, and the tumorigenic capacity of tumors treated with DT was similar to tumors treated with control vehicle. Overall, my data suggest that TPCs is likely a dynamic or transient state, rather than a stable cell population, and the elimination of TPCs alone may not be able to eradicate tumor self-renewal.

Cell extrinsic cues play vital roles in determining the cancer cell heterogeneity and may contribute to the dynamic or transient state of TPCs. Specifically, TPCs has been shown to reside in specific microenvironments. For example, TPCs in head-and-neck cancers and glioblastoma has been shown to reside in the perivascular niche, and vascular endothelial cells support TPCs by preventing apoptosis (Calabrese et al., 2007; K. E. Ritchie & Nor, 2013). Cancer-associated fibroblasts and hypoxia is known to stimulate stemness and promote TPC survival via activation of self-renewal pathways (Plaks et al., 2015; Vermeulen et al., 2010). Cross-talk among cancer cells can contribute to phenotypic diversity of cancer cells as well. In breast cancer, a minor subpopulation of cells expressing IL-11 can stimulate the growth of other cancer cells through a non-cell autonomous manner (Marusyk et al., 2014). Moreover, breast cancer cells with a mesenchymal phenotype can promote tumorigenicity and self-renewal of nearby tumor cells via paracrine activation of Wnt signaling. Future studies into the interactions between cancer cells or cancer cells, and the tumor microenvironment may reveal the mechanisms that enable non-TPCs to convert to TPCs. Targeting TPCs while preventing compensation from non-TPCs may offer novel therapeutic approaches against cancer progression.
3.5 METHODS

Mouse Models

All animal studies were performed following approved protocols from the Duke University Institutional Animal Care and Use Committee. The R26R-Confetti (Stock number: 013731) and R26R-LSL-DTR (Stock number: 007900) mice were obtained from Jackson Laboratories (ME). The KRAS\textsuperscript{1LSL-G12D} mice were provided by Dr. Tyler Jacks (MIT), and the p53\textsuperscript{f/f} mice were provided by Anton Berns (NKI). All mice were on a mixed genetic background. To induce tumors, adenovirus expressing Cre recombinase mixed with 50\textmu l of 2M CaCl\textsubscript{2} in DMEM were directly injected into the hind gastrocnemius muscle of 7-10 weeks old mice. Growth of tumors was confirmed by physical examination at the injection site and by histology.

Tumor Digestion

Primary sarcoma tumors from KPCC and KP-DTR mice were mechanically and enzymatically dissociated into single-cell suspensions as previously described (Wei et al., 2015). Briefly, the tumors were mechanically dissociated into small pieces and enzymatically digested with a mixture of 10mg/mL of collagenase IV (Worthington), 2.4U/ml of Dispase (Becton Dickinson), and 0.05% trypsin. Red blood cells were excluded with ACK lysis buffer (ThermoFisher Scientific). Single cell suspensions are frozen down using Synth-a-Freeze media following manufacture’s protocol (ThermoFisher).

Histology

For frozen sections, the tumor samples were fixed in 4% paraformaldehyde at 4°C for 1 hour, followed by incubation in 15% and 30% w/v sucrose for 24 hours. The tissues were then embedded
in Optimum Cutting Temperature medium (OCT, Tissue-Tek) on dry ice. 10-12\(\mu\)m sections were cut using a Cryostat (Leica CM1950). The tissue sections were mounted with Prolong Diamond Antifade Mountant with or without DAPI (ThermoFisher). H&E images were taken with Zeiss Axio Imager Z2 (Zeiss, Germany) at either 10X or 20X, and the entire sections were scanned.

**Confocal Microscopy**
Images of frozen KPCC tumors were taken using Leica SP8 confocal microscope with a 20X or 40X objective. The argon laser 488nm line was used to excite nuclear GFP, 514nm line for YFP, 561nm red diode laser for RFP, 458nm laser line for CFP, and UV laser line for DAPI. The collection wavelengths for GFP, YFP, RFP, CFP, and DAPI were: 493-510nm, 537-565nm, 612-650nm, 468-498nm, 417-497nm, respectively. Airy unit 1 was used to take images. The images were taken using Leica Application Suite X.

**Flow Cytometry**
Flow cytometry assisted cell sorting (FACS), or analysis were performed on BD FACSDiva Cell Sorter at the Duke Flow Cytometry Core (Duke University, NC). For isolation of SP and NSP cells, single-cell suspensions were treated with 2.5 mg/mL of Hoechst 33342 dye (Sigma) alone, or in combination with 50 mmol/L of verapamil (Sigma) as a negative control, for 90 minutes at 37\(^\circ\)C. SP cells were identified using dual wavelength analysis (blue, 424–444nm; red, 675 nm) after excitation with 360 nm UV laser. SP cells from the hematopoietic lineage were eliminated by staining for CD45 (1:500 dilution, BD Biosciences). RFP and YFP cells from the KPCC tumors are identified using the blue 488nm laser. Dead cells are eliminated with propidium iodide.
Tumor Transplantation and Diphtheria Toxin (DT) Treatment

For tumor transplantation experiments, 6 to 7-week-old Foxn1\textsuperscript{nu/nu} were purchased from Duke University Division of Laboratory Animals and Resources. Tumor cells are diluted into 1XPBS at different numbers, as indicated in the study. To inject the cells, the nude mice were anesthetized under 2% isoflurane. The cell solutions were injected into the left gastrocnemius muscle of the animals. The animals were examined every two days. Tumor formation was confirmed by physical examination and histology. All animals were euthanized when the tumors reached 1.5 cm in diameter. Enrichment of TPC population is determined by Extreme Limiting Dilution Method based on the number of tumors out of the total number of mice transplanted at each cell dilution (Hu & Smyth, 2009). To generate the tumors mixed with KPC and KP-DTR cells, SP cells and NSP cells were mixed at 1:2.5 and injected into gastrocnemius muscle of nude mice. DT was obtained from Sigma (MO) and resuspended in 1XPBS. The animals were administered intraperitoneally at 250 μg/kg diluted in 100μl. Injections were performed every other day for a total of 6-8 doses. The animals were monitored for signs of weight loss, posture, and mobility.

Statistics

Graphs and statistical calculation are generated using GraphPad Prism (version 6). Statistical significance is determined by two-tailed student’s t-test unless described otherwise.
CHAPTER 4

TRACING TUMOR EVOLUTION IN SARCOMA REVEALS CLONAL ORIGIN OF ADVANCED METASTASIS

4.1 Authors Contributions

I performed most of the experiments and data analyses with the help of lab members and collaborators. J.G. Huang (Duke University, NC, USA) assisted with PCR for CRISPR-barcoding, D. Corcoran (Duke University, NC, USA) assisted with bioinformatics and X. Ding (Duke University NC, USA) assisted with developing the statistical theory. The experiments are performed in the laboratory of Dr. Benjamin Alman in collaboration with the laboratory of Dr. David Kirsch (Duke University, NC, USA).
4.2 Introduction

Most cancers are composed of heterogeneous populations of cells. During tumor progression, genomic instability of cancer cells and differences in the tumor microenvironment can confer tumor clones with different genetic and epigenetic alterations (Marusyk, Almendro, & Polyak, 2012; McGranahan & Swanton, 2017; Swanton, 2012). These alterations may endow phenotypic differences in tumor clones which are subjected to competition and selection. Cancer progression can be viewed as an evolutionary process where tumor clones compete and are selected upon over the course of the disease (Greaves & Maley, 2012). While the phenomenon of clonal heterogeneity at the morphological, histological and genetic level has been well characterized in certain cancer types, the functional impact of different tumor clones during tumor evolution is not well understood, particularly in mesenchymal cancers (de Bruin et al., 2014; Gerlinger et al., 2012; Gulati et al., 2014; Makohon-Moore et al., 2017). In particular, how different tumor clones contribute to primary tumor growth, recurrence, and metastasis is not fully defined. During primary tumor progression, it remains unclear whether growth is driven by the continuous expansion of a dominant clone or the maintenance of various clones. Furthermore, the clonal composition of distant metastasis is poorly understood. Specifically, it is not known whether all tumor clones are capable of forming distant metastases and whether the spatially distinct metastatic lesions found in the receptive organ are clonally related. Insights into the clonal composition and evolution of tumors throughout the major stages of disease progression are essential to understanding the biological consequences of intratumoral heterogeneity and developing new therapy to impede cancer progression.

Our understanding of tumor clonal dynamics is largely derived from next generation sequencing of tumor genomes (Gundem et al., 2015; Hao et al., 2016; Y. Wang et al., 2014; Yates
et al., 2015). Based on shared somatic mutations, the phylogenetic relationships of tumor clones and their life history can be retrospectively constructed (Schwartz & Schaffer, 2017). These studies captured snapshots of tumor heterogeneity and clearly demonstrated that most tumors are composed of multiple clones. While genomic analyses of clonal dynamics have revealed important insights into tumor evolution, such as frequency of mutations and the order in which specific somatic mutations may arise, their ability to prospectively track tumor evolution in an unperturbed manner is limited (Landau et al., 2013; Nik-Zainal et al., 2012; Welch, 2014). This is especially the case for many solid cancers, where longitudinal sampling is difficult, and patient-matched primary tumor and metastases samples are scarce. Moreover, it is difficult to ascribe the genetic variations used to identify tumor clones to specific cellular phenotypes. To functionally investigate the impact of various tumor clones during different stages of tumor progression requires in vivo models of lineage tracing.

In this study, I traced tumor evolution in a mouse model of undifferentiated pleomorphic sarcoma (UPS), one of the most common and lethal soft tissue sarcomas in human adults (Burningham et al., 2012; Kirsch et al., 2007). Using the R26R-Confetti lineage tracing system and endogenous barcoding with CRISPR-Cas9, I investigated clonal dynamics during tumor initiation, local recurrence after therapy, and distant metastasis.

4.3 Results

4.3.1 Modeling Clonal Evolution in Sarcoma

To track tumor evolution in sarcoma, I utilized two complementary lineage tracing mouse models. First, I crossed the homozygous R26R-Confetti system with a spatially and temporally restricted model of UPS, carrying a Cre-inducible oncogenic Kras<sup>LSL-G12D</sup> allele and homozygous
p53<sup>−/−</sup> alleles (Kras<sup>LSL-G12D</sup>; p53<sup>−/−</sup>; Confetti/Confetti; KPCC mice in Figure 4.1A) (Snippert et al., 2010). Ras/MAPK signaling is commonly activated in human soft tissue sarcomas, including UPS, and inactivating p53 mutations or silencing of the p53 pathway is one of the most common mutations in UPS (Cancer Genome Atlas Research Network. Electronic address & Cancer Genome Atlas Research, 2017; Perot et al., 2010; Serrano et al., 2016; Yoo, Robinson, & Lee, 1999). Tumors formed by activation of oncogenic Kras<sup>G12D</sup> and deletion of p53 were similar to human UPS in histology and molecular profile (Blum et al., 2013; Kirsch et al., 2007; Mito et al., 2009; Sato et al., 2016). In the KPCC animals, following injection of Cre-expressing adenovirus into the gastrocnemius muscle, primary sarcomas expressing different fluorescent reporters were formed at the injection site between 9-12 weeks with 100% penetrance similar to previous reports (Figure 4.1B) (Kirsch et al., 2007; Mito et al., 2009). I observed up to 8 distinct fluorescent protein combinations, labeling on average ~60% of cells in tumors (Figure 4.1C). Expression of the R26R-Confetti allele did not affect tumor histology (Figures 4.1B). Consistent with other publications using the R26R-Confetti allele (Baggiolini et al., 2015), fewer cells recombined to express GFP compared to the other fluorescent reporters (Figures 4.1C). Long term culture experiments showed that the expression of fluorescent proteins was permanent and stable (Figure 4.1D). Distinct tumor clones can be identified by the expression of different fluorescent reporters, and these clones share founder mutations and grew in close proximity within the tumor (Figure 4.1E). Intermixing of tumor clones is commonly observed in KPCC tumors (Figure 4.1F).
**Figure 4.1.** KPCC and K-sgP53-Cas9 mouse models of sarcoma for tracing clonal evolution.

**Primary tumor progression is associated with a loss of clonal heterogeneity.**  
A. Schematic of KPCC and K-sgP53-Cas9 mouse models.  
B. Photo of sarcoma in the left gastrocnemius muscle in a KPCC mouse. Tumor histology of different genotypes. Scale 100μm.  
C. Percentage of different fluorescent reporter combinations in KPCC mice. Data represented as mean ± SD of total DAPI+ cells in tumors.  
D. *In vitro* passage of KPCC tumor cells expressing RFP and CFP.  
E. PCR analysis demonstrating recombination of LSL-KrasG12D and p53f/f alleles in different tumor clones from
a KPCC tumor and negative control. Non-recombined LSL-KrasG12D band is 612bp and
recombined KrasG12D band is 650bp. To check recombination of p53, the recombined p53+/f
alleles has a band at 600bp and non-recombined p53+/f alleles do not have any bands. The bottom
panel shows PCR for the same samples to check for unrecombined p53+/f band in the negative
control, and absence of bands in p53 recombinant tumor clones. F. Confocal microscopy image of
a KPCC tumor, Scale 1mm. Inset Scale 100μm.

Because the R26R-Confetti alleles randomly labeled tumor cells with 6-8 colors, and a
given tumor may theoretically have more than 8 clones, it is possible that different clones may be
labeled by the same color. To complement the KPCC model, I utilized another KP sarcoma model
induced by CRISPR-Cas9 mediated genome editing (J. Huang et al., 2017). The CRISPR-Cas9
system can produce a vast array of indels, and thus is not limited to the maximum number of colors
observed in the R26R-Confetti alleles (McKenna et al., 2016; Shin et al., 2017). In this model, the
Cre-inducible KrasG12D allele is retained, and the Rosa26 locus expressed Cas9 endonuclease.
Sarcomas were initiated by injecting adenovirus expressing Cre and sgRNA targeting p53 (K-
sgP53-Cas9 in Figure 4.1A). The K-sgP53-Cas9 tumors were histologically similar to KP and
KPCC tumors (Figure 4.1B). The genetic edits generated by CRISPR-Cas9 at the targeted region
of p53 can serve as barcodes marking independent tumor clones.

4.3.2 Primary Tumor Progression is Associated with A Loss of Clonal Heterogeneity

Using the KPCC and K-sgP53-Cas9 models, I first examined how clonality changed during
primary tumor progression. Tumors were harvested between 250mm³-300mm³ to determine the
clonality at a relatively early stage of sarcomagenesis and were then compared to late stage tumors
that were approximately 2000mm$^3$ in volume (Figure 4.2A). The cytology observed in early stage tumors consisted of pleomorphic spindle cells with nuclear atypia, which had a similar cytological appearance to late stage tumors (Figure 4.2A). I observed significantly more fluorescent colors in early stage tumors compared to late stage tumors (Figures 4.2B-D). Furthermore, I scored each image from tumor sections as monochromatic, consisting of mostly 1 fluorescent color, as dichromatic, consisting of 2 fluorescent colors or as polychromatic, consisting of 3 or more fluorescent colors. The early stage tumors had significantly higher polychromatic scores compared to late stage tumors (Figures 4.3A-B).

For the K-sgP53-Cas9 model, I harvested tumors at similar early and late stages and performed deep sequencing of the $p53$ alleles using genomic DNA extracted from tumor masses. Early stage tumors have significantly more barcodes compared to late stage tumors (Figures 4.2E and Figure 4.4). This is consistent with our findings in the KPCC model, indicating that primary sarcoma progression is associated with the sweeping growth of a dominant clone.

The KPCC and K-sgP53-Cas9 models require separate animals to study clonal evolution. To trace tumor clones in the same animal over time, I developed an intravital microscopy model in the mouse pinnae. Differently labeled tumor cells were injected into the pinnae of nude mice. These transplanted tumors maintained the histological features of primary sarcoma (Figure 4.5A). Consistent with the autochthonous models, the growth of transplanted tumors is associated with a reduction in clonal heterogeneity (Figure 4.5B).
Figure 4.2. Primary tumor progression is associated with a loss of clonal heterogeneity. A.

Tissue and tumor histology from H&E staining at different time points after injection of Ad-Cre
virus into the gastrocnemius muscle. Scale 100µm. B. H&E staining and confocal microscopy images of a representative early stage tumor. GFP (green), YFP (yellow), RFP (red), CFP (cyan), DAPI (blue). Scale 1mm, Inset scale 50µm. C. H&E staining and confocal microscopy images of a representative late stage tumor. Scale 1mm, Inset scale 25µm. D. Graph comparing the number of fluorescent reporters observed in early and late tumors of KPCC model. Data are represented as mean ± SD (*, P<0.05). E. Graph comparing the number of barcodes from the K-sgP53-Cas9 mice between early and late stage tumors. The barcodes are generated by deep sequencing of the p53 allele. Data are represented as mean ± SD (*, P<0.05).
**Figure 4.3** Loss of clonal heterogeneity during primary tumor progression.  
A. Percent of monochromatic, dichromatic, and polychromatic regions in early and late stage tumors. (***, P<0.005).  
B. Representative images of different categories of tumor regions. Scale 25µm.

**Figure 4.4.** Deep sequencing of DNA barcodes in primary tumors of K-sgP53-Cas9 mice. 
Representative DNA edits at the p53 alleles induced by CRISPR-Cas9 in early and late stage tumors.
Figure 4.5. Intravital microscopy of tumor progression. A. Photo of tumor engraftment in the pinna of a nude mouse and H&E staining of the engrafted tumor. Scale 50μm. B. Representative intravital microscopy images of the engrafted tumor over a period of 14 days showing a decrease in clonal heterogeneity over time. Scale 50μm.
### Figure 4.6. Local recurrence in sarcoma is driven by multiple clones

A. Schematic of experiments for local recurrence (LR). Tumor-bearing KPCC mice were either amputated or focally irradiated. B. Kaplan-Meier curve representing the percentage of mice and time to local recurrence following surgery. C. Confocal microscopy images of locally recurrent tumor after surgery. GFP (green), YFP (yellow), RFP (red), CFP (cyan). Scale 1mm, Inset scale 100µm. D. Confocal microscopy images of recurrent tumor after irradiation. GFP (green), YFP (yellow), RFP (red), CFP (cyan), DAPI (blue). Scale 1mm, Inset scale 100µm. E. Two polyclonal tumors were transplanted into multiple nude mice and stereotactic irradiation delayed tumor growth. Graph show effects of irradiation on tumor growth over time. Data represented as mean ± SD of relative tumor volume to the starting volume. F. Confocal microscopy images of control and recurrent transplanted tumors after radiation therapy. Scale 50µm.

#### 4.3.3 Multiple Tumor Clones Give Rise to Local Recurrence

Local recurrence following therapy causes substantial morbidity and is an important prognostic risk factor for survival in many cancers including sarcoma (Daigeler et al., 2014). To investigate whether local recurrence is derived from a single or multiple clones (Figure 4.6A), I treated polyclonal early stage KPCC tumors by either surgically amputating the tumor-bearing leg (Sachdeva et al., 2014) or by delivering focal image-guided radiation (Moding et al., 2015). Amputation of the tumor-bearing legs was performed at the femoral head, and the entire leg was removed. In the amputated mice, 5/26 (19%) developed locally recurrent tumors near the pelvic bone at the surgical margins with a median of 30-days (Figure 4.6B). I analyzed the recurrent disease at approximately 300mm³-800mm³ and found that the recurrent tumors consisted of multiple clones (Figure 4.6C). Recurrence after irradiation was defined by the continued growth
of the tumor post irradiation and the recurrent tumors were harvested between 800mm$^3$-1500mm$^3$. Similar to local recurrence following surgery, the recurrent tumors (n=6) following radiation therapy consisted of multiple clones (Figure 4.6D). To assess the effect of radiation on recurrence for the same tumor, I transplanted 2 polyclonal tumors into nude mice and gave half of the mice a single dose of 25 Gy radiation therapy. By imaging the control tumors and irradiated tumors, I found that similar to the autochthonous model, recurrence following irradiation in the transplant model was derived from multiple tumor clones (Figure 4.6E-F). Together, these data demonstrate that multiple clones give rise to local recurrence after surgery or radiation therapy.

4.3.4 Distant Metastases are Clonal in Origin

Metastasis accounts for the vast majority of cancer related mortality. To investigate the clonality of metastases, I amputated the tumor-bearing limb and monitored the animals for up to 6 months (Figure 4.7A). In the KPCC mice, approximately 40% of mice developed metastasis after amputation with a median survival of 54 days post-surgery (Figures 4.7B-C). This is similar to the rate of metastasis in human patients (Digesu, Wiesel, Vaportcian, & Colson, 2016). The primary tumors were resected at early stage (250mm$^3$ - 350mm$^3$) when they were polyclonal (Figure 4.7D). The majority of the metastasis I observed was to the lungs (n=18/19), the most common site of metastasis in human patients; and 1/19 animal developed metastasis to the kidney (Figure 4.8A). Upon examination of the lung metastases, every lung lesion within a mouse with multiple lesions, ranging from 3-52 lesions per mouse, expressed the same fluorescent reporter (Figure 4.7E). I examined the metastases from KPCC animals by confocal microscopy and found that the percentages of different fluorescent reporters expressed in the metastases across animals were similar to the percentages of different fluorescent reporters observed in the primary tumors.
Figure 4.7E). Two animals also developed liver metastases in addition to their lung metastases. The tumor cells in the liver expressed the same reporter as tumor cells in the lungs (Figure 4.8B).

Since labeling by the R26R-Confetti allele is random, it is possible for distinct clones to be labeled by the same fluorescent reporter. I statistically determined the probability of this occurring over multiple animals with metastases. I reasoned that the likelihood of two or more fluorescently indistinguishable tumor clones forming metastases would decrease with an increasing number of animals that had metastases comprised of cells expressing the same reporter (Supplement Statistical Theory, Section I). In other words, if metastasis is polyclonal in origin, the probability of not observing multi-colored lesions would decrease with an increasing number of mice. For our sample size \( n=19 \), the probability that two or more clones expressing the same reporter contributing to lung metastasis is less than \( 9.71 \times 10^{-15} \). Therefore, advanced metastasis from sarcoma is most likely derived from a single clone in the primary tumor.

To further establish the clonal contribution of metastases experimentally without ambiguity, I used the K-sgP53-Cas9 sarcoma model. I amputated the extremity of these animals when the tumors were at an early stage (350mm\(^3\)). The amputated mice developed metastases at a similar latency and frequency as the KPCC model \( n=7/18 \), Figure 4.7C). Deep sequencing of the \( p53 \) alleles in matched bulk tumors and multiple metastases showed that while the primary sarcomas were polyclonal containing an average of 4-5 barcodes, the metastases from each mouse showed only 1 \( n=6 \) mice) or 2 barcodes \( n=1 \) mice) that matched the barcodes from the original primary tumor (Figure 4.7G, Figure 4.9). The WT \( p53 \) reads observed in the primary tumors and metastases were likely derived from normal stromal or immune cells in the primary tumors and at the metastatic site. In the metastases with 2 barcodes, the percentage of reads for each barcode was similar, suggesting that different edits were likely introduced to each \( p53 \) allele of the same
clone (Figure 4.9). The combined strength of lineage tracing data from two independent and complementary mouse models indicate that advanced lung metastases originated from a single metastatic clone (MC) in primary sarcoma. Moreover, in these models, distinct metastatic lesions in the same animal are clonally related.
Figure 4.7. Lung metastases is driven by a single clone.  A. Schematic of amputation experiments and development of lung metastases. B. Gross examination of harvested lungs with arrows indicating the presence of metastases. C. Kaplan-Meier curve representing the percent metastasis-free survival of amputated KPCC and K-sgP-53-Cas9 mice. D. H&E staining and confocal microscopy images of a representative tumor after amputation. GFP (green), YFP (yellow), RFP (red), CFP (cyan), DAPI (blue). Scale 500µm, Inset scale 50µm. E. Stacked bar graph of the number of KPCC mice that developed metastases with corresponding color; and stacked bar graph of the percentage of differently labeled cells in the primary tumor. F. H&E staining and confocal microscopy images of representative lung metastases from a KPCC animal. Scale 1mm, Inset scale 25µm. G. Graph showing the number of p53 barcodes in K-sgP53-Cas9 primary tumor and metastases.
Figure 4.8 Clonality of metastasis to other organs in KPCC animals. A. H&E and confocal image of kidney metastasis. Scale 1mm. Inset Scale 100μm. B. Gross picture and confocal image of metastasis to the liver. Scale 1mm. Inset Scale 100μm.
Figure 4.9. Deep sequencing of DNA barcodes in primary tumors and metastases of K-sgP53-Cas9 mice. Barcode sequences of amputated primary tumors matched metastases, and additional metastases are shown. The percent representation of each barcode is provided. Arrow indicates the matched barcode in the primary tumor and metastases.

4.3.5 Clonal Metastasis is a Deterministic Process

The emergence of a dominant clone in tissues can occur through either stochastic neutral drift or deterministic clonal selection (Lopez-Garcia, Klein, Simons, & Winton, 2010; Morrissy et al., 2016; Snippert et al., 2010). To determine whether the cellular mechanism of clonal origin of metastases is due to neutral drift where a single clone overtakes the lungs by chance, or selection of a tumor clone that is most capable of metastatic colonization, I orthotopically transplanted polyclonal primary KPCC tumor cells into the left gastrocnemius muscle of multiple nude mice (Figure 4.10A). I hypothesized that if metastases were due to neutral drift, then the metastases in each nude mouse engrafted with the same KPCC tumor would be formed by different clones, which can be observed by the expression of different fluorescent reporters. On the other hand, if the clonality of metastases was due to biological selection of a clone with enhanced metastatic capacity, then the same clone would form metastases in all the nude mice transplanted with the same KPCC tumor.

I transplanted 5 primary KPCC tumors into the gastrocnemius muscle of 10-16 nude mice per KPCC tumor (Figure 4.10A). After the orthotopic transplant, I amputated the tumor-bearing legs from the nude mice at early stage (250 mm$^3$-350mm$^3$) and followed the mice for the development of metastases (Figure 4.10B). I confirmed the transplanted tumors were polyclonal, indicating that the differently labeled clones were tumorigenic, and the transplanted tumors
recapitulated the clonal heterogeneity of the primary tumor (Figure 4.10C). In mice that were transplanted with the same KPCC tumor, metastases in each nude mouse expressed the same fluorescent reporter, indicating they were derived from the same clone (Figure 4.10D). This result was consistent across 5 groups of animals transplanted with different primary tumors (Figure 4.10B, Figure 4.11). Statistically, I considered the possibility of a single clone forming metastases in all the nude mice by random chance. I used an intracluster correlation coefficient approach to determine the probability of chance giving rise to our results (Kerry & Bland, 1998). Based on our sample size, the probability that all of the metastases in different nude mice randomly expressed the same fluorescent reporter is less than 3% (Supplement Statistical Theory, Section II). Therefore, the consistent presence of the same fluorescent reporter found in the metastases of nude mice indicates that metastasis is likely driven by a deterministic process via clonal selection.

Since late stage primary tumor and advanced metastasis were dominated by a single clone, I sought to determine whether MCs were derived from the dominant clones in each primary tumor. Because amputation of late stage tumors resulted in overwhelming mortality and rapid onset of local recurrence, I orthotopically transplanted 3 independent polyclonal KPCC tumors in which the identity of the MC cells was known, into nude mice and allowed these tumors to grow to late stage. I analyzed the transplanted tumors by confocal microscopy and FACS to determine the abundance of each tumor clone. The results showed that in KPCC-1, the metastatic clone was present as a minor clone when the transplanted tumor reached late stage (Figure 4.12A). In the other 2 transplanted KPCC tumors, the MCs were dominant in late stage primary tumors (Figure 4.12B, C). However, in KPCC-4, the dominance of MC in the late stage transplanted tumor was not consistent in every nude mouse, wherein one of the transplanted tumors MC was not the dominant population (Figure 4.12C). Taken together, these results suggest that MCs can be either
the dominant or minor clone in late stage primary tumors. Thus, growth advantage in the primary location is not always associated with enhanced metastatic ability.

Figure 4.10. Monoclonal origin of metastases is driven by selection. A. Schematic of orthotopic transplantation and amputation experiment. Independent primary KPCC tumors were injected into the left gastrocnemius muscle of multiple nude mice, and the tumors were amputated. B. Kaplan-Meier curve representing the percentage of metastasis-free survival for each transplanted KPCC tumor. C. H&E staining and confocal microscopy images of representative transplanted tumors at the time of amputation. GFP (green), YFP (yellow), RFP (red), CFP (cyan). Scale 1mm, Inset scale 50µm. D. H&E and confocal microscopy images of lungs from nude mice transplanted with tumor KPCC1. GFP (green), YFP (yellow), RFP (red), CFP (cyan). Scale 1mm, Inset scale 100µm.
Figure 4.1. Metastasis in KPCC tumors is driven by selection. Lung lobes with advanced metastasis in nude mice transplanted with different KPCC tumors. The metastases are formed by the same clone in each nude mouse transplanted with the same tumor. Scale 1mm.
Figure 4.1. Metastatic Clones (MCs) can derive from minor or dominant clones in the primary tumor. A-C. FACS analysis and confocal image of orthotopically transplanted KPCC tumors (n≥5 for each KPCC tumor) harvested at late stage (~1800mm³). The percent of each tumor clone from individual nude mouse transplant are plotted as bar graphs. Data are represented as mean ± SD. In panel C, red points indicate data from a nude mouse transplanted with KPCC4 tumor, where MC less than the RFP population, and is a minor clone in this tumor-bearing nude mouse.

4.3.6 Clonality of Early Metastatic Process

Next, I sought to determine the clonality of early metastatic seeding. Advanced lesions in distant organs are the end result of a multi-step cascade that involves intravasation into the bloodstream, survival in circulation, extravasation, and engraftment, and ultimately, metastatic colonization (Lambert et al., 2017). Selective pressure can act on each step to influence clonal dynamics (Naxerova & Jain, 2015). I examined the intravasation step by determining the clonal composition of circulating tumor cells (CTCs). Blood from KPCC mice (n=5) with primary sarcomas (approximately 300mm³) was collected by cardiac puncture and stained with hematopoietic lineage marker CD45 and analyzed by flow cytometry. CTCs were identified as cells in the blood expressing R26R-Confetti reporters. I found CTCs in 4/5 mice. On average 1.34% (±0.61%SEM) of CD45- cells expressed fluorescent reporters, and in all 4 animals, multiple reporters were present (Figure 4.13A-D). Nonetheless, in 2/4 animals, there was dominance of a single color in the CTCs (Figure 4.13B), whereas in the other 2/4 animals, multiple fluorescent cells were represented (Figure 4.13C).

I looked further in the metastatic cascade by examining the lungs of KPCC mice before the
onset of weight loss or labored breathing. In contrast to earlier experiments where the lungs were harvested when animals displayed clear symptoms of metastasis (ranging 54-164 days post amputation), lungs were harvested from mice 35-45 days post amputation. I examined the lungs histologically for the presence of early metastatic lesions. The early metastatic lesions were defined as clusters of less than 500 cells and were found in 5 of 15 amputated mice. In 3/5 mice, the early lesions were monoclonal (Figure 4.14A), but in 2 mice the lungs contained polyclonal cell clusters, with 3 colors in each mouse (Figure 4.14B). Furthermore, in the polyclonal animals, the relatively larger early metastatic lesions, which are defined as a cluster of 30-500 cells, were mostly monoclonal (Figure 4.14B). In an additional 2 animals, I observed single KPCC cells in the lungs that expressed the same reporter (Figure 4.13E). Taken together, these data suggest that early metastatic seeding into the lung can be polyclonal or monoclonal.

To determine the ability of each individual clone to colonize the lungs, I FACS sorted different clones from 4 primary KPCC tumors. The individual clones were injected into separate groups of 3-5 nude mice via the tail vein. I followed the animals for 3-6 weeks and compared the metastatic ability of the different tumor clones. The MCs that formed metastases in the amputated primary KPCC mouse formed metastases in 21/21 nude mice (Figure 4.14C-D). Most of the other primary tumor clones did not form metastases. Overall, 4/39 nude mice developed advanced metastases after tail vein injection of the other clones, which I termed Non-MCs. These results suggest that in the polyclonal tumor, most tumor clones have limited capacity to form overt metastases, consistent with the notion that advanced metastasis is largely driven by a single MC.
Figure 4.13. Analysis of circulating tumor cells and early monoclonal metastatic seeding. A. Flow cytometry analysis of nucleated blood cells from a non-tumor bearing KPCC mouse. PI- and CD45- cells are analyzed for cells expressing fluorescent reporters. B. Representative KPCC mouse with polyclonal CTCs. C. Representative KPCC mouse with a dominant RFP+ CTC population. D. Bar graph summarizing the percentage of CTCs in KPCC animals. Data represented as mean ± SEM. E. Confocal image of a lung lobe from a KPCC mouse following
amputation, prior to the onset of metastatic symptoms. Scale 1mm. Inset images are representative single tumor cells in the lung. Inset scale 50µm.

Figure 4.14. Clonality of early metastatic process. A. H&E and confocal images of monoclonal early metastatic lesions from a KPCC mouse. Scale 1mm, Inset Scale 100µm. B. H&E and confocal images of polyclonal early metastatic lesions from a KPCC mouse. Scale 1mm. Inset 1 Scale 50µm. Inset 2 Scale 100µm. C. Table indicates the number of tail-vein injected mice that developed metastases for each sorted population and their totals. (* P<0.05, Chi-Square Test comparing the number of mice with and without metastases between different clones). D. H&E staining images of representative lung metastases from each sorted population. Scale 1mm.
Figure 4.15. Transcriptome analysis of tumor clones from primary KPCC sarcomas.  A. Hierarchical clustering heatmap of MCs (n=6) vs Non-MCs (n=8).  B. GSEA analysis of molecular pathways enriched in MCs.  C. CINSARC signature in MCs vs Non-MCs.  D. Quantitative RT-PCR validation 6 candidate down-regulated genes in MCs in RNA-seq.  Data represented as mean ± SEM.  * P<0.05, **P<0.01, One-way ANOVA.

4.3.7 Transcriptome of Metastatic Clones and Non-Metastatic Clones

To gain insight into the molecular mechanism that distinguishes MCs, which are capable of forming advanced metastatic growth, from Non-MCs, I performed RNA-sequencing on MCs and Non-MCs from the same primary tumors. When the primary sarcomas were amputated in the autochthonous model, the mice were followed for the development of lung metastases. By
matching the color of the fluorescent reporter in the lung metastases to the clones in the primary tumor, I was able to define primary tumor clones as MCs or Non-MCs. I performed RNA-sequencing on MCs, Non-MCs and matched lung metastases (Lung-Mets). Clustering of samples by gene expression patterns distinguished MCs from Non-MCs (Figure 4.15A). The expression of 853 genes was differentially expressed in the MCs relative to Non-MCs. Gene set enrichment analysis (GSEA) of MCs versus Non-MCs showed positive enrichment of processes involved in stemness, cancer aggressiveness, cell cycle, and DNA repair (Subramanian et al., 2005). Cellular processes involved in extracellular matrix interaction, collagen synthesis, and integrin were negatively enriched (Figure 4.15B). Several gene signatures associated with poor cancer outcome were positively enriched in MCs. CINSARC, a well-established gene set prognostic of poor survival in human sarcomas (Chibon et al., 2010) was enriched in MCs (Figure 4.15C). Because the majority of (74%) of the differentially regulated genes were suppressed in MCs compared to Non-MCs, I sought to investigate potential candidate suppressors of metastasis. Since the top negatively enriched processes in our GSEA involved cell-matrix interaction, integrin, and cell adhesion, I focused on genes that are associated with these processes. I took a candidate approach and selected six top differentially suppressed genes in MCs relative to Non-MCs that were not previously known to play a role in sarcoma metastasis, namely: Chad, Podn, Fibin, Rasd1, Reck, and Aldh1a2. The down-regulation of these candidate genes in MCs was confirmed by quantitative RT-PCR (Figure 4.15D).

Comparison of overall gene expression patterns between Lung-Mets, MCs, and Non-MCs showed that the Lung-Mets were distinct from clones in the primary tumor (Figure 4.16A, C). The gene expression differences between lung metastases and Non-MCs were greater compared to the differences between MCs and Non-MCs, with 5,162 genes differentially expressed. GSEA
analysis between Lung-Mets and Non-MCs showed pathways specifically enriched in the Lung-Mets (Figure 4.16B). Notably, processes involved in hypoxia, cell metabolism, and Wnt signaling were positively enriched in metastases, whereas, TNF-α, EGF and immune response were negatively enriched. In addition, I compared the gene expression profile of Lung-Mets versus MCs, and the expression of 3,519 genes was significantly different (Figure 4.16C). Many of these genes are associated with mitochondrial metabolism, DNA damage response, and protein degradation (Figure 4.16D). These differences likely reflect additional molecular changes in the MCs in response to the lung microenvironment as they grew and colonized the tissue over time. Although the CINSARC gene signature was enriched in Lung-Mets compared to Non-MCs, the enrichment was not statistically significant between Lung-Mets and MCs. This suggests that the CINSARC enrichment is specific for differences between metastatic and non-metastatic tumor cells (Figure 4.16D).
Figure 4.16. Transcriptome analysis of Non-MCs, MCs, and Lung-Mets. A. Clustering heatmap of Lung-Mets (n=5) versus Non-MCs (n=8). The CINSARC signature is enriched. B. Summary of GSEA analysis of positively and negatively enriched pathways in Lung-Mets vs Non-MCs. C. Clustering heatmap of Lung-Mets (n=5) versus MCs (n=6). The CINSARC signature is not enriched. D. Summary of GSEA analysis of pathways positively and negatively enriched pathways in Lung-Met vs MCs.
4.3.8 Overexpression of Suppressed MC Genes Inhibits Metastasis

To investigate the potential for the six candidate metastasis suppressor genes (Figure 4.15D) to inhibit metastasis in vivo, I over-expressed each gene in metastatic KPCC cells (Figure 4.17A, 4.18A). Cells over-expressing each candidate gene or empty vector control cells were injected into nude mice (n=3-5 for each construct) via the tail-vein. Metastatic ability was then assessed by averaging the relative total area of the metastatic lesions across the lungs, and by enumerating the number of metastatic lesions. Three of the six genes (Rasd1, Reck, and Aldh1a2) showed 70-75% reduction in metastatic burden (Figure 4.17A, Figure 4.18B).

From the 3 candidate suppressors of metastasis, I focused on Aldh1a2, one of the least characterized members of the aldehyde dehydrogenase family that functions in retinoic acid biosynthesis. MCs were enriched for the down-regulated gene signature associated with knockout of retinoic acid receptors, the downstream transcription factors of retinoic acid signaling (Al Tanoury et al., 2014) (Figure 4.18C). In addition, loss of ALDH1A2 copy number in human soft tissue sarcoma was correlated with decreased survival (Figure 4.18D). I further validated the overexpression of Aldh1a2 in transduced cells by Western blot (Figure 4.18E) and performed tail vein injection in an independent cohort of nude mice. The results showed a consistent decrease in lung metastases (Figures 4.17B-C, 4.18F). Because MCs have increased competitive capacity to establish colonies in the lungs compared to Non-MCs, I hypothesized that over-expression of Aldh1a2 may reduce their competitiveness. To test this, I performed a competition assay by co-injecting GFP+ Aldh1a2-OE cells with RFP+ control cells at a 1:1 ratio through the tail vein of nude mice. In these animals, the resulting metastases were robustly dominated by control cells, where less than 10% of the metastases included Aldh1a2-OE cells (Figures 4.17D-E). I then tested the competitiveness between Aldh1a2-OE cells and the empty vector control cells in the
primary tumor site. An equal number of Aldh1a2-OE cells and control cells were orthotopically engrafted into the gastrocnemius muscle of nude mice. The resulting tumors showed relatively similar numbers of Aldh1a2-OE cells and control cells (Figures 4.1F-G), suggesting that overexpression of Aldh1a2 preferentially impairs competitive fitness in the lungs.

**Figure 4.17. Identification of in vivo metastasis suppressors.** A. Effect of stable overexpression of each candidate gene on the metastatic ability of KPCC cells after tail-vein injection. Each point represents the relative metastatic area in the lungs of an individual nude mouse. Data
represented as mean ± SD. (*P<0.05, One-Way ANOVA). B. Metastatic burden following tail-vein injection of control or Aldh1a2 overexpressing cells in an independent set of animals quantified by relative area of metastatic lesions. Data represented as mean ± SD. (** P<0.01). C. Representative H&E images of lungs from mice injected with Aldh1a2-OE and Empty-Vector control cells. Scale 1.5mm. D. Confocal lung metastasis after competition between Aldh1a2-OE (labeled green) and empty vector control cells (labeled red) injected at 1:1 ratio into the tail vein of nude mice. Scale 1.5mm. Inset Scale 100µm. E. Quantification of metastases area. (*** P<0.001). F. Confocal microscopy of orthotopic competition between Aldh1a2-OE cells (labeled green) and control cells (labeled red) injected at 1:1 ratio into the muscle of the extremity. Scale 1mm. Inset Scale 100µm. G. Quantification of tumor tissue area for each fluorescent reporter in tumor sections.
**Figure 4.18. Functional validation of suppressors of metastasis.**

**A.** Quantitative RT-PCR data validating the overexpression of candidate genes using lentiviral constructs in metastatic KPCC cells. Each dot represents an individual animal. Data represented as mean ± SD. (* P<0.05).  

**B.** The number of metastatic lesions in cells over-expressing different candidate genes. Data represented as mean ± SD. (⁎ P<0.05).  

**C.** GSEA of MCs using the list of genes that are significantly affected by knockout of retinoic acid receptor genes found in Tanoury et al, 2014.  

**D.** Kaplan-Meier survival data of human patients from TCGA sarcoma dataset segregated based on the copy number of ALDH1A2.  

**E.** Western blot Aldh1a2 in metastatic KPCC cells transduced with empty vector or over-expression vector for Aldh1a2.  

**F.** Number of metastatic lesions in cells over-expressing Aldh1a2 in an independent mouse cohort. Each point represents an individual animal. Data represented as mean ± SD. (⁎⁎ P<0.01).
4.4 Discussion

Competition and selection of tumor clones are essential elements of cancer evolution (Marusyk & Polyak, 2010). In this study, I prospectively traced tumor clones over the course of disease progression, from sarcoma initiation, growth, and local recurrence to metastasis. Using models of multicolor fluorescent reporters, intravital microscopy, and CRISPR-Cas9 based barcoding, I demonstrated the functional roles of different tumor clones throughout the major stages of sarcoma progression. These models closely recapitulated important aspects of tumor evolution. First, in our models, tumor clones were initiated with the same founder mutations, Kras<sup>G12D</sup> and homozygous deletion of p53. Second, contrary to many other autochthonous tumor models, such as non-small cell lung cancer, where the individual tumor clones are spatially separated throughout the organ with limited contact, the tumor clones in our models grow in close proximity as part of a single tumor mass where clonal intermixing is common (Chuang et al., 2017). Third, our model captured the stepwise progression observed in human cancers, where tumor clones from one primary location grow, disseminate, and metastasize to form multiple lesions in distant organs. Taken together, the models in this study mimic tumor evolution where distinct tumor clones occupy overlapping tissue habitats and progress via competition and selection.

The K-sgP53-Cas9 model generated different indels targeted by the sgRNA within the p53 gene, which served as barcodes for different tumor clones. Although the majority of the indels induced are frameshift mutations or gene truncation (Pinello et al., 2016; Ran et al., 2013), it is still possible that different indels may confer distinct functional consequences on the gene. In our previous work, I tested the expression of p53 in cells generated from sarcomas in the KP-sgP53-Cas9 model and found that p53 expression is lost as a result of the various indels (Huang et al,
Furthermore, because our results with the KP-sgP53-Cas9 model mirror the results from the KPCC model, where p53 alleles were deleted by Cre-Lox recombination, I concluded that the sarcomas in the K-sgP53-Cas9 model has lost p53 function. Based on the fluorescent reporter model and DNA barcoding, I showed that primary tumor progression is associated with the expansion of a dominant clone. Moreover, I demonstrated that local recurrence and metastasis are driven by distinct clonal dynamics, where local recurrences are polyclonal and advanced distant metastases are clonal in origin. Importantly, to investigate the clonal origin of metastasis, the tumors were amputated at an early stage, prior to the establishment of a dominant clone in the primary tumor, indicating metastatic ability is gained before clonal sweep in the primary tumor. Amputation of late stage tumors where clonal dominance has occurred could not be performed effectively because of mortality during surgery and rapid onset of local recurrence in the survivors, which prevented the study of gross metastases. By transplanting polyclonal tumors that were metastatic into nude mice and allowing the tumors to reach late stage showed MCs could be either the minor or the dominant clones in late stage primary tumors. This is consistent the notion that metastatic ability may be independent of primary tumor dominance and is in agreement with other reports sequencing matched tumors and metastases in human patients where metastatic seeding can derive from rare or truncal clones (Bao et al., 2018; Fidler, 2003; Gundem et al., 2015; Makohon-Moore et al., 2017).

Through transplantation of the same polyclonal tumor into multiple animals, I demonstrated that the metastatic clone arises through a deterministic selection process instead of neutral population drift. The statistical methods I used to exclude the probability of multiple clones being labeled with the same reporter by chance will be valuable for future investigations of clonal dynamics. By tracing the early steps of the metastatic cascade, our data demonstrated
polyclonal CTCs and polyclonal early metastatic lesions in some animals. This suggests that metastatic seeding can be polyclonal. As such, the critical selective bottleneck leading to the monoclonality of clinically relevant, advanced metastases may be in the adaptation and colonization of the lungs. Although I cannot rule out the possibility that a few cells from other clones in the early polyclonal metastatic seeds may be residing below the threshold of detection in advanced stage metastases, in these models the vast majority of animals with detectable, clinically relevant metastases are clonal. Our data also suggest that selective pressure and competition between clones may be acting at each successive step of the metastatic cascade. Furthermore, it is conceivable that in certain types of cancer a primary tumor could harbor more than one clone each with the MC phenotype and clonally derived metastases may be generated by multiple tumor clones. Nonetheless, in our models, when the clones were transplanted individually, the inability of most clones to form metastases or at a reduced rate to form metastases suggests that the MC cells likely can outcompete the other clones over time, leading to the monoclonality of advanced metastases.

Previous lineage tracing studies for metastases were in epithelial cancers and found that metastases can derive from multiple clones (Cheung et al., 2016; Maddipati & Stanger, 2015; Reeves, Kandyba, Harris, Del Rosario, & Balmain, 2018). In the breast cancer study, the labeled tumor cells were generated from single-cell derived tumor organoids. The progenies of these organoids may not faithfully represent the divergent properties that exist between clones of an autochthonous tumor developed over the course of the disease (Cheung et al., 2016). Reeves et al labeled tumor cells in clonally derived skin squamous cell carcinoma after the tumors had progressed and found polyclonal metastases (Reeves et al., 2018). The difference in the clonality of sarcoma metastases in our model may be related to the biology of mesenchymal tumors, or more
likely, in the method in which tumor clones are defined. In some of the previous studies, tumor
cells are randomly labeled late during tumor progression (Cheung et al., 2016; Reeves et al., 2018).
In these studies, it is unclear whether distinctly labeled tumor cells are different tumor clones that
exhibit phenotypic diversity or if they are progenies of the same clonal population with similar
biological properties but are labeled differently by chance. In our model, tumor clones were
distinguishable at tumor initiation, so that the competition and selection of different clones can be
traced throughout different stages of diseases progression. In the pancreatic cancer model, where
tumor clones are traced from initiation, even though regional metastases were polyclonal, the large
distant metastases to the lungs were frequently monoclonal, consistent with our observations
(Maddipati & Stanger, 2015).

Genes that are associated with metastasis are usually identified by comparing matched
pairs of established metastases and primary tumor (Lambert et al., 2017). Differential expression
of genes that underlie the metastatic properties within phenotypically different primary tumor
clones has not been studied and may shed light on the early metastatic process. By focusing on
genes altered in MCs, I identified suppressors of metastasis, namely Aldh1a2, Reck and Rasd1.
The expression of these genes is altered in the primary tumor preceding cells reaching distant organ
and may represent the early gene expression changes that contribute to metastasis initiation.
Aldh1a2 is a member of the aldehyde dehydrogenase family of enzymes that catalyze oxidation of
aldehydes to carboxylic acids and participates in the biosynthesis of retinoic acid (RA)
(Niederreither & Dolle, 2008). Its role in cancer metastasis has not been described previously.
During tissue development, members of the Aldh1a family, including Aldh1a2 is known to play
pivotal roles. Loss of Aldh1a2 disrupts mesodermal patterning and formation of the forelimb bud
(Niederreither & Dolle, 2008). More recently, genetic knockout of retinoic acid receptors, the
transcriptional factors downstream of RA signaling, is reported to inhibit cell adhesion to matrix substrates (Al Tanoury et al., 2014). In our study, Aldh1a2 expression was suppressed in MCs, and genes associated with RA signaling were negatively enriched. Pathways involved in cell adhesion and cell-matrix interaction were among the top cellular processes negatively enriched in MCs. Rasd1 encodes a member of the Ras superfamily of GTPases and may play a role in cell-matrix interactions and cell adhesion (Tu & Wu, 1999). The RECK protein is a membrane bound glycoprotein that functions in extracellular matrix remodeling by inhibiting the activity of metalloproteinases (Walsh et al., 2015). Expression of Reck is down-regulated in breast cancer via hypermethylation and contributes to tumor invasion and angiogenesis (Walsh et al., 2015). Future studies into the molecular mechanisms in which these candidate genes regulate metastasis may reveal novel pathways and potential interactions between metastasis suppressor genes that contribute to the metastatic process.

Overall, by modeling clonal competition and selection, I describe in detail the clonal dynamics during each step of sarcoma progression, from tumor initiation, local recurrence after therapy, and to advanced metastasis. The clonal origin of advanced metastases offers the possibility of targeting specific clonal populations within a heterogeneous tumor to impede disease progression. The model I used here traced clonal evolution from the beginning of the tumorigenic process; however, during tumor evolution, each clone may give rise to distinct subclones, creating additional heterogeneity. In future studies, approaches with single-cell resolution such as single-cell RNAseq may help identify additional transcriptional heterogeneity within individual clonal populations. Furthermore, the metastatic process involves a complex interplay between the tumor intrinsic alterations, such as changes in the genome and epigenome with extrinsic alterations such as the tumor microenvironment. Systematic investigation of the tumor genome and
microenvironmental interactions at the clonal level may reveal new insights into the selective forces that enabled certain clones to gain the metastatic advantage.

4.5 Methods

Mouse Models

Animal studies were performed in accordance with approved protocols from the Duke University Institutional Animal Care and Use Committee. The R26R-Confetti mice were obtained from Jackson Laboratories (Stock No: 013731). These mice were crossed with previously described KRAS^{LSL-G12D}; p53^{lof} mice (Kirsch et al., 2007) to generate the KRAS^{LSL-G12D}; p53^{lof}; R26R-Confetti/Confetti (KPCC) mice. The KRAS^{LSL-G12D} mice were provided by Dr. Tyler Jacks (MIT, MA) and the p53^{lof} mice were provided by Anton Berns (NKI, Netherlands). To induce tumors, adenovirus expressing Cre recombinase were mixed with 2M CaCl$_2$ in DMEM and were directly injected into the hind gastrocnemius muscle of 7-10-week-old mice (Kirsch et al., 2007). Tumor volume was calculated using the formula $V = (\pi \times L \times W \times H)/6$, with L, W, and H representing the length, width, and height of the tumor in mm. All mice were on a mixed genetic background. The K-sgP53-Cas9 mice were generated by crossing Kras^{LSL-G12D} mice with Rosa26-LSL-Cas9 (Stock No: 024857, Jackson Laboratory) mice or Rosa26-Loxp-Cas9 (J. Huang et al., 2017). Tumor initiation in K-sgP53-Cas9 mice was as described previously (J. Huang et al., 2017). For tumor transplantation experiments, 6 to 7-week-old Foxn1^{nu/nu} were purchased from Duke University Division of Laboratory Animals and Resources. Tumor cells were diluted in 1XPBS for injection. The cells were injected into the hind gastrocnemius at 500,000 cells per mouse or the pinnae at 100,000 cells per mouse.
METHOD DETAILS

Amputation Surgery, Orthotopic Transplantation, Tail-Vein and Pinna Injections

For amputation surgery, primary tumors in KPCC and K-sgP53-Cas9 mice were amputated when their size reached an average of 250-350mm$^3$. The tumor-bearing limb was amputated under 2% isoflurane induced anesthesia. Subcutaneous injection of 0.1mg/kg buprenorphine and 5mg/kg ketoprofen and 0.25% bupivacaine were administered to provide analgesia. The mice were followed for the development of lung metastases. The amputated tumors were digested into single cells using a mixture of collagenase, dispase, and trypsin (Wei et al., 2015). Mice were examined every 2 days following amputation until they developed signs of systemic illness, such as hunched posture, lethargy, or significant weight loss. When clear symptoms were observed, the mice were euthanized, and the lungs were harvested and analyzed for the presence of metastases. To observe advanced metastasis, usually complete local control of the tumor is required. The animals that developed local recurrence following surgery were euthanized and the recurrent tumors were harvested for histology and microscopy. For orthotopic transplants, the tumors were amputated at early stage. The nude mice were monitored using the same protocol as the autochthonous model.

Tail-vein injection experiments were performed by injecting 250,000 to 500,000 tumor cells diluted in 1XPBS into the tail-vein of nude mice. The mice were euthanized when they developed clear symptoms of metastases, and the lungs were harvested. Pinna injections were performed by injecting single cell suspensions of primary KPCC tumor cells with a 27G needle into the pinna of 6-7-week-old nude mice. Successful injections were confirmed by swelling immediately after injection. The pinnae were imaged using a two-photon microscope (FluoView FV1000, Olympus, Central Valley, PA).
**Isolation of Circulating Tumor Cells**

Blood from KPCC mice was harvested via cardiac puncture and the animals were euthanized. The blood was then centrifuged in Ficoll-Paque PLUS (GE-Health Care, PA) and the nucleated cells were isolated from the buffy coat. The nucleated cells were immediately stained with CD45-PECy7 anti-mouse antibody (BD, NJ) and analyzed using MoFlo Astrios Flow cytometer (Beckman Coulter) for RFP, YFP, CFP, and GFP expression. Propidium Iodine (ThermoFisher, MA) was used as live/dead indicator.

**In Vivo Competition Assay**

The empty vector puro cells were generated by transducing parental cells with lentivirus carrying the pCDH-CMV-MCS-EF1α-Puro vector (System Biosciences, CA). For competition in metastasis, 250,000 cells with the over-expression vector were mixed with 250,000 empty vector cells in 1XPBS. The mixture was injected into nude mice via the tail-vein. For competition in orthotopic tumors, 150,000 cells from each group were mixed in 1XPBS and injected into the left gastrocnemius muscle of nude mice. Quantification of the metastases area and tumor area was done using FIJI (Version 2) by calculating the area of each fluorescent reporter with intensity threshold.

**Radiation Treatment**

Sarcoma irradiation was performed using the X-RAD 225Cx small animal image-guided irradiator (Precision X-Ray) (Newton et al., 2011). The tumors were irradiated at an average of 350mm³. The radiation field was centered on the target by fluoroscopy with 40-kilovolt peak (kVp), 2.5 mA x-rays using a 2-mm aluminum filter. Radiation therapy was performed with parallel-opposed
anterior and posterior fields at an average dose rate of 300 cGy/min prescribed to midplane with 225 kVp, 13 mA x-rays using a 0.3-mm copper filter and a collimator with a $40 \times 40\,\text{mm}^2$ radiation field at treatment isocenter (Moding et al., 2013). The mice were placed under anesthesia with continuous 2% isoflurane with 2L of oxygen per minute. The mice were also treated with 0.1mg/kg buprenorphine and 5mg/kg of ketoprofen, and local injection of 0.25% bupivacaine for analgesia. For autochthonous sarcomas, an average of 35 Gy was delivered in a single fraction, and for transplanted tumors in nude mice, the total dose was reduced to 25 Gy in a single fraction because of increased normal tissue injury observed in Foxn1nu/nu mice.

**Histology**

For frozen sections, the lung samples were inflated with 4% PFA through the trachea. The tumor and lung samples were fixed in 4% paraformaldehyde at 4°C for 1 hour and 3 hours respectively, followed by incubation in 15% and 30% w/v sucrose for 24 hours. The tissues were then embedded in Optimum Cutting Temperature medium (OCT, Tissue-Tek) and frozen on dry ice. 10-12µm sections were cut using a Cryostat (Leica CM1950). Sections were taken 50-80µm apart across 1-2mm of tissue to limit sampling bias. The tissue sections were mounted with Prolong Diamond Antifade Mountant with or without DAPI (ThermoFisher). For formalin-fixed and paraffin-embedded sections, samples were fixed in formalin overnight, washed with PBS and ethanol, followed by embedding in paraffin. The tissues were cut at 5µm. At least 1-2 mm of tissue from each sample were analyzed. H&E images were taken with Zeiss Axio Imager Z2 (Zeiss, Germany) at either 10X or 20X, and the entire sections were scanned.

**Confocal and Intravital Microscopy**
Fluorescent images of primary and transplanted tumors were taken using Leica SP8 confocal microscope with a 20X and 40X objective. The argon laser 488 nm line was used to excite GFP, 514 nm line for YFP, 561 nm red diode laser for RFP, 458 nm laser line for CFP and UV laser line for DAPI. The collection wavelengths for GFP, YFP, RFP, CFP, and DAPI were: 493-510 nm, 537-565 nm, 612-650 nm, 468-498 nm, and 417-497 nm, respectively. Airy unit 1 was used for all images. The images were taken using Leica Application Suite X. Intravital imaging of sarcoma ear allograft was performed using a two-photon microscope (FluoView FV1000, Olympus, Central Valley, PA). This microscope is equipped with a 680–1050 nm tunable ultrafast laser (Spectra-Physics, CA). To image transplanted tumors in the ear pinna, tumor-bearing mice were anesthetized with 2% isoflurane, and the ear pinna was fixed to the stage using double-sided tape. YFP was excited by 860–980 nm laser and detected using a 495/540 nm bandpass filter. RFP was excited by 980 nm laser and detected using a 575/630 nm bandpass filter. CFP was excited by 830-850 nm laser with a 460-500 nm bandpass filter. Collagen was imaged by second harmonic generation using 860–920 nm excitation and 420/460 nm detection. Image processed with Imaris Microscopy Image Analysis Software (v 9.30).

Tumor Digestion, Cell Sorting, and Cell Culture

Primary sarcoma tumors from KPCC mice were mechanically and enzymatically dissociated into single cell suspensions as previously described (Wei et al., 2015). Briefly, the tumors were mechanically dissociated into small pieces and enzymatically digested with a mixture of 10 mg/mL of collagenase IV (Worthington), 2.4U/ml of Dispase (Stem Cell Technologies, BC), and 0.05% trypsin (ThermoFisher). Red blood cells were excluded with ACK lysis buffer (ThermoFisher). FACS for isolating the R26R-Confetti fluorescent proteins was performed using
Moflo Astrios Sorter (Beckman Coulter, CA) at the Duke Cancer Institute Flow Cytometry Core (Duke University, NC). CFP was detected by 405 nm violate laser with a 488/59 bandpass filter, and RFP was detected by 561 nm red laser with a 579/16 bandpass filter. GFP and YFP were detected by 488 nm argon laser with a 510/20 bandpass filter for GFP and a 546/20 bandpass filter for YFP. Propidium iodide was used to remove dead cells. The sorted cells were collected in PBS and cultured in DMEM (ThermoFisher) supplemented with 10% FBS.

Cell Lines
Sorted cells for tail-vein injections and metastatic cells used for over-expression experiments were from digested primary KPCC animals. The purity of the different sorted populations was checked by FACS after the cell lines were passaged once in vitro. All cell lines were confirmed to be negative for mycoplasma using the MycoAlert Detection Kit (Lonza, Brazil).

DNA Isolation, PCR and Targeted Deep Sequencing
Multiple pieces of the primary tumor masses and multiple lesions from the lung metastases of K-sgP53-Cas9 mice were harvested. The bulk DNA from the harvested tissues were isolated using DNeasy Blood and Tissue Kit (Qiagen) following manufacturers’ protocol and then sequenced. To sequence the CRISPR-Cas9 induced barcodes, the 200bp region near the sgRNA target site was PCR amplified using proof-reading AccuPrime Taq Polymerase (ThermoFisher). The forward and reverse primer sequence are 5’CAGAAGCTGGGGAAGAAACA3’, and 3’GTAGGGAGCGACTTCACCTG5’ respectively. The PCR products were sequenced by Massachusetts General Hospital DNA Core (Boston, MA). Sequencing depth were >5,000X. Sequences representing 1% of total reads for each sample was used as a cut-off. PCR to check
the recombination of LSL-Kras\textsuperscript{G12D} and p53 deletion were performed as described previously (DuPage, Dooley, & Jacks, 2009; Marino, Vooijs, van Der Gulden, Jonkers, & Berns, 2000). DNA was extracted from sorted cells of primary KPCC tumors. The primers for LSL-Kras\textsuperscript{G12D} are 5’ GTC TTT CCC CAG CAC AGT GC 3’, 5’ CTC TTG CCT ACG CCA CCA GCT C 3’, 5’ AGC TAG CCA CCA TGG CTT GAG TAA GTC TGC A 3’. The primers for p53 are 5’ CAC AAA AAC AGG TTA AAC CCA G 3’, 5’ GAA GAC AGA AAA GGG GAG GG 3’, 5’ AAG GGG TAT GAG GGA CAA GG 3’. Tail genomic DNA is used as the negative control.

**RNA Isolation and RNA-Sequencing**

Primary KPCC tumor cells were FACS sorted into MCs and Non-MCs, and the RNA was extracted using Norgen RNA Purification Kit (Norgen, ON). RNA from bulk metastatic lesions were isolated using Directzol RNA kit (Zymogen, CA). Standard manufacturer’s protocol is used for both kits. The total RNA for RNA-sequencing was performed at the Duke Center for Genomic and Computational Biology (Duke University, NC). The RNA was prepared using KAPA stranded mRNA-seq kit (Sopachem, Belgium). The first 50 bases of these libraries were sequenced on a HiSeq 4000 system (Illumina, CA). Over 50 million reads were generated per sample.

**Molecular Cloning, Lentivirus Generation and Transduction**

To generate viral vectors over-expressing the six-candidate metastasis suppressor genes (Chad, PodN, Fibin, Rasd1, Reck, and Aldh1a2), cDNA constructs were ordered from Origene (TrueORF, Origene, MD). The cDNA for each gene was cloned into EcoRI and FseI sites of pLenti-MCS2-CMV-GFP-2A-Puro or pCDH-CMV-MCS-EF1α-Puro vector (System Biosciences, CA). Successful cloning was confirmed via Sanger sequencing. Lentivirus was generated using the
Delta8.2 and VSV-G packaging plasmids. To generate stable over-expression cells, virus carrying the cDNA constructs and empty vectors were transduced into the cells with 8μg/mL of polybrene. After two days, the cells were selected with puromycin and over-expression was validated by qPCR. Over-expression of Aldh1a2 was also validated with Western Blot.

**RT-qPCR**

Reverse-transcription of RNA to generate cDNA was done using iScript™ gDNA Clear cDNA Synthesis Kit following the manufacturer’s instructions (Bio-Rad, CA). Quantitative PCR (qPCR) was performed using TaqMan Fast Advanced Master Mix (ThermoFisher, MA) with QuantStudio Flex 6 (ThermoFisher, MA). The denaturation step was carried out at 95°C for 20 seconds; the annealing/extension step was carried out at 60°C for 20 seconds, and the total cycle is 40. Relative gene expression was determined using ΔΔCt with beta-Actin as the internal control gene.

**Western Blot**

Whole-cell protein extracts from cells were denatured, separated on SDS–PAGE gels, and dry-transferred to nitrocellulose membranes. After blocking in 5% milk in Tris-buffered saline–Tween, membranes were probed overnight at 4°C with either vinculin (1: 5,000, Cell Signaling Technology) or Aldh1a2 (1:500; ProteinTech, IL). After incubation, the secondary antibody was added and incubated for 1 hour in room temperature. Results were detected using SuperSignal™ West Femto Maximum Sensitivity Substrate, according to the manufacturer's instructions (ThermoFisher, MA).
QUANTIFICATION AND STATISTICAL ANALYSIS

Quantitative Image Analysis

Quantification of images was performed by manually counting the fluorescent reporter positive cells and total DAPI+ cells using FIJI (version 2.0 with cell counter plugin). For each tissue section, I either scanned the entirety of the slide or examined approximately 20-30 areas of view. For categorization of the images, each picture was categorized as monochromatic if ≥90% of fluorescent cells were of a single color, as dichromatic if the two colors represented ≥90% of total fluorescent cells, and as polychromatic if at least 3 colors were present, each representing 10-20% of total fluorescent cells. Quantification of metastases in the lungs was done by measuring the number of metastatic lesions and the total area of the metastatic lesions using FIJI (version 2.0). For both measurements, approximately 5 sections were randomly selected across 1-2 mm of lung tissues and the average number of metastases and the average area of metastatic lesions per animal was used for statistical comparison. For in vivo competition assays, approximately 5 sections from across 1-2 mm of lung tissues or tumor tissues were randomly examined. The area of each fluorescent reporter was calculated by color threshold using FIJI (version 2.0) and divided by the total tissue area of the sections.

RNA-Seq Analysis

RNA-seq analysis was performed on MCs (6 biological replicates with 5 technical replicates), Non-MCs (8 biological replicates with 1 technical replicate), and Lung-Mets (5 biological replicates and 1 technical replicate). The RNA sequencing reads were processed using the Trimmomatic (v0.36) (Bolger, Lohse, & Usadel, 2014) to trim low-quality bases and Illumina sequencing adapters from the 3’ end. Only reads that were 20nt or longer after trimming were
kept for further analysis. Reads were mapped to the mouse transcriptome using the STAR aligner (v2.5.2b) (Dobin et al., 2013). Gene counts were compiled using the HTSeq tool implemented in STAR program (Anders, Pyl, & Huber, 2015). The reference sequence and GTF file were obtained from the UCSC GRCm38 bundle available from the iGenomes collection. Normalization and differential expression were carried out using the DESeq2 (v1.16.1) (Anders & Huber, 2010). Gene differential expression was performed within the framework of a negative binomial model using R (v3.4.4) (R Core Team, 2016). Statistical analyses were adjusted for multiple testing within the framework of control of the false-discovery rate (Bass, Swcf, Dabney, & Robinson, 2015; Benjamini & Hochberg, 1995). Statistical analyses were mainly scripted using the R statistical environment (R Core Team, 2016) along with its extension packages from the comprehensive R archive network (CRAN; https://cran.r-project.org/) and the Bioconductor project. (Gentleman et al., 2004). I considered genes with P<0.005 for hypothesis generation. Gene set enrichment analysis (GSEA) was performed to identify differentially regulated pathways and gene ontology (Subramanian et al., 2005). Statistically significant gene sets are defined as Nom p-val<0.05 and FDR q-val<0.05. Gene sets describing similar processes were grouped together, and the average NES and FDR q-val were determined for each group. The genes list for CINSARC and Rar-KO enrichment were obtained from Chibon et al, and Tanoury et al (Al Tanoury et al., 2014; Chibon et al., 2010). For selection of genes in the Rar-KO list, genes that are differentially expressed between Rar-KO and WT cells in RNA-seq were selected (P_adj<0.01 was used as a cutoff). Copy number information for ALDH1A2 and corresponding survival analysis from The Cancer Genome Atlas (TCGA) was accessed using UCSC Xena (https://doi.org/10.1101/326470).

**Statistical Analyses**
Graphs and statistics for most biological experiments were generated using the GraphPad Prism software (Prism 6) unless indicated otherwise. Statistical significance was calculated using the two-tailed unpaired Student's t-test for comparison between two groups or one-way ANOVA for comparison between multiple groups unless described otherwise. Holm-Sidak method was used post-hoc correction of multiple comparisons in ANOVA, adjusted P<0.05 is considered significant.

Statistical Theory

I. The Probability of Monoclonal Origin of Advanced Lung Metastases in the KPCC Model

I. Background

Cell labeling by R26R-Confetti allele is a stochastic process with a slight bias in the representation of each color combination (Baggiolini et al., 2015; Maddipati & Stanger, 2015; Snippert et al., 2010). In our study, I observed a total of 8 different color combinations. A limitation of fluorescent reporter based lineage tracing methods is that it is plausible that there are more tumor clones than represented by reporter labeling. However, in our study, the number of clones by fluorescent reporters matches closely with the number of clones with CRISPR-Cas9 barcoding. Nonetheless, since cell labeling is stochastic, there may be ≥ 2 different clones labeled by the same reporter. As such, I considered the possibility that although I observed that all of the advanced metastasized lesions in each animal expressed the same reporter, they may reflect different tumor clones. To test this, I reasoned that the probability of multiple tumor clones expressing the same reporter colonizing the lungs would decrease with increasing sample size. In other words, with a sufficient number of animals, I would either observe some animals with multi-
colored lung metastases; otherwise I can reasonability accept that the single colored lung metastases are originated from a single tumor clone. Here, I describe the statistical methods used to test this notion.

I hypothesized that advanced lung metastases are derived from a single tumor clone ($H_1$). Our null hypothesis is that the lung metastases are derived from multiple tumor clones expressing the same fluorescent reporter ($H_0$). It is important to note that to demonstrate $H_1$, I would theoretically need an infinite number of observations. As such, I will conservatively test the probability that $\geq 2$ tumor clones expressing the same reporter occurring over a set of observations. I expect that with an increasing number of observations where lung metastases in mice are labeled by one fluorescent color, the probability of multiple clones expressing the same reporter leading to metastasis will decrease.

I. II Notations:

- $N$: number of mice, assuming they are all identical and independent
- $q_i$: the probability of color $i$ for a tumor clone
- $m$: number of tumor clones expressing 1 or more reporters, $m = 1, 2, 3, ...$
- $p_{i,m}$: for mouse $i$, the probability that $m$ different tumor clones have the same color.

And marginally, $p_m$ stands for the probability that $m$ different tumor clones have the same color in all mice.

III Derivations

For each mouse $i$, there are $m$ number of different tumor clones. Assume each clone is independent, then the probability that all $m$ clones have the same color in mouse $i$ is:

$$p_{i,m} = \sum_{i=1}^{\theta} q_i^m$$
I assume that all $N$ mice are independent, the probability that the lung metastases are the same color in each mouse over $N$ mice is:

$$p_m = \left( \sum_{i=1}^{8} q_i^m \right)^N$$

The observed fluorescent reporter frequency in our study is (Figure 3E):

<table>
<thead>
<tr>
<th>Color</th>
<th>RFP ($q_1$)</th>
<th>YFP ($q_2$)</th>
<th>GFP ($q_3$)</th>
<th>CFP ($q_4$)</th>
<th>RFPYFP ($q_5$)</th>
<th>RFPCFP ($q_6$)</th>
<th>YFPCFP ($q_7$)</th>
<th>YFPGFP ($q_8$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage</td>
<td>0.3883</td>
<td>0.2181</td>
<td>0.0034</td>
<td>0.0594</td>
<td>0.1385</td>
<td>0.1386</td>
<td>0.0525</td>
<td>0.0012</td>
</tr>
</tbody>
</table>

Practical estimates for $q_i$ can be calculated using from the above table. This $q_i$ estimate is based on a total of 19 mice that developed lung metastases observed in our study.

$N = 19$ The following plot shows the relationship between $p_m \sim m$:
Since the probability of multiple clones expressing the same color forming metastases decreases with increasing \( m \) (i.e. it is less likely for 3 clones to express the same reporter compared to 2 clones expressing the same reporter), the most conservative calculation of probability \( p_m \) can be based on \( m=2 \), where 2 clones expressing the same fluorescent reporter form metastases.

IV Calculations of Probability from Observations

From our experimental observations, 1 fluorescent reporter was observed within each mouse for all 19 mice. Let \( n_i \) be the number of mice that have color \( i \), \( \sum_{i=1}^{8} n_i = 19 \). Then \( n_i \) follows a multinomial distribution with parameter \( (n_1, n_2, \ldots, n_8; N, q_1, q_2, \ldots, q_8) \). Thus, the probability of our observation by chance is:

\[
p_m = \binom{N}{n_1} q_1^{n_1} \cdot \binom{N-n_1}{n_2} q_2^{n_2} \cdot \binom{N-n_1-n_2}{n_3} q_3^{n_3} \cdot \ldots \cdot \binom{N-n_1-n_2-\ldots-n_7}{n_8} q_8^{n_8} = \frac{N!}{n_1! \cdot n_2! \cdot \ldots \cdot n_8!} q_1^{n_1} \cdot q_2^{n_2} \cdot \ldots \cdot q_8^{n_8}
\]

The distribution of colors found in animals with metastasis, \( n_i \), is as follows (Figure 3E):

<table>
<thead>
<tr>
<th>Color</th>
<th>RFP</th>
<th>YFP</th>
<th>GFP</th>
<th>CFP</th>
<th>RFPYFP</th>
<th>RFPCFP</th>
<th>YFPCFP</th>
<th>YFPYFP</th>
</tr>
</thead>
<tbody>
<tr>
<td>((q_1))</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>((q_2))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>((q_3))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>((q_4))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>((q_5))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>((q_6))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>((q_7))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>((q_8))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Assuming \( m = 2 \), the probability of all 19 mice following color distribution above is:

\[
p_2 = \frac{N!}{n_1! \cdot n_2! \cdot \ldots \cdot n_8!} q_1^{2n_1} \cdot q_2^{2n_2} \cdot \ldots \cdot q_8^{2n_8} = 9.71 \times 10^{-15}
\]
I also considered the probability of more than 2 clones expressing the same reporter contributing to metastases, where \( m \geq 2 \). The value of \( p_m \) decreases logarithmically with increasing \( m \). Inclusion of probabilities \( m > 2 \) does not significantly deviate from \( p_2 \), as described below:

<table>
<thead>
<tr>
<th>( m )</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>( 9.71 \times 10^{-15} )</td>
</tr>
<tr>
<td>3</td>
<td>( 2.94 \times 10^{-25} )</td>
</tr>
<tr>
<td>4</td>
<td>( 8.90 \times 10^{-36} )</td>
</tr>
<tr>
<td>5</td>
<td>( 2.70 \times 10^{-46} )</td>
</tr>
<tr>
<td>6</td>
<td>( 8.17 \times 10^{-57} )</td>
</tr>
<tr>
<td>7</td>
<td>( 2.47 \times 10^{-67} )</td>
</tr>
<tr>
<td>8</td>
<td>( 7.49 \times 10^{-78} )</td>
</tr>
<tr>
<td>9</td>
<td>( 2.27 \times 10^{-88} )</td>
</tr>
</tbody>
</table>

I. Conclusion

Given our sample size \( N = 19 \), the probability of more than one clone expressing the same reporters forming lung metastases is approximately \( 9.71 \times 10^{-15} \). Therefore, I am confident that statistically, advanced lung metastasis in sarcoma is likely driven by a single clone.

II. Deterministic (i.e. Selection) versus Stochastic Process (i.e. Neutral Drift) Driving Metastasis

II. I Background

To determine whether the monoclonality of metastases arise through biological selection, which is a deterministic process, or neutral drift, which is a stochastic process, I orthotopically transplanted 5 primary KPCC sarcomas into nude mice. I reasoned that in neutral drift, different tumor clones have equal capacity to colonize the lungs. Therefore, I would observe lung metastases in different individual nude mouse transplanted with the same tumor, express different fluorescent reporters, resulting from random chance. On the other hand, if the tumor clone that
gives rise to metastases is selected due to biological differences among tumor clones, I would observe the opposite. The lung metastases will express the same reporter in each nude mouse transplanted with the same tumor. Indeed, I observed the latter. To statistically determine the likelihood of our observation by chance, I used the intracluster correlation coefficient (ICC, denoted as Rho or ρ). ICC assesses the relatedness between subjects within the same cluster. In our case, each set of nude mice transplanted with the same primary KPCC tumor is considered as one cluster. ICC can range from 0 to 1, where 0 indicate the subjects within a cluster is not related, and 1 indicate the subjects within a cluster is completely related. For our purpose, 0 represent the scenario where different clones formed metastases in nude mice of the same cluster (i.e. the lung metastases would express different reports), and 1 indicate a single clone generated metastases in all the nude mice of the same cluster (i.e. the lung metastases would express the one reporter).

I hypothesize that the same tumor clone colonizes the lungs in nude mice transplanted with the same primary tumor (H₁). In other words, the fluorescent cells found in the lungs will be the same for each nude mouse in a cluster. Our null hypothesis (H₀) is that the tumor clone forming lung metastases in each nude mouse will be random.

II. Notation

(1) N: number of clusters

(2) mi: number of mice developed lung metastases in a cluster i, where i = 1, 2, 3, ..., N, mi ≥ 2

(3) pk: the probability that a mouse has color k, where k = 1, 2, 3, ..., 8

(4) xik: number of mice with color k among mi mice in cluster i

(5) ρ: the intracluster correlation coefficient (ICC)

III Hypothesis
**H₀**: The metastatic clone in each mouse is randomly selected

**H₁**: One tumor clone is responsible for advanced metastasis

Under **H₀**, $X_i \sim (x_{i1}, x_{i2}, ..., x_{i8})$ follows a multinomial distribution with probability $\pi = (p_1, p_2, p_3, ..., p_8)$:

$$ X_i \sim \text{Multi} (m, \pi) $$

Then, the hypothesis is equivalent to

**H₀**: $\rho = 0$

**H₁**: $\rho = 1$

### II. Derivation and Calculation

Under the null hypothesis, the probability that all $m_i$ mice in cluster $i$ share the same color is:

$$ P_0(i, \rho) = \left( \sum_{k=1}^{8} p_k^{m_i} \right)^{1-\rho} $$

Thus, the probability that in all clusters, all $m_i$ mice share the same color is:

$$ P(N, \rho) = \prod_{i=1}^{N} P_0(i, \rho) = \prod_{i=1}^{N} \left( \sum_{k=1}^{8} p_k^{m_i} \right)^{1-\rho} $$

I reject H₀ if I observe matching colors from all $N$ mice with a $\rho$ approaching 1. Specifically, I reject H₀ if I observe matching colors from all $m$ mice when $P(N, \rho) < \alpha$.

Thus, the number of clusters ($N$) and the number of mice in each cluster can be represented as the following:

$$ P(N, \rho) = \prod_{i=1}^{N} \left( \sum_{k=1}^{8} p_k^{m_i} \right)^{1-\rho} < \alpha $$

$p_k$ Could be estimated using the observed florescent reporter frequency:
<table>
<thead>
<tr>
<th>Color</th>
<th>RFP</th>
<th>YFP</th>
<th>GFP</th>
<th>CFP</th>
<th>RFPYFP</th>
<th>RFPCFP</th>
<th>YFPCFP</th>
<th>YFPGFP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>((q_1))</td>
<td>((q_2))</td>
<td>((q_3))</td>
<td>((q_4))</td>
<td>((q_5))</td>
<td>((q_6))</td>
<td>((q_7))</td>
<td>((q_8))</td>
</tr>
<tr>
<td>Percentage</td>
<td>0.3883</td>
<td>0.2181</td>
<td>0.0034</td>
<td>0.0594</td>
<td>0.1385</td>
<td>0.1386</td>
<td>0.0525</td>
<td>0.0012</td>
</tr>
</tbody>
</table>

Using \(p_k\) I determined the clusters (N), and the number of mice (m) in each cluster needed to achieve a certain ICC (\(\rho\)) level, given a 0.05 significance level:
<table>
<thead>
<tr>
<th>N</th>
<th>rho</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>0.4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>0.6</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>0.8</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>0.9</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>0.95</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>0.99</td>
<td>157</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>0.4</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>0.6</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>0.8</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>0.9</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>0.95</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>0.99</td>
<td>105</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0.4</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>0.6</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>0.8</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>0.9</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>0.95</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>0.99</td>
<td>79</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0.2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0.4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0.6</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>0.8</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>0.9</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>0.95</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>0.99</td>
<td>63</td>
</tr>
</tbody>
</table>
The corresponding plot is as follows:

**II.**

**IV**

**Conclusion**

Based on the number of KPCC tumors I transplanted (N=5) and that all the mice in each cluster (i.e., transplanted with the same KPCC tumor) developed metastases expressing the same fluorescent reporter (m=24), I conclude that ICC ($\rho$) >0.97 for P<0.05. Therefore, the metastatic ability of a tumor clone is probably not stochastic, but due to biological selection.
CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Conclusions

Intratumoral heterogeneity is a crucial, clinically relevant feature common to most cancers. The ability of cancer cells to continuously evolve in response to intrinsic genomic instability, extrinsic environmental and selective forces likely underpin therapy failure. Nonetheless, whether distinct tumor cell subpopulations can alter their cellular state, specifically between a stem-like state and a differentiated state, and how distinct tumor clones change throughout disease remains poorly understood. Answers to these questions have important biological and clinical implications. First, biological insights into the state of TPCs reveals the cellular organization of a tumor. A model of fixed TPCs suggests that the tumor is organized in a unidirectional hierarchy with TPCs at the apex. On the other hand, a dynamic or transient TPC state would suggest that the relationship between TPCs and non-TPCs is bidirectional. From a therapeutic perspective, TPCs are believed to be the most malignant subpopulation in a tumor because of their capacity for long-term self-renewal and proliferation (Burrell, McGranahan, Bartek, & Swanton, 2013; McGranahan & Swanton, 2017). They are also thought to be more chemoresistant and more metastatic (Kreso & Dick, 2014; O'Brien, Kreso, & Jamieson, 2010). Selective elimination of these cells may be more effective than destroying other cancer cells. However, in a dynamic or transient model of TPCs, the plastic non-TPCs can contribute to TPCs, and may render selective targeting of TPCs alone ineffective. Second, understanding the evolution of distinct tumor clones during cancer progression uncovers the functional contribution of the various tumor clones to different aspects of cancer. Despite our ability to capture snapshots of tumor evolution through deep sequencing of tumor genomes, the lack of in vivo tracing of tumor cells over time has limited
our knowledge into the functional roles and interactions of tumor clones. The identification of clones that contribute to different aspects of cancer such as growth, local recurrence, and distant metastasis allows the investigation of more precise molecular mechanism that governs each of these facets and identify novel therapeutic targets. Third, the combined information of stem cell state and clonal evolution describes an in-depth picture of cancer cell dynamics at a given time point and throughout its evolutionary history.

I hypothesized that sarcomas are comprised of heterogeneous cancer cell populations of dynamic states, and the cancer clonal architecture are subject to biological selection during different stages of tumor progression. In this thesis, I tested this hypothesis by systematically interrogating cancer cell dynamics in sarcoma using both the hierarchical model and the clonal evolution model. I identified a cell surface marker, CD146, which is highly enriched for tumor propagating capacity in common bone and soft tissue sarcoma subtypes. Furthermore, using serial transplantation of a combination of SP and CD146+ cells subpopulation, I showed that the TPC population are heterogeneous. Transcriptional profiling of the SP and CD146 cells identified common pathways such as Notch and TGF-β that were activated in both TPC populations. Targeting Notch signaling was able to significantly inhibit self-renewal. To investigate TPC dynamics, I generated KPC animals, where the tumor cells are permanently labeled with distinct fluorescent reporters. By transplanting NSP and SP cells of different colors, SP cells were shown to exist in a dynamic or transient state, where NSP cells can contribute to the SP population. The unstable state of TPCs is supported by single-cell RNAseq experiments that demonstrated transitional states common to the SP and NSP cells. The shared transitional states contributed by NSP cells may represent cells that are “reversing” to an SP state. Genetic ablation of SP cells in independent transplant experiments indicate that NSP cells can compensate for the loss of SP
population. Consistent with this, transplantation of tumors with ablated SP cells did not show any difference in tumor-propagating capacity. Taken together, these results suggest that the SP population, which are enriched for TPCs is a dynamic or transient state rather than a stable cell population.

The study of TPCs and non-TPCs examine tumor cell dynamics at a single time point during tumor development. To examine cellular dynamics throughout cancer progression, I used the labeled tumor subpopulations with different fluorescent reporters from the KPCC animals. These cells share identical founder mutations and are in growing in the same microenvironment, and therefore models distinct tumor clones. I also generated a complementary model that labels tumor clones with CRISPR-Cas9 based DNA barcodes. Lineage tracing experiments revealed that clonal diversity decreases during primary tumor progression. When the primary tumor is challenged with surgical or radiation therapy, local recurrence of the tumor is driven by multiple clones. In striking contrast, distant metastasis resulted from a single tumor clone that arose through biological selection. This clone, named metastasis-initiating clone (MC) rapidly forms metastatic lesions upon tail-vein injection, while the other clones have the minimal metastatic ability. MCs are characterized by up-regulation of genes related to more aggressive tumor behavior such as increased expression of genes associated with stem cells, and cell cycle, while genes associated with cell-cell and cell-matrix interactions are decreased. Forced expression highly suppressed genes such as Aldh1a2 significantly reduced metastatic colonization.

Collectively, the work described in this thesis is the first in-depth investigation of in vivo tumor cell dynamics in sarcoma. It revealed tumor cells are highly dynamic in their cell states and as clones; the clonal architecture is significantly altered during different aspects of disease progression. The dynamic or transient TPC state revealed in this work adds to the CSC model by
demonstrating that sarcomas may not be organized by a unidirectional hierarchy, but by bidirectional changes between cell states. It has been previously described that the CSC model and clonal evolution model are not mutually exclusive, where evolutionary pressures and genetic mutations can act on TPCs and their progenies to generate additional intratumoral heterogeneity.

My thesis work reconciles the two models further as models that describe intratumoral heterogeneity in relation to disease progression. Specifically, the clonal evolution model examines the tumor cell dynamics across the life history of the tumor, where selective pressures act to change the composition of tumor cells and their phenotype over time. Specifically, selection led to the monoclonal origin of all metastatic lesions in sarcoma. In contrast, the CSC model examines the different state of tumor cells at a particular given time point (i.e., when the tumor is harvested). My work demonstrated that the TPCs is a dynamic or transient state, where non-TPCs can give rise to TPCs, with cells existing as transitional states. From a translational view, an unstable TPC state challenges the concept of selectively targeting these cells as an effective therapeutic approach but identifying the molecular mechanism that underlies cell state transition may offer opportunities of combination therapy. Furthermore, the identification of monoclonal origin of metastasis provided novel targets for the most lethal stage of cancer progression. The distinct gene expression profiles and functionally essential genes such as Aldh1a2 in metastatic clones before the formation of distant lesions may represent vulnerabilities in the early metastatic process. Future investigations into the molecular and genetic mechanisms that control the TPC states and the MCs phenotype is likely to uncover novel therapeutic vulnerabilities against the evolving and heterogeneous tumor cells.

5.2. Future Directions
5.2.1 The Epigenome of TPCs and Metastatic Clones

Epigenetic mechanisms are responsible for the diverse cellular phenotype in normal physiology. Cell fate determination from quiescent pluripotency in stem cells to differentiated progeny is determined by modulation of the chromatin and DNA methylation landscape (Keller, 2005). Likewise, epigenetic aberrations play crucial roles in the phenotype of cancer cells. Indeed, disruptions in DNA methylation and chromatin modifications are recognized hallmarks of cancer development (Hanahan & Weinberg, 2011). It is widely accepted that many functionally important changes in gene expression that drive tumor progression are independent of genetic mutations. For example, upregulation of Notch and TGF-β signaling are necessary for self-renewal in TPCs of sarcoma and other cancers, but mutations in these pathways are very rare (Cancer Genome Atlas Research Network. Electronic address & Cancer Genome Atlas Research, 2017; Kovac et al., 2015). Furthermore, WGS of TPCs and non-TPCs, and in functionally distinct clones of sarcoma revealed little differences, strongly implicating a role of the epigenome in regulating intratumoral heterogeneity. The recent identification of mutations affecting epigenetic regulators driving tumorigenesis provides direct evidence for the importance of epigenetic dysregulation in generating the TPC phenotype. In leukemia, chromosomal rearrangement in the KMT2A/MLL gene induces TPC formation in both hematopoietic stem cells and downstream progenitor cells. The KMT2A/MLL gene encodes a histone methyltransferase that regulates enhancer accessibility, and when mutated can directly confer de novo self-renewal capacity to differentiated progenitor cells. Furthermore, in glioblastoma, mutations in the H3F3A gene inhibits the Polycomb Repressive Complex 2 (PRC2) and leads to the re-establishment of more immature developmental programs in neural precursor cells, and acquisition of more stem-cell like features. Dysregulation of epigenetic regulators independent of genetic mutations contributes to
TPC phenotype as well. In multiple cancer types, the linker histone H1.0 level is highly variable between cells of the same tumor. Low H1.0 cells demonstrated a chromatin landscape that promotes self-renewal. Forced expression of H1.0 significantly reduced self-renewal and promote differentiation.

Alterations in the epigenome are emerging as an important driver of metastasis. The lack of genetic differences between MC and non-MC is consistent with the notion of epigenetic heterogeneity, resulting in differences in metastatic ability. In pancreatic cancer, heterochromatin loss tends to occur in tumor cells at distant sites rather than in primary tumors. This loss of heterochromatin domain is associated with increased expression of genes that promote malignancy, such as regulators of epithelial to mesenchymal transition. Importantly, a subpopulation of tumor cells with reduced heterochromatin can be identified in the primary tumor, suggesting that clones with these epigenetic changes are evolutionarily selected for during metastasis. Restoration of heterochromatin in the metastatic cells via 6-phosphogluconate dehydrogenase inhibition was able to inhibit metastasis. A recent study profiling enhancer elements in isogenic metastatic and non-metastatic cells of osteosarcoma demonstrated clusters of enhancers coordinated waves of gene expression to promote metastasis. These enhancer clusters induce the expression of multiple known genes that contribute metastasis such as members of the AP-1 complex. Specific disruptions of these enhancer elements can reduce the metastatic burden.

To understand the mechanism that drives the TPC state and clonal dynamics in sarcoma, large scale profiling of the epigenome is likely required. Correlating gene expression patterns with ChIP-seq may reveal the mechanism that orchestrates cell state switching and clonal behavior. ChIP-seq of canonical histone markers that promote gene expression for SP, NSP and NSP derived SP cells, combined with RNA-seq, can map the network of genes and enhancers essential to de
Novo switching of cell states. Similarly, a strategy using ChIP-seq of repressive histone markers may identify the epigenetic mechanism that drives the repression metastatic suppressor genes. Once these genes and their corresponding histone modifications are identified, the upstream histone modifiers that modulate metastasis and the signaling pathways the metastasis suppressor genes participate can be elucidated.

5.2.2 Identification of Druggable Targets in Sarcoma Metastasis

Distant metastasis accounts for the majority of cancer-related mortality. Previous studies to understand the mechanism of metastasis and to identify druggable targets has primarily relied on comparisons between bulk metastatic lesions and primary tumors. This approach is limited to studying late stage metastases, in which the cancer cells have established overt colonies in the distant microenvironment. As such, the genes and mechanisms identified using this approach likely only capture the cumulative result of long-term adaptation of the metastatic cells. By tracing tumor clones throughout the metastasis process, I was able to isolate the clones responsible for metastasis in the primary tumor and study their gene expression. As proof of principle, I was able to identify multiple genes that are suppressed in the primary tumor that contribute to metastasis before overt metastatic lesions are established. These genes represent the early molecular events that enabled certain clones to outcompete others during metastasis.

Targeting the early metastatic clones have several advantages. First, my results showed that there are relatively fewer differences in gene expression between MC and non-MC, compared to between non-MC and established metastases. This suggests that there may be fewer genes and pathways that are essential to their enhanced metastatic ability. Second, tumor cells continue to evolve throughout the metastatic process, and the established lesions may harbor additional genetic
or epigenetic heterogeneity, enabling the established lesions to bypass targeted therapy more easily. Therefore, targeting metastatic cells early in their evolution may be more effective by delaying or preventing the formation of overt metastasis. Third, a major challenge in targeting metastasis is due to having multiple lesions throughout the distant organ, which complicates drug delivery due to potential site-specific differences in drug retention (Ganapathy, Moghe, & Roth, 2015). Furthermore, drug delivery to metastatic sites such as the brain can be particularly difficult due to the blood-brain-barrier. In contrast, MCs are largely concentrated in the primary tumor site and can ease drug delivery.

In the KPCC model, the MCs can be readily isolated by flow cytometry. To identify druggable targets against MCs, a high throughput multiplexed small molecule screening platform can be applied (Gruner et al., 2016). The MCs cells can be labeled with DNA based barcodes and cells with different barcodes can be matched to receive a different drug, and then transplanted into animals. Deep sequencing of the barcoded cells after drug treatment can be used to identify the most effective drug for eliminating MCs. Another approach is to identify functional genes in the MCs that govern their metastatic ability and survival and use targeted clone to specifically inhibit the action of these genes. Therapies targeting the most metastatic clone in cancer may provide an effective method of treating metastasis early during cancer progression.

5.2.3 The Role of the Tumor Microenvironment in Intratumor Heterogeneity

The tumor microenvironment (TME) is the product of continuous cross-talk between the tumor cells and their surrounding normal cells. As tumors grow, they are constantly reshaping and being reshaped by the neighboring cells, and alterations in the microenvironment frequently result in alterations of tumor cell phenotype. Therefore, heterogeneity in the microenvironment may
play important roles in tumor cell heterogeneity. Specifically, TPCs are believed to reside in specific niches. For example, cancer-associated fibroblasts (CAFs) can stimulate TPC properties via activation of Wnt and Notch signaling pathways (He et al., 2004; Vermeulen et al., 2010). Wnt ligand secreted by CAF is sufficient to induce stemness in the colon and other cancer cells. In addition to CAFs, the TME is characterized by infiltration of inflammatory cells such as tumor-associated macrophages (TAM) and CD4+ T cells (Plaks et al., 2015). These cells secrete TNFα, which upregulates NF-kB signaling pathways to induce Snail homolog 2 (Slug), Snail homolog 1 (Snail), and Twist family basic helix-loop-helix transcription factor (Twist) expression in cancer cells (X. Liu & Fan, 2015). Slug, Snail and Twist are critical transcription factors that induce epithelial-to-mesenchymal transition (EMT), which can confer self-renewal and stem-cell-like properties in cancer cells (X. Liu & Fan, 2015; Mani et al., 2008). Mesenchymal stromal cells are also known to stimulate TPC phenotypes through secretion of chemokines IL-6 and IL-8 and BMP antagonists Ghrelin-1 (Cabarcas, Mathews, & Farrar, 2011; Davis et al., 2015).

Cell-independent microenvironment factors can influence cell phenotypes as well. Hypoxia promotes TPC survival and EMT through reactive oxygen species (ROS)-activated stress response pathway and induction of TGF-β and TNF-α signaling pathways (L. Liu, Wise, Diehl, & Simon, 2008; Pavlides et al., 2010). Moreover, hypoxia-inducible factors can directly stimulate NOTCH signaling, a key self-renewal pathway (Marignol, Rivera-Figueroa, Lynch, & Hollywood, 2013). Extracellular matrix (ECM), which is the large collection of distinct proteins, proteoglycans, and polysaccharides in the tumor microenvironment are known to interact with TPCs (Plaks et al., 2015). Mechanical cues from the ECM has been shown to activate YAP/TAZ, transcriptional co-activators of the Hippo pathway (Basu-Roy et al., 2015; Dupont et al., 2011).
YAP expression has been shown to mark stem-like populations in cancers and maintains self-renewal properties through Sox2 (Basu-Roy et al., 2015).

The ability of cell-dependent and cell-independent cues from TME to influence cancer cell phenotype implicate these cues in TPC state dynamics. Systematic investigation of the tumor microenvironment can be performed by single-cell RNA-sequencing (Sc-RNAseq) of non-tumor cells. In the KPCC model, non-tumor cells can be isolated by FACS for cells that do express fluorescent reporters. Functional dissection of various TME cell type or factors derived from these cells can be tested using dual-recombinase animals that enable targeted manipulation of gene expression outside of the tumor (Moding et al., 2014).

5.2.4 Tracing Clonal Heterogeneity at Single Cell Resolution

In this thesis, lineage development of different tumor clones and cell states are traced by labeling distinct tumor cells at the initiation of tumorigenesis. However, tumor clones continue to evolve throughout cancer progression, giving rise to progenies that will likely possess different genotypes and phenotypes. This continued clonal evolution could only be captured by progressively labeling tumor cells in vivo. Recently, advances in genome editing technologies have enabled in vivo generation of complex barcodes over a sustained period (Gulbahce et al., 2017; Kalhor, Mali, & Church, 2017; McKenna et al., 2016). This technology utilizes a synthesized locus of a contiguous array of multiple CRISPR-Cas9 targets for mutagenesis such that a diverse set of outcomes is generated and can be used as unique barcodes (McKenna et al., 2016). The differences in the efficiency of the CRISPR-Cas9 targets at this locus allows different barcodes to be generated over time. Proof-of-principle studies in zebrafish and mouse have demonstrated this in vivo barcoding can accurately construct the lineage relationship from tissues
to whole-organism at single-cell resolution. In zebrafish, the “genome editing of synthetic target arrays for lineage tracing” (GESTALT) was able to simultaneously analyze barcodes thousands of individual zebrafish cells to construct a lineage tree from fertilization of the egg to a developed larva (McKenna et al., 2016). This method has been adapted to the mouse, referred to as the Mouse for Actively Recording cells (MARC1), which delineated the development of whole embryonal and adult tissues (Kalhor et al., 2018).

To trace clonal evolution at single-cell resolution in tumors, future experiments can utilize the progressive in vivo barcoding strategy. For example, the KP-Cas9 sarcoma mouse can be crossed with the MARC1 mouse. Injection of Ad-Cre will activate the oncogenic mutations and barcoding by inducing the expression of Cas9. Progressive barcoding of clones and their progenies may reveal a map of continuous evolution. Another possibility is to leverage dual recombinase technology and trace specific cell lineages after the tumor has been established, such as markers of TPCs, to investigate the lineage trajectory TPC progenies over time.

5.3 Concluding Remarks

The thesis described here is the first systematic investigation of intratumoral heterogeneity in sarcoma. Using primary human samples and transgenic mouse models, the two prevailing models of tumor cell heterogeneity, namely the CSC model and the clonal evolution model are studied. I set out to understand the cellular dynamics that underlie functionally different tumor cell phenotypes. I hypothesized that sarcoma progression is driven by dynamic or transient changes between TPCs and Non-TPCs at fixed points in tumor progression, and by dynamic changes between different tumor clones during different stages of disease progression. I present evidence that 1) sarcoma TPCs are heterogeneous, consist of distinct but overlapping populations;
2) TPC is a dynamic or transient cell state, and targeted elimination of TPCs can lead to compensation from non-TPC cells; 3) over the course of sarcoma progression, tumor clones undergo dynamic changes from polyclonal early stage tumors to monoclonal late stage metastasis. These data provided new insights to the functional roles of the heterogeneous sarcoma cells, and possible routes to target these cells. My work also opens up many new questions with regard to the in-depth genetic/epigenetic mechanisms that drive cellular dynamics, and the evolutionary trajectory of clones at single cell resolution. The models that I have built can be used to explore these new questions with the aid of increasingly sophisticated multiplexed cellular and molecular platforms. However, it is important to note that most of this work is built on either transplants of patient derived sarcoma in immunocompromised mice or GEMM models. These model systems may not fully capture every aspects of human tumor biology such as its genomic complexity and \textit{de novo} tumor microenvironment. Overall, the work described here adds to the growing understanding that understanding intratumoral heterogeneity may hold valuable opportunities toward building a complete roadmap of tumorigenesis and that predicting and targeting the dynamic evolution of cancer cells may improve therapy.
REFERENCES


152


by stereotactic body radiation therapy. *Sci Transl Med*, 7(278), 278ra234. doi:10.1126/scitranslmed.aaa4214


162


167


