Interrogation of Changes in Cell State during Tumor Evolution of a Genetically Engineered Mouse Model of Lung Adenocarcinoma

by

Amanda Margarita Cruz

B.S. Cell Biology
University of California, Davis, 2016

Submitted to the Department of Biology
In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September 2021

© 2021 Massachusetts Institute of Technology. All Rights Reserved

Signature of Author ……………………………………………………………………………..
Department of Biology
Sept 7, 2021

Certified by ………………………………………………………………………………………..
Tyler Jacks
David H. Koch Professor of Biology
Daniel K. Ludwig Scholar for Cancer Research

Accepted by ……………………………………………………………………………………..
Amy Keating
Professor of Biology and Biological Engineering
Co-Director, Biology Graduate Committee

Interrogation of Changes in Cell State during Tumor Evolution of a Genetically Engineered Mouse Model of Lung Adenocarcinoma

by

Amanda Margarita Cruz

Submitted to the Department of Biology on September 7, 2021 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biology

ABSTRACT

In genetically engineered mouse models of lung adenocarcinoma (LUAD), tumors become more heterogeneous and dysregulate cell identities as they progress and evolve. In this thesis, single-cell RNA-sequencing technology was utilized to understand dynamic changes that occur during tumor evolution both with respect to tumor cells and tumor-specific cytotoxic CD8 T cells. In tumor cells, expression of *Etv4* and *Etv5*, which belong to the Pea3 family of transcription factors, vary as a consequence of tumor progression. *Etv5* regulates the identity of the cells that give rise to KP tumors, and its expression is lost as tumors evolve. Conversely, *Etv4* is not expressed in the adult lung, but becomes latently expressed in aggressive tumors. Interestingly, we find that both *Etv4* and *Etv5* are required for lung tumor initiation. In addition, we also profile CD8 T cells that specifically recognize experimentally defined tumor neoantigens and provide evidence for an antigen dominance hierarchy that creates competition between T cell responses to tumor neoantigens. Critically, we find that this hierarchy influences the functionality of CD8 T cells and describe novel differentiation trajectories that distinguish subdominant and dominant antigen

responses. Together, findings from these studies were used to propose analytical methodologies to model tumor evolution.

CV

Amanda Margarita Cruz

amcruz@mit.edu (530) 574 7681
linkedin.com/in/amacruz/

Educational Background

University of California, Davis (2012 – 2016) with Honors

Grade Point Average: 3.8

B.S. in **Cell Biology**

Massachusetts Institute of Technology (2016 – present)

Department of Biology, PhD Candidate

Thesis Advisor: **Tyler Jacks**

Coursework:

Principles of Biochemical Analysis Methods & Logic: Molecular Biology Systems Biology Graduate Genetics	Advanced Cancer Biology Case Studies in Expt. Design Molecular Biology

Relevant Work Experience

- PhD Intern, Molecular Biology Team and Computational Biology and Bioinformatics Team, *Skyhawk Therapeutics, Platform Team (Sept 2020 – Present)*

Awards and Honors

- Dean's List Honors, *College of Biological Sciences (Winter 2015, Fall 2014, Spring 2014, Winter 2013, Fall 2013, Spring 2013)*
- Outstanding Sophomore *(Biology Undergraduate Scholars Program, 2014)*
- Outstanding Poster Presentation, *Cancer Biology (Annual Biomedical Research Conference for Minority Students, 2014)*
- Provost's Undergraduate Fellowship *(University of California, Davis, 2014)*
- Outstanding Junior *(Biology Undergraduate Scholars Program, 2015)*

- Garfield Fellowship *(Undergraduate Research Program, Cold Spring Harbor Laboratory, 2015)*
- Outstanding Senior Advisor *(Biology Undergraduate Scholars Program Peer Advising, 2016)*
- Departmental Citation Award *(Department of Molecular and Cellular Biology, College of Biological Sciences, University of California, Davis, 2016)*
- Phi Sigma *(Biological Sciences Honor Society, University of California, Davis, 2016)*
- Honorable Mention *(Ford Foundation Predoctoral Fellowship, June 2018)*

Program Affiliations and Organizations

- Biology Undergraduate Scholars Program *(UC Davis, 2012-2014)*
- Biology Undergraduate Scholars Program Summer Honors - IMSD *(UC Davis, 2014)*
- Biology Undergraduate Scholars Program Honors – IMSD *(UC Davis, 2014-2015)*
- Biology Scholars Advanced Research Program – MARC-USTAR *(UC Davis, 2015-2016)*
- Undergraduate Research Program *(Cold Spring Harbor Laboratory, 2015)*

Academic Research Experience

- Research Assistant, *Sweeney Lab, UC Davis Medical School (2014 - 2014)*
- Undergraduate Research Fellow, *Sweeney Lab, UC Davis School of Medicine (2014 – 2016)*
- Summer Undergraduate Research Fellow, *Mills Lab, Cold Spring Harbor Laboratory (2015)*
- Graduate Researcher, *Jacks Lab, Koch Institute for Cancer Research at MIT (2017 – Present)*

Academic Research Projects & Conference Presentations

“Loss of Lrig1 Expression as a Mechanism for Tamoxifen Resistance in Breast Cancer”, Sweeney Lab

Biology Undergraduate Scholars Program Summer Symposium (Aug 2014), Chancellor’s Fall Convocation Poster Session (Oct 2014), Annual Biomedical Research Conference for Minority Students (Nov 2014)

“Lrig1 Regulation of Canonical Wnt signaling in Triple Negative Breast Cancer”, Sweeney Lab
Undergraduate Research Conference at the University of California, Davis (May 2015)

“Lrig3 Regulation of Canonical Wnt signaling in Triple Negative Breast Cancer through interaction with the Wnt co-receptor Lrp6”, Sweeney Lab
Undergraduate Research Conference at the University of California, Davis (2016)

“Loss of Chd5 Leads to Aberrant Differentiation of Neural Stem Cells in *mus musculus”*, Mills Lab
Undergraduate Research Program Final Symposium (2015), Biology Undergraduate Scholars Program Summer Symposium (2015), Annual Biomedical Research Conference for Minority Students (2015), Molecular and Cellular Biology T32 Retreat Poster Session (2015)

“Transcriptional and Epigenetic Mechanisms of Tumor Evolution in Genetically Engineered Mouse Models of Lung Adenocarcinoma”, Jacks Lab

“The Roles of ETS Transcription Factors in Disease Progression of Genetically Engineered Mouse Models of Lung Adenocarcinoma”, Jacks Lab

“Antigen Dominance Hierarchies Shape TCF1+ Progenitor CD8 T Cell Phenotypes in Tumors”, Jacks Lab

Teaching Experience

- Chemistry Tutor*, Student Academic Success Center at UC Davis (Aug 2014 – June 2015)*
- Chemistry Tutor, *STEP, Education Enrichment & Outreach Program at UC Davis (Summer 2014)*
- Graduate Teaching Assistant*, Introduction to Biology (7.013) at MIT (Spring 2018)*
- Graduate Teaching Assistant, *Hallmarks of Cancer (7.45/7.85) at MIT (Fall 2019)*

Diversity Outreach, Advocacy & Volunteer Work

- Volunteer, *Medical Intensive Care Unit, UC Davis Health System (2013)*
- Peer Advisor, *Biology Undergraduate Scholars Program (2014-2015)*
- Volunteer, *Yolo County Hospice (Fall 2015)*
- Lead Peer Advisor, *Biology Undergraduate Scholars Program (2015-2016)*
- Graduate Diversity Ambassador, *Office of Graduate Education at MIT (2017 – 2020)*
- Diversity Co-Chair, *Biology Graduate Student Council at MIT (2017 – 2018)*
- Department of Biology Conduit Representative, *Diversity Conduit Initiative, Graduate Student Council at MIT (2017 – 2020)*
- Statement of Purpose Advisor, *CONVERGE, Diversity in Graduate Education Initiative at MIT (Fall 2017)*
- Statement of Purpose Advisor, *Diversity in Graduate Education Initiative at MIT (Fall 2019)*
- Application Reviewer**,** *MIT Summer Research Program (MSRP) (2017, 2019)*
- Member**,** *Graduate Students for a Healthy MIT (G4HMIT) (2019 – 2020)*
- Co-Founder, *Biology Diversity Community at MIT (2018 – 2020)*

(continued)

Skills

Benchwork:
Tissue Culture (2D Immortalized, Primary Cells, and Organoids), Cell Migration Assays, Transient Transfections, Lentiviral & Retroviral production and transductions, Plasmid extractions, Nucleic acid isolation of tissue samples, Western Blotting, Transformations, Co-Immunoprecipitations, Colorimetric Assays, Gateway/Gibson/Restriction Enzyme Cloning, shRNA/siRNA knockdown, Library Prep, CRISPR/Cas9 Knockouts, PCR, qRT-PCR, Immunohistochemistry, HTS-based assays
Animal Work:
Mouse Husbandry and Colony Maintenance, Genotyping, Mouse Genetics, Intratracheal Delivery of virus, IVIS imaging, micro-CT imaging, Intraperitoneal injections, Handling of NSG/Immunocompromised mice, Basic Surgical Techniques
Computational Work:
Python Programming, Basic MATLAB, R/R-Studio, RNA-seq analysis, bulk RNA-seq analysis, single cell ATAC-seq analysis (sciATAC), single-cell RNA sequencing analysis (10X, SmartSeq2), High-dimensional data analysis, Basics of data structures and manipulation (PANDAS and Tidyverse), TCR/VDJ single-cell sequencing analysis, Cluster-based computing, UNIX, Basic utilization of Google Cloud Storage & Computing Infrastructure, Basic utilization of Terra (Firecloud) computing infrastructure, Basics software development operations, Benchling Enterprise.

Publications

Megan L. Burger, **Amanda M. Cruz**, Grace E. Crossland, Giorgio Gaglia, Cecily C. Ritch, Sarah E. Blatt, Arjun Bhutkar, David Canner, Tamina Keinka, Sara Z. Tavana, Alexia L. Barandiaran, Andrea Garmilla, Jason M. Schenkel, Michelle Hillman, Izumi de los Rios Kobara, Amy Li, Alex M. Jaeger, William L. Hwang, Peter M. K. Westcott, Michael J. Manos, Marta M. Holovatsk, F. Stephen Hodi, Aviv Regev, Sandro Santagata, Tyler Jacks. **Antigen Dominance Hierarchies Shape TCF1+ Progenitor CD8 T Cell Phenotypes in Tumors**. (Manuscript Accepted July 2021, *Cell*)
http://dx.doi.org/10.2139/ssrn.3770111

Peter M K Westcott, Francesc M Remolar, Olivia Smith, Haley Hauck, Nathan J Sacks, Zackery A Ely, Alex M Jaeger, William M Rideout, Arjun Bhutkar, Daniel Zhang, Mary C Beytagh, Roderick T Bronson, David A Canner, Santiago Naranjo, Abbey Jin, JJ Patten, **Amanda M Cruz**, Isidro Cortes-Ciriano, Tyler Jacks **Mismatch repair deficiency is not sufficient to increase tumor immunogenicity.**
https://www.biorxiv.org/content/10.1101/2021.08.24.457572v1

References

Colleen Sweeney, PhD	Connie Champagne, PhD	Alea Mills, PhD	Tyler Jacks, PhD
Department of Biochemistry and Molecular Medicine University of California, Davis School of Medicine Sacramento, CA 95817 (916) 734 0726 casweeney@ucdavis.edu	Biology Undergraduate Scholars Program MARC-USTAR IMSD University of California, Davis (530) 754 8727 cechampagne@ucdavis.edu	Professor Cold Spring Harbor Laboratory Cold Spring Harbor, New York mills@cshl.edu	Professor, Department of Biology Director, Koch Institute for Integrative Cancer Research at MIT Cambridge, MA 02139 tjacks@mit.edu

ACKNOWLEDGMENTS

Tyler - I have always told you that Tumor Evolution was my dream project. I will always be grateful for your encouragement to explore new fields and approach questions in tumor biology in sometimes unconventional ways. You were able to provide me the rare opportunity to focus solely on the research that was most interesting to me, rather than conforming to convention, and without distraction. You have also been an outstanding advisor to me and have supported me through difficult and complex situations, even when it was not easy to do so. Thank you for being a genuinely kind and caring person, and for investing so deeply in my professional success.

To my thesis committee:
Eliezer - Thank you for being an invaluable resource of scientific and graduate school advice for success in graduate school. I could always trust that the advice that you ever gave me had my best interests in mind, and every interaction that I have had with you has been extraordinarily helpful.
Mike Yaffe - You were the first faculty member that I interacted with who not only encouraged me to learn entirely new skills in different fields, but also rigorously question every assumption and methodology incorporated into my experimental design. I am so grateful to have had the opportunity to think about tumors and cell signaling with you. Thank you for believing in me.
Cole Trapnell - Thank you for being willing to serve as my external thesis committee member, which is not something I could have imagined, even in my wildest dreams.

I am extremely grateful for the efforts of many members of the faculty in the Department of Biology, notably: **Jackie Lees, Steve Bell, Salil Garg, Amy Keating, Angelika Amon, and Matt Vander Heiden** for encouragement, support and for showing kindness to me at critical points during graduate career.
Graham Walker - Thank you for being incredibly kind and encouraging after unexpectedly being asked to step in to administer my qualifying exam. Your advice led me to think more critically about the molecular properties of the genes I studied.
AJ Bhutkar & Sarah Blatt- Thank you both for giving me technical (and statistical) advice while learning the ropes of computational biology. I had a blast collaborating with both of you and have nothing but the utmost respect for both of you.
Judy - Thank you for all of the support you have given me over the years. You always went above and beyond to help me, both professionally and personally, when I needed it most. I would not have made it through this journey without you and I am eternally grateful for you.
Kim - The mouse queen of the Jacks lab! Thank you for encouraging me to approach my animal work with compassion and kindness. You also knew exactly what to say to

help me at my lowest points and give me encouragement when I needed it most. It always meant so much coming from you. You are extraordinary.
Margaret, you were a fantastic baymate! If it was not for you, I (and the whole lab) would fall apart. Your commitment to supporting this lab is unparalleled and I was always so grateful that you were there alongside me through many late nights.
Carmen Villatoro - There are not enough words to thank you for the kindness you showed me. There were countless moments in graduate school that you were my biggest cheerleader. Thank you for being there for me and for helping me through many difficult situations.
Luz - Thank you for always being a friendly and warm person to be around, even on the night shift! You are so hardworking and I will miss being able to chat with you all the time.
Karen and Kate - Thank you both for being so hardworking and keeping the lab functional, particularly during the COVID-19 pandemic!
Britt - Thank you for being a valuable resource for just about everything. I got so lucky with my baymates. Also, thanks for not laughing at my stupid mistakes (even when I really deserved it). I am extremely grateful that the vast majority of my late nights in the lab were not spent alone because you were there with me! Long live the night owl bay!
Sheng Rong and Peter, thank you both for getting me started in the lab and helping me through all of my stupid mistakes. I cannot wait to see all of the discoveries that you will make. Thanks for always going out of your way to support me, celebrate my successes, and help me through my failures - particularly when I didn’t make it very easy to do that.
Toni-Ann Nelson and Michelle Yin, thanks for being super awesome students to mentor. You are both so brilliant and capable. I cannot wait to see how your careers progress in the future!
Lindsay - Thank you for convincing me that my scientific ideas were worth investing resources into. There were countless moments where it felt like you were the only person that understood my scientific perspectives. I am so grateful that I got to collaborate and think about tumor evolution with you. You are a creative and thoughtful scientist and will be an amazing investigator.
Issy DelPriore, Olivia Smith, Katherine Wu, Caterina Colon, Demi Sandel, Ana Paola Garcia, Alexia Barandiaran, Grace Crossland, Michelle Hillman, Sophie Dielek and all of the techs of the Jacks lab - You are all some of the most hardworking, brilliant, and inspired scientists that I have ever worked with. I have so much respect for and have learned so much from each one of you.
Will Hwang - there are so many things that I could thank you for, but foremost, thanks for being a great friend. You are equally as kind as you are brilliant.

Nemo - you were one of the first people I interacted with extensively in the Jacks lab and are a big reason why I joined in the first place. Thank you for getting excited about tumors with me!
Laurens - Thanks for looking out for me, pushing me to improve when needed, and just being an overall great person to share a work environment with!
Grace Phelps - It was truly the best being just down the hall from you - thanks for being such a great person to work with and supportive friend.
Helen - thank you for your advice before and throughout grad school at MIT. If it was not for you, I would not be here!
Colin, Swanny, Sophia and the rest of the Lees lab - thanks for being like a second lab family to me!
The next generation of the Jacks lab: Sean-Luc Shanahan, Emma Dawson, and Nick Mathey-Andrews - thank you for helping me rediscover all of the reasons why I got excited about tumor evolution in the first place. I think you are all so smart and each bring so much to the table; I cannot WAIT to see how you all approach your projects in the future and bring your own perspectives to our models of tumor evolution.
Christina Cabana - I could write an entire acknowledgements page just for you, but I am so grateful that you have been here to support and think about tumor evolution with me. I think that you are going to change the world, and you will have my support indefinitely. It will be so exciting to see where you lead tumor evolution in the future.
Amy Li - you were such an amazing role model for me, in research, general academics, and personally during graduate school. I am so grateful for the overlap we had in the lab. I can't wait to see all of the things you will accomplish in your career.
Grissel - an amazing companion in the Jacks lab, through thick and thin. Thanks for always having my back and helping me navigate some of the hardest challenges in grad school!
David Canner, Ryan Kohn, Banu Eskokiak, Chengcheng Jin, Leanne Li, Carla Concepcion, Alex Jaeger, Jason Schenkel, Marianna Trakala, the rest of the Jacks Lab Grad students & postdocs, Dian Yang, Bashar Hamza, and Tobiona Oni - Thank you for all of the discussions, career advice, and interactions we had in (and out) of the lab!
Megan Burger - Thank you for being an inspired and enthusiastic collaborator. I had the best time thinking about and talking about science with you. You have pushed me to grow both in terms of learning immunology but also as a data analyst, thank you for giving me the opportunity to work with you. Additionally, thank you for the invaluable help you provided on other projects that we did not formally collaborate on - the fact that you were so willing to help me without any expectation of benefit to yourself speaks to your collaborative and inclusive approach to science. I am so excited for what the future holds for you.

Allegra Terhorst, Kim Nguyen, Ellen Duong, Arish Shah and my BioGrad cohort - thanks for all of the support, and for helping me (try) to have a work-life balance. I wouldn't have made it without you all.

To my systems biology friends: **Byron Lee, Constantin Weisser, Sameer Abraham, Chantal Geugler, Thy Pham, and Champ Taweewat Somboonpanyakul** - thank you all, for helping me learn physics, math, statistics, programming, and systems biology. You all made what was one of the hardest experiences in graduate school also one of the most fun. You've been awesome friends.

Bobby Baraldi and Ethan Baker - Thank you both for being such incredible and supportive friends, both personally and professionally, to me for many years. I feel so fortunate to have had you both so actively encouraging and supporting my personal and professional growth. You are both amazing humans.

To **Meghann Kasal, Alexandra Navarro, Yamilex Acevedo-Sanchez**, **Swanny Lamboy Rodriguez, Nicole Aponte Santiago, Alicia Zamudio, Gerardo Manuel Perez Goncalves, Sheena Vasquez, Emma Sedivy, Samantha Weber, Talya Levitz,** and **Claudia Variela** - I couldn't have asked for better teammates in diversity, equity, and advocacy efforts. Being able to work alongside you all was inspiring.Thanks for helping me create balance during graduate school and advocate for meaningful change.

To Alea Mills, Anbu, Dongwoo, Ellen and the rest of the Mills Lab - Thank you for encouraging and allowing me to conduct research in a curiosity-based approach. The summer I spent with you all changed my life and I am so incredibly grateful that I had the opportunity to learn from all of you.

To Colleen Sweeney and Kermit Carraway - Thank you for taking a chance on me and providing my first research experiences. I look up to you both in so many ways, both as scientists and as mentors.

Jason Hatakeyama - My accomplishments and successes in research are entirely because you believed in me and mentored me even though you were never obligated or asked to. In addition to training me to think about biology in both a practical and abstract sense which laid the foundations for much of the work discussed in this thesis, you also taught me how to design experiments that accommodate complex biological systems. You are the kind of mentor who goes above and beyond to help your students (helping me even when we were at different institutions). I am eternally grateful to you.

Frank Mercado, Catalina Simion, Jessica Toan, Anastasia Berg, Kacey VanderVorst, Matthew Saldana, Ignat Prinstev, Maria Cedano, Maxine Umeh-Garcia - Thank you all for your mentorship and for training me at the bench. The quality of the training you all provided for me facilitated my work in graduate school and beyond. Thanks for not killing me after I spilled acetic acid in the lab and for every other mistake I made.

Manojit Swamynathan, Shadd Cabalatungan, Antonio Cuevas Navarro - Thanks for all of the (very) late night scientific discussions, company while doing experiments, friendship, and laughter.
Alexis Wrate, Ekaterina Tangog, Tiani Calip, Jessica West, Daniel Curiel & Anastasia Liebler, Priya Garcha, Lily Richie, Veronica Ruiz, and Ayda Sadeghian - Thank you for being amazing friends during graduate school.
Connie Champagne, Heather Lawrence, and Barbara Horwitz - There will never be enough words to thank both of you for the impact you have both had on my career, but foremost, thank you for your substantial emotional, professional, and scientific investments in my career and training. If it was not for you three women, it is very likely I would never have gotten involved in biological research. In particular, thank you for providing endless support during some of the darkest years of my life.
Penny Shelton, Laura Brun, Jon Bryant, Bruce Coleman, Lenore Shimota, and the rest of my childhood teachers - Thank you for your support, kindness, and attentiveness.
Jeanette Natzle, Jawdat Al-Bassam, John Albeck, Bruce Draper, Jonathan Eisen, Enoch Baldwin, Dan Starr - Thank you for providing an amazing foundation of cell biology, which was instrumental throughout graduate school.
Sara Hawkes - Thank you for teaching me math in a way that I could intuitively understand and for laying the foundations for my conceptual understanding of mathematical models.
To my co-workers at Skyhawk Therapeutics: Yi Ying, Anne Deconick, Botao Liu, Boyko Kakaradov, Woody Lin, Zach Dwyer, Zhlaleh Safikhani, Kathleen McCarthy, and the rest of the company - thanks for creating a supportive and inclusive environment that allowed me to develop skills which, although unrelated to my academic research, facilitated much of my graduate work. I am so excited to continue to work with all of you in the future.
In memoriam - Rolf Unterleitner. I will never have the opportunity to tell you how important you were to my success in graduate school, but if you were still here, I would tell you that, because of you, I was able to persevere without taking time off during some of the most difficult situations of my life. You went so far above and beyond to mentor and help countless students like myself. I miss you all of the time and will never forget all that you did for me.
To my biological family - thank you for your undying support, even from across the country, during graduate school and my academic training. I love you all so much.
To my (future) Aoude family: **Liana, Nassim, Mona, Kyle, Fareed & George, George, Elias, and the Wettengels**- Thank you for providing support and balance in my life during some very stressful moments. I am so grateful for all of you.
To my (future) Mother and Father in law, Marie and Farid Aoude - If it was not for you both, I am not sure I would have survived graduate school. Thank you for taking

care of me like one of your own children and for helping me get established in the Boston area, from the very start of graduate school! I am so fortunate that I will be able to call you officially my family later this year.
To my Lolo, Jose B. Cruz - thank you for pushing me to think about modeling and behavior of systems outside of biology. By pushing me to incorporate many of these ideas, you allowed me to scientifically connect with you and my late father in a way that I never thought would be possible. It is a very unique and privileged experience to be able to have high-level and abstract scientific discussions with family. Additionally, thank you for your feedback on many of the concepts introduced in this thesis.

To my mom, Barbara Vinsant - everything that I am, everything in this work, and everything I will ever accomplish is because of you. There is never a way to thank someone who has quite literally dedicated so much of their life to my growth and development. I love you so much.
To my twin Sister, Donna Cruz - The direct current to my alternating current. Thanks for believing in me and supporting me so much during graduate school (and in life). We used to have a saying - “you quit, I quit”. Guess we didn’t need to. Can’t wait to (hopefully) be Dr. Cruz^2.
To my future husband, Patrick Aoude - your unrelenting support has made all of this possible. From rescuing me when debugging code, your patience as I ramble on about tumors for hours, feeding me, and supporting me through extremely late nights, you have loved me unconditionally and wholeheartedly. You have had a profound influence on all of my research and the work that is in this thesis. I love the life that we have built for ourselves and cannot wait to spend the rest of my life with you.

This work is dedicated to and inspired by my late father, Rene Leonardo Cruz. I hope that this work would have made him proud.

TABLE OF CONTENTS

ABSTRACT **2**

CV **4**

ACKNOWLEDGMENTS **9**

TABLE OF CONTENTS **15**

CHAPTER 1, PART 1 23
1. Properties of Tumor Heterogeneity 24
1.1 Classification of mutations that underlie cancer development 24
1.2 Implications of oncogenic mutations 27
1.3 The Dynamic Tumor Microenvironment 31
1.4 Perspectives of Tumor Heterogeneity 33
1.4.1 Mutational Intra-Tumoral Heterogeneity 35
1.4.2 Functional Tumor Heterogeneity 38
1.4.3 Tumor Heterogeneity and Metastasis 42
1.5 Emerging Evidence for the Role of the Epigenome in Tumor Heterogeneity 46
2. Tumor immunology 54
2.1 Inflammation and cancer 54
2.2 Influence of solid tumor biology on immune response 56
2.3 Characterization of cytotoxic T cell dysfunction 57
2.3.1 Differentiation of cytotoxic T cells 57
2.3.2 Hallmarks of T cell dysfunction and exhaustion 61
2.4 A Brief History of Immunotherapy 63
2.5 Immune Checkpoint Blockade 65
2.5.1 Predicting Response to Checkpoint Blockade 66
2.5.2 CD8 T Cell Progenitors are the Target of Immune Checkpoint Blockade Therapy 67
CHAPTER 1, PART 2 68
1. Clinical Characteristics of Non-Small Cell Lung Cancer 68
1.1 Epidemiology 68
1.2 Disease Characteristics and Staging 69
1.3 Treatment 73
2. Lung Development 75
2.1 Physiology 75
2.2 Anatomy 76
2.3 Genetics of development 77

2.4 Lung Development and Non-Small Cell Lung Cancer 80
3. Genetically Engineered Mouse Model of Lung Adenocarcinoma 81
3.1 The KP Model 81
3.1.1 Kras 81
3.1.2 Trp53 82
3.1.1 Behavior of the KP model 83
3.1.2 Cell of origin in the KP model 84
3.2 Developments in experimental tools for studying the KP model 85
CHAPTER 1, PART 3 86
High-Dimensional Biological Data 86
1. Technological Advancements in Sequencing Technology 86
2. Structures of High-Dimensional Biological Data 88
2.1 Advancements in discovery of complex data structures 88
2.2 Relationships between data structures and underlying noise 89
3. Dimensionality Reductions and Interpretability 90
3.1 Modern approaches to dimensionality reduction in biology 90
3.1.1 Linear Dimensionality Reduction 91
3.2.2 Nonlinear Dimensionality Reduction 93
3.2 Limitations of dimensionality reductions 95
4. Clustering and Classification of Cell Types 96
5. Biological Variation 99
5.1 Biological Noise 99
5.2 Biological Metadata 100
5.3 Biological Pseudotime 101

CHAPTER 2 **106**

ABSTRACT **106**

INTRODUCTION **107**

RESULTS **111**

KP Tumor progression occurs reproducibly over time across mice and tumors 111

Figure 1. KP Tumor progression occurs reproducibly over time across mice and tumors. **117**

Figure S1. KP Tumor progression occurs reproducibly over time across mice and tumors. **119**

Biological pseudotime approximates gene expression changes in KP lung tumor progression 120

Figure 2. Biological pseudotime approximates gene expression changes in KP lung tumor progression. **124**

Sequential changes in cell identity occur throughout KP lung tumor cell evolution and progression. 125

Figure 3. Sequential changes in cell identity occur throughout KP lung tumor cell evolution and progression. **128**

Supplemental Figure 3. Expression of epithelial markers associated with metastasis in non-small cell lung cancer. **129**

Evaluation of Pea3 transcription factors in KP lung tumor progression. 130

Figure 4. Evaluation of Pea3 transcription factors in KP lung tumor progression. **133**

Etv5 is required for KP lung tumor initiation in vivo but not in vitro. 134

Figure 5. Etv5 is required for KP lung tumor initiation in vivo but not in vitro. **138**

The ETS domain of Etv4 is likely required for KP lung tumor initiation in vivo. 140

Figure 6. The ETS domain of Etv4 is required for KP lung tumor initiation in vivo. **142**

DISCUSSION **143**

MATERIALS AND METHODS **153**

scRNA-seq analysis 153

Figures and Visualizations 154

Lentiviral Production 154

Lentiviral Titering 155

Lentiviral Vectors and sgRNA Cloning 155

CRISPR Guide & Targeting Validation 155

sgRNA sequences 156

PCR primers for TIDE analysis 157

Animal Studies 157

Antibodies Used 158

Histology 158

ACKNOWLEDGEMENTS **158**

AUTHOR CONTRIBUTIONS **158**

CHAPTER 3 **160**

ABSTRACT **161**

INTRODUCTION **161**

RESULTS **165**

Temporal dynamics of CD8 neoantigen response in Lung Adenocarcinoma 165

Figure 1. Temporal dynamics of CD8 neoantigen response in Lung Adenocarcinoma. 168

SIY-Specific CD8 T Cells are Enriched for a TCF1+ Progenitor Phenotype 170

Figure 2. SIY-Specific CD8 T Cells are Enriched for a TCF1+ Progenitor Phenotype 174

Analysis of TCR Clonotypes from SIIN versus SIY scRNA-seq Data Delineates Relationships Between CD8 T Cell States 176

Figure 3. Analysis of TCR Clonotypes from SIIN versus SIY scRNA-seq Data Delineates Relationships Between CD8 T Cell States 180

Supplemental Figure 3. SIY-specific CD8 T cells are enriched for TCF1+ cells and are less differentiated than SIIN-specific CD8 T Cells. 181

The Subdominant CD8 T Cell Response is Enriched for a CCR6+ TCF1+ Progenitor Cell Subset with a Tc17 Differentiation Trajectory 182

Supplemental Figure 4. Tc17-like TCF1+ Progenitor CD8 T Cells in KP LucOS Tumors 185

DISCUSSION **186**

MATERIALS AND METHODS **192**

Lentiviral Tumor Induction 192

Tissue Collection and Flow Cytometry 193

scRNA-seq Data Processing 195

Cell Clustering, and Differential Expression Analysis 196

scRNA seq T-Cell Subtype Classification 198

Gene Signature Scores 199

TCR Clonotype Analysis 199

ACKNOWLEDGEMENTS **201**

AUTHOR CONTRIBUTIONS **202**

CHAPTER 4 203

SPECULATIONS AND FUTURE DISCUSSIONS **203**

Models for Evolution of Tumor Heterogeneity 203

Tumor evolution as a chaotic process 206

Tumors may evolve as a result of sensitivity to tumor-initiating signals 206

Tumor evolution is deterministic and bounded 207

Biological pseudotime is a rescaled unit of time 209

Comparisons to models of mutational tumor cell evolution 213

Transcriptional Bursting May Underlie Heterogeneity of Cell State 215
Relationships between Cell Signaling and Cell Identity **217**
Signaling Architecture 217
Refining models of the functional differentiation of tumor specific CD8 T Cells **222**
Conclusions 224
REFERENCES **229**

CHAPTER 1

INTRODUCTION

Discoveries about developmental and homeostatic biological processes have been largely connected to advancements in cancer biology. The opportunity to compare healthy and cancerous tissues provides an ideal experimental system to identify critical regulatory processes that prevent or create oncogenic stress. However, identifying the changes that drive transitions between cell states associated with early stages of disease progression to those of late stages have been more difficult. Identifying these changes requires a comparison amongst a continuum of heterogeneous cell states that arise over an extended period of time. Additionally, differences in how cells of different tissue types respond to intra- or extracellular changes further highlight the importance of cell state on disease progression.

Dysregulation of cell identity and differentiation status has long been implicated in cancer with respect to tumor cells, and manifests clearly in tumor histology. More recently, processes that lead to differentiation and activation of infiltrating immune cells have also been implicated in cancer progression. Still, it remains unclear how changes of cell state occur and contribute to overall disease progression. Developmental and homeostatic differentiation processes dictate the identity of a cell and can have a large influence over how it responds to a change or stimulus. Cell state, which is defined by the transcriptome, epigenome, proteome, and/or behavior of a cell, is a product of global permutations of gene expression related to cell identity. Genes, transcription factors, and signaling machinery are typically characterized in experimental systems that isolate each component as distinct and separable entities. Large datasets produced by improvements in biological profiling technologies have revealed intricate, combinatorial, and context-specific regulation of signaling pathways, chromatin

topology, and gene regulatory networks. Together, they orchestrate higher order biological systems that control both the behavior and inputs of these systems to generate a specific biological outcome. As a consequence, it is difficult to predict overt biological behavior of a given cell state from its transcriptional profile alone.

Consequently, the objectives of my thesis research have been to 1) integrate findings in developmental biology or non-malignant tissues to better understand the genetics and context-specific nature of cell biology in cancer, and 2) determine how these findings affect the hallmarks of cancer which characterize the biology and progression of the disease. In particular, I have sought to understand how transcriptional profiles associated with differentiation status of both tumor cells and immune cells of the tumor microenvironment reflect dysregulated functionality over time or in response to differentiation signals. Moreover, I have further compared heterogeneous subpopulations of cells that result from differentiation or loss of cell identity in the tumor microenvironment with cell states that arise in other biological contexts, such as embryonic development or chronic viral infection. These findings have led me to identify transcription factors that may be implicated in regulation of cell states that ultimately influence and orchestrate the development of tumor heterogeneity.

To begin, I will give a historical context of fundamental discoveries made about cancer biology and tumor immunology, which have largely occurred by challenging assumptions made about the underlying nature of these diseases. I will then describe how these discoveries have influenced experimental models and approaches of research in tumor biology and immunology. In addition, I will give a background of the behavior and characteristics of lung adenocarcinoma, the disease I have chosen to

study. Finally, I will describe approaches taken to elucidate biologically informative interpretations of high-dimensional data structures produced from high-throughput sequencing technologies.

CHAPTER 1, PART 1

A History of Tumor Biology and Tumor Heterogeneity

The primary cause of cancer related mortality is complications that arise due to metastatic disease (Gupta and Massagué 2006). Unfortunately, studies that aim to understand tumors in the context of longitudinal disease progression are limited by availability of patient samples. The majority of longitudinal patient biopsy samples are taken before and after treatment (Rye et al. 2018; Chicard et al. 2018; Hata et al. 2016). For patients who do not elect to undergo treatment, longitudinal patient biopsies are rarely acquired due to lack of individual clinical benefit. As a consequence, our understanding of tumor progression in the absence of therapeutic intervention in many disease contexts is limited to experimental models of cancer and their associated technical limitations.

Understanding cancer progression can provide opportunities to develop rationally designed and biologically informed therapies. In most cases, patients with cancers that are less progressed have better associated clinical outcomes (Knudsen et al. 2016; de Koning et al. 2014; Humphrey et al. 2002). By identifying the mechanisms through which cancer progression occurs, tumor behavior can be better predicted and can preemptively inform cancer treatments.

1. Properties of Tumor Heterogeneity

1.1 Classification of mutations that underlie cancer development

Some of the first recorded descriptions of cancer were written circa 1500 BC (Breasted 1930). Many generations later, cancer was more comprehensively described and categorized by Henri de Mondeville (Fell 1857). By that time, and in the many years that followed, surgical excision of tumors became increasingly more recognized as an effective form of cancer treatment. The associated procedures also became increasingly more aggressive (Hildanus, n.d.).

For most of human history, the cause of cancer was largely unknown. In the 16th century, associations were made between observations of cancer incidence and exposure to industrial chemicals (Hajdu 2011). In some of the earliest reports of hereditary cancer predisposition, the disease of cancer was mistakenly reported as contagious (Lusitani, n.d.; Tulp 1716). Until James Nooth challenged this conclusion and proved cancer is not a contagious disease in humans by injecting himself with pieces of breast cancer (Hajdu 2012), cancer patients were treated the same as those with infectious disease. Eventually, additional insight into the causes of cancer was provided in the late 1800s, when an observation was made about the unusually high prevalence of breast cancer in a woman’s family (Broca 1866). These observations and discoveries laid the foundation for later studies of cancer development and progression. Arguably, some of the most important findings from this early era of cancer research made clarifications that led to better understandings of what causes cancer and how to treat it.

The modern era of cancer research largely began when a theory emerged that proposed cancer was a result of mitogenic signals that can then cause chromosome abnormalities (Boveri 1914). Many years later, this theory was inadvertently proven when it was discovered that a tumor causing avian virus contained DNA that encoded a mutant version of an endogenous avian gene (Stehelin et al. 1976). The unmutated, endogenous gene was termed a “proto-oncogene”, and the mutant gene an “oncogene”, defined its ability to transform a cell (Eva Y. H. P. Lee 2010). Thereafter, cancer became appreciated as a disease which is caused by abnormal genetic sequences.

Later, oncogenes were discovered to be able to transform a cell even if wild-type alleles are still expressed, and their behavior was described as genetically dominant. Shortly thereafter, Rb1, named for its association with retinoblastomas, was discovered and characterized (Knudson 1971). However, it was distinct from previously characterized oncogenes because tumor formation required inactivating mutations of both alleles of Rb (J. M. Dunn et al. 1988). Confusingly, point mutations in other genes whose wild-type counterparts seemed to play tumor suppressive roles, in some cases, were able to confer oncogenic phenotypes even when wild-type alleles were still expressed in the cell (Willis et al. 2004). Later, these mutations were discovered to produce protein products that act in dominant negative fashion. Genes that, when deleted, cause formation of tumors and act to suppress proliferative processes were then later termed “tumor suppressor” genes.

In the years that followed, many other genes went on to be classified as oncogenes and tumor suppressor genes. Importantly, characterization of tumor associated mutants compared to their wild-type counterparts led to advancements in

our understanding of cell signaling machinery, DNA damage, apoptosis, and more (Hirota 1998; A. J. Wong et al. 1992; Laken et al. 1997; Nigro et al. 1989). Many of these studies were conducted in the context of development and other biological settings (Luetteke et al. 1999; Lin, Skapek, and Lee 1996; L. Liu et al. 2002). Generally, these studies showed that most mutations that underlie development of a tumor act by disrupting regulatory processes in the cells that give rise to the tumor. Over time, it has become clear that cancer progression is a manifestation of programs that result from dysregulation.

Despite these discoveries, many genes, particularly those which control tissue specific behavior, are poorly understood in development and in disease. Cancer treatments that were discovered in the mid to late 1900s work by causing tissue damage (Chabner and Roberts 2005). One of the most effective treatments, radiation therapy, revolutionized cancer treatment and has become increasingly more effective at tumor control and better tolerated by patients due to advancements in dosing regimen and delivery (Yifan Wang et al. 2018). Similar to other treatments, radiation causes tissue damage and DNA damage (Ward 1986). During this era, many drugs were discovered on the basis of how they affect cancer cell growth, rather than targeting a specific gene or process, and several of these are still utilized today. Early iterations of these therapies were poorly tolerated in patients, and in some cases had side-effects that made treatment more harmful than beneficial (Laviano and Fanelli 2012). Even amongst patients who benefited from early therapies, as well as those who benefit from modern versions of them, these drugs delayed, but did not cure, cancer (Hanahan

2014). In patients who do respond to these therapies, a fraction of tumor cells almost invariably become resistant to it.

Modern approaches to drug discovery in cancer research generally aim to create targeted and less toxic treatments. Many drug discovery methodologies seek to disrupt features of tumors that are either less important or do not affect normal tissue by specifically targeting highly recurrent tumor-specific mutations (F. Cheng et al. 2019). Improvements in our understanding of biology and the genes responsible for biological processes can then be used to inform approaches to cancer therapy (F. Cheng et al. 2019). Furthermore, cancer treatment has moved away from heuristics that governed earlier eras of cancer treatment, which promoted highly aggressive, invasive, and damaging treatments. Today, data driven efforts have stratified patients into groups based on clinical or epidemiological observations to identify patients that are most likely to benefit from treatments. These remarkable efforts have informed treatment guidelines for numerous kinds of cancers and treatments, including pharmacological, surgical, and radiological treatments (Sparano et al. 2018; Stearns 2018; Hamdy et al. 2016; Temel et al. 2010).

1.2 Implications of oncogenic mutations

Advancements in DNA sequencing have led to rapid molecular characterization of tumors across many cancer types. These efforts have revolutionized cancer research and treatment through identification of tumor-specific (and sometimes therapeutically actionable) mutations (F. Cheng et al. 2019). Cancer driver genes promote tumor progression by conferring evolutionarily advantageous phenotypes when mutated, relative to other cells around it (Martínez-Jiménez et al. 2020). It has recently been

appreciated that more tissue specific cancer driving genes exist than those that are found across many different kinds of cancers (Martínez-Jiménez et al. 2020). While cancers that arise from different tissue types have long been recognized as different diseases (National Institutes of Health (US) and Biological Sciences Curriculum Study 2007), the tissue specific behavior of some oncogenic mutations is still not well understood.

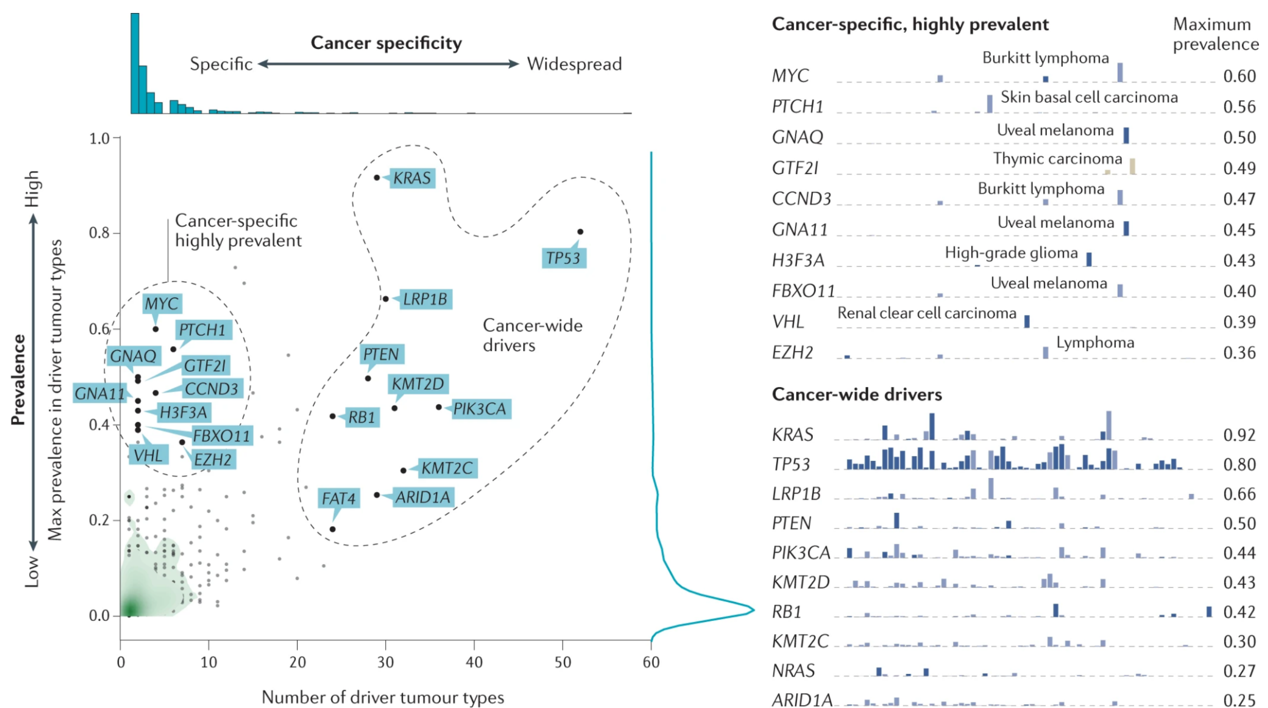


Figure 1. Prevalence of recurrent oncogenic mutations across cancer types. Scatter plot of cancer driver mutation prevalence within a given cancer type (y axis, “Prevalence”) versus the number of cancer types a gene acts as a driver gene for (x axis, “Number of driver tumor types”) (Left). Prevalence of cancer-specific highly prevalent drivers and cancer-wide drivers are labeled in the scatter plot (left) with specific breakdowns for each driver gene prevalence (y axis) in each cancer type (x axis) (right). Derived from (Martínez-Jiménez et al. 2020).

Many oncogenic signaling pathways are organized in a tissue specific manner that is specialized for homeostatic processes of that tissue type (A. Sharma and Sen 2013; Feng et al. 2007; K. Luo 2017). With the exception of de-novo activity conferred by rare gain-of-function mutations and translocation events, the majority of oncogenic mutations affect signaling and transcriptional networks that govern homeostatic processes in normal tissue (Martínez-Jiménez et al. 2020). As such, tissue specific specialization of signaling and transcriptional networks may influence how tissue types respond to oncogenic mutations. This idea is further supported by the frequency of mutations of particular genes across cancers of certain tissue types (**Figure 1**), and is most evident in phenotypes of hereditary oncogenic mutations that confer predispositions to specific kinds of cancer (Schneider et al. 2017).

Some widespread cancer driver mutations may confer different phenotypes across different types of tissues (Schneider et al. 2017; Garber and Offit 2005). Still, the effects of any individual oncogenic driver mutation all share an ability to give rise to a tumor in at least one tissue. Many of the phenotypes associated with cancer driver mutations in cancer cells are described as hallmarks of cancer, which are repeatedly observed across many different kinds of tumors (Hanahan and Weinberg 2011). Thus, although most cancer driver mutations do exhibit tissue specificity (Martínez-Jiménez et al. 2020), their associated phenotypes do share similarities that are frequently observed in many kinds of cancers (Hanahan and Weinberg 2011). Still, the mechanism by which cancer driver genes produce these phenotypes (Hanahan and Weinberg 2011) may be different.

Despite recognized and central regulatory roles in cell proliferation, some widespread cancer drivers, such as TP53 and RB, are thought to act in pathways with many functionally redundant regulators (Lipinski and Jacks 1999; Hanahan and Weinberg 2011; Ghebranious and Donehower 1998). Many of these genes drive several cancer types and are recognized as ubiquitous regulators of system stability and regulatory processes (Martínez-Jiménez et al. 2020), particularly those involved in proliferation. This finding is not surprising given that the defining feature of cancer is uncontrolled cell division (Weinberg 2013). However, these genes still have some degree of cancer type specificity (Martínez-Jiménez et al. 2020) (**Figure 1**). The roles of these genes in tissue homeostasis, which is defined as biological systems designed to respond to signals and maintain normal state of the tissue (Cannon 1929), and development is extremely important. In many cases, other genes in prevalent driver gene pathways or regulatory circuits can act in a functionally redundant manner (Schmale and Bamberger 1997; Van Nostrand et al. 2017). In these circumstances, functional redundancy must be overcome in order for the phenotypic effects of these mutations to manifest in cancer cells. Even the most widespread cancer driving mutations exhibit some degree of tissue specificity, suggesting that proliferation, or any processes implicated in widely shared hallmarks of cancer are at least partially regulated in a tissue specific manner (Castellano and Santos 2011). If functional redundancy of key regulatory genes is achieved through tissue specific gene expression, the mechanisms by which this redundancy is overcome in cancer cells may also be tissue specific. As such, some tissues may be more sensitive to a particular

oncogenic driver mutation than another, and could explain the prevalence of oncogenic driver genes across cancer types.

Components of oncogenic signalling networks are expressed broadly across tissue types. In most cases, there is very little understanding of how these networks are specialized for certain tissue types and it is not understood why certain mutations are able to transform some tissues but not others. It is possible that further characterization of cancer-specific driver genes may lead to further insight into the specialization of tissue specific gene expression and signaling networks. For more prevalent driver genes, a better understanding of tissue specific regulatory architecture and topology may also lead to insights of tissue specific functional redundancy.

One of the shared features of cancer driver genes commonly found across cancers of many tissue types is their involvement in processes that stabilize some aspect of a cell. As an example, P53 and DNA repair machinery confer genomic stability (Agarwal et al. 1998). In any case, dysregulation of these homeostatic processes creates opportunities for cancer cells to acquire new mutations or modify gene expression which can be advantageous to the cell or tumor by diversifying responses to selection pressures. Importantly, this can facilitate adaptation to changes in the environment of the tumor.

1.3 The Dynamic Tumor Microenvironment

For many cancer types, a tumor mass consists of tumor cells and a wide variety of other cell types (Whiteside 2008; M.-Z. Jin and Jin 2020). These include fibroblasts, blood vessels, nerves, and immune infiltrating cells (M.-Z. Jin and Jin 2020). Jointly, cells of the tumor microenvironment create chemical microenvironments that can impart

selective pressures on tumor cells and otherwise influence the behavior of other, non-cancerous adjacent cells. Additional biophysical or mechanical features of the microenvironment that are dictated by the organs where a tumor develops can also influence cells of the tumor microenvironment (X. Li and Wang 2020).

Tissue microenvironments are formed by an array of diverse cell types. As tumors grow and progress, the tumor microenvironment changes as a consequence of changes in physical properties, nutrient availability, metabolites, oxygenation, pH, etc. (M.-Z. Jin and Jin 2020). Many of these are a result of having an imbalance of cells proliferating or obstructing tissue structures, or tumor-mediated dysregulation of untransformed cells (Sugimoto et al. 2006; F. R. Balkwill, Capasso, and Hagemann 2012). In any case, as tumors grow the tissue they arise in is progressively pushed to a state of disequilibrium.

The tumor microenvironment, defined as the environment where tumors form and develop, exposes tumor cells to interactions with many different kinds of cells that constitute homeostatic regulatory systems. Importantly, this environment is thought to impart strong enough selective pressures on tumor cells that continually lead to selection of tumor cells which can exploit non-cancer cells of the tumor microenvironment for tumor-promoting functions (Baghban et al. 2020). Intercellular communication between tumor cells and cells of the tumor microenvironment is mediated through secretion of molecules which affect survival, inflammation, matrix remodeling, growth, and more (Baghban et al. 2020). Consequently, it is important to understand how cells of the tumor microenvironment respond to changes that result from tumor development and progression.

Many of the regulatory systems implicated in cancer conventionally act in tumor-suppressive ways, including as a surveillance mechanism for aberrant cells (Swann and Smyth 2007). As such, it is fascinating that tumor cells adapt to the effects of these systems in a manner that not only allows tumor cells to evade detection by these regulatory systems, but also exploit them (Hanahan and Weinberg 2011). The contrast between overt regulatory responses that are canonically anti-tumor with those that are tumor-promoting suggests that tumor progression requires changes in dynamic interactions to redefine their functional relationships. Thus, the tumor microenvironment has a strong influence over how a tumor progresses; iterative changes in the interaction between tumor cells and cells of the microenvironment likely govern tumor evolutionary processes (Lorusso and Rüegg 2008). Importantly, this creates the need to characterize cells of the tumor microenvironment in a longitudinal fashion, or in the context of tumor progression. Conversely, tumor cells should be characterized in the context of the changing tumor microenvironment when studying tumor progression.

1.4 Perspectives of Tumor Heterogeneity

Over the course of tumor progression, changes in the tumor microenvironment impart dynamically changing selective pressures on tumor cells. As such, mutations or changes in gene expression which can lead to further destabilization of the cell promote plasticity of cancer cell states; tumor cells with destabilized states may have better evolutionary fitness than those which do not (Hanahan and Weinberg 2011; Nemanja D. Marjanovic, Weinberg, and Chaffer 2013). Over time, this may lead to cells with dysregulated or destabilized states outcompeting those which are more stable. Because the tumor microenvironment is not uniform, the tumor cells which are favored in one part

of the microenvironment may not be favored in others. Additionally, this phenotypic diversification can lead to co-evolution of tumor clones which play functionally distinct roles in tumors (Tabassum and Polyak 2015). Globally, this leads to evolution of heterogeneous tumors and provides a possible explanation for variation in tumor therapeutic responses. In order for a productive therapeutic response, cancer therapies must simultaneously target multitudes of functionally distinct populations in a heterogeneous tumor. Overall, selective pressures and dysregulation of cell state together orchestrate changes which underlie tumor heterogeneity, evolution, and progression.

Histological heterogeneity in tumors has been observed for many years (Hanahan and Weinberg 2011) and is generally described as a feature of tumors that is dependent on time or tumor progression (Janiszewska 2020). Functionally, clonal subpopulations of heterogeneous tumors were known to have different characteristics before mutational heterogeneity was widely described (Heppner and Miller 1983). Tumor heterogeneity has been most heavily implicated in and studied in the context of therapeutic resistance (Dagogo-Jack and Shaw 2018) because tumor heterogeneity is intricately related to tumor cell plasticity and adaptation to selective pressures of therapeutic treatment. In the absence of therapeutic intervention, selective pressures from the tumor microenvironment, from other subclonal tumor cell populations, or from other sources, act dynamically. Thus, tumor heterogeneity has an extensive relationship with tumor progression. As such, understanding the processes by which tumors become heterogeneous is crucial to understanding how tumors progress.

1.4.1 Mutational Intra-Tumoral Heterogeneity

Loss of genome stability is one of the most widely recognized mechanisms by which tumors acquire heterogeneous mutations. It has multiple causes; however, irrespective of how genomic instability occurs, it produces a positive effect on the overall fitness of the tumor (Dagogo-Jack and Shaw 2018), within a certain limit. Several people have proposed that in certain contexts, genomic instability is required for tumor formation (Negrini, Gorgoulis, and Halazonetis 2010). In models for many kinds of cancer, tumor formation is thought to require a sequential series of defined mutations (Martincorena and Campbell 2015) that arise across many cell divisions and is a process by which intra-tumor heterogeneity inherently becomes created (P. C. Nowell 1976).

Because mutations are heritable, when a mutation arises in an evolutionary branch during tumor progression, it creates a clone. Clonal evolution in cancer is largely believed to occur through the Nowell model of tumor progression, in which sequential mutations create complex branched mutational trajectories (Greaves and Maley 2012) that are driven by selective pressures (**Figure 2**). Through studies of clonal mutations in cancer with whole exome sequencing and next generation DNA sequencing, in contexts where strong selective pressures exist, this model for subclonal selection has been supported (A. W. Zhang et al. 2018; Janiszewska et al. 2015). In some specific cancers, tumor progression has been defined by sequential acquisition of mutations which drive cancer (Martincorena and Campbell 2015). However, in reality, not all mutations confer an effect on the evolutionary fitness of a cell, but still arise in a clonal manner alongside mutations which do affect cell fitness. These mutations which do not affect cell fitness

are classified as passenger mutations (McFarland et al. 2017). In models of neutral evolution, after acquiring the mutations required for transformation, subclone evolution occurs through acquisition of passenger mutations (M. J. Williams et al. 2016), and the observed allelic frequencies of mutations which define subclones are driven by probability and exponential growth. In this neutral evolution model, these subclones do not evolve through selection. Notably, this model assumes there are no strong selective pressures which drive evolution during timescales in which neutral growth dynamics are observed.

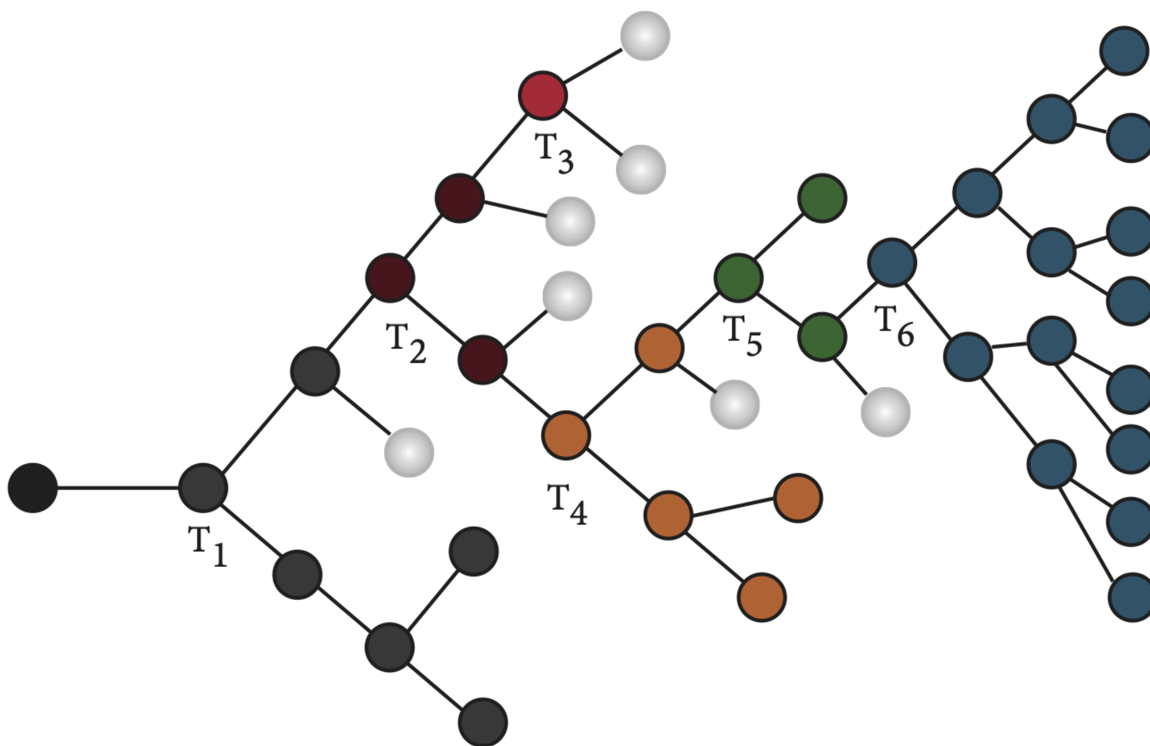


Figure 2. Clonal evolutionary tree of tumor cells. Adapted from (P. C. Nowell 1976). Tumor subclones are denoted as T1-T6. Cells in light grey depict subclones which undergo negative selection.

For many years, it was difficult to distinguish mutations in cancer driver genes from passenger mutations. While these efforts are largely ongoing, advancements in DNA sequencing technologies have afforded greater resolution of mutational patterns and their temporal dynamics in tumors (Goodwin, McPherson, and McCombie 2016). Clonal outgrowth of tumor cells which acquire mutations in cancer driving genes gives cells a selective advantage relative to those without these mutations (Martínez-Jiménez et al. 2020). Modern approaches to molecular profiling of mutations in tumors utilize the prevalence of a given mutation in clonal expansion of a tumor or its prevalence across patients to identify these cancer driving genes. Thereafter, the frequency at which these candidate driver genes are observed are compared to those expected for passenger and driver mutations in computational models of clonal evolutionary dynamics (Foo et al. 2015). In combination with these models, data derived from other sources (e.g. across multiple patients with the same cancer type) help to prioritize mutations in putative driver genes that are highly recurrent across patients and those known to play functionally important roles in cancer progression (Raphael et al. 2014). Today, many genetic mutations observed in cancer have been classified in a systematic manner.

The assumptions that underlie many models of tumor evolution limit their utility in describing generalized processes of tumor evolution, because tumors can evolve through a variety of mechanisms that are not mutually exclusive (Martincorena and Campbell 2015). In addition, the effect of selective pressures on tumor cells may not be constant throughout the course of tumor evolution. It is possible tumor cells evolve as an adaptation to selective pressures; in the absence of additional selective pressures, subclones may reach a point of equilibrium in which neutral evolution accurately models

tumor progression (M. J. Williams et al. 2016). However, additional selective pressures develop in response to therapy, in response to changes in the microenvironment that result from expansion of the tumor or disruption of its surrounding tissue, in response to the influence of latent variables that are unknown, or in response to the effects of spontaneously arising mutations which alter the fitness landscape of tumor cells. As such, when additional selective pressures are imposed upon tumors after a period of neutral growth, tumor evolution may instead follow the Nowell model until the tumor has adapted. Still, these models of tumor evolution have been essential in current understandings of tumor dynamics, heterogeneity, and evolution.

1.4.2 Functional Tumor Heterogeneity

Additional models of tumor evolution have more recently emerged that are formed on the basis that tumor subclones may have variable functional and phenotypic plasticity. The cancer stem cell model proposes that cancer stem cells are a functionally plastic group of cells within a tumor that produce and differentiate into populations that form subclones of the tumor (Clevers 2011). With respect to mutational status, under this model, mutations which define a subclone are generated in and from parental cancer stem cells (Greaves and Maley 2012). This model has been particularly influential in studying resistance to therapy (N. Y. Frank, Schatton, and Frank 2010; Ishikawa et al. 2007). In some cases, cancer stem cells have been implicated as a source of genetic diversity in a tumor (Greaves 2010) that can underlie therapeutic resistance (J.-K. Kim, Jeon, and Kim 2015). Cancer stem cells are defined by their functional plasticity, which can be a consequence of genomic instability, but can also arise through changes in gene expression that are not driven by mutations (Kreso and

Dick 2014). In fact, not all histological heterogeneity can be explained by heterogeneity in genetic mutations (Hlubek et al. 2007; Stanta and Bonin 2018).

Irrespective of mutation status, under the cancer stem cell model, a cancer stem cell has the capacity to give rise to functionally distinct subclones, each of which can promote, facilitate, and/or drive tumor progression (Shackleton et al. 2009). Classically, a single cancer stem cell must also be able to give rise to all of the cell types that are observed in a heterogeneous tumor (Shackleton et al. 2009). From mutational status alone, it is difficult to distinguish cancer stem cell driven tumor heterogeneity from subclonal selection without empirical knowledge of the functional plasticity of parental subclones. For this reason, the cancer stem cell model is largely ignored in studies of intra-tumoral mutational heterogeneity. Instead, gene expression of tumor subpopulations have been used to further interrogate the cancer stem cell model of tumor evolution.

In cancers such as leukemia, this model was largely supported when it was found that only a small subpopulation of tumor cells, cancer stem cells, had the capacity to form new tumors (Bonnet and Dick 1997). Cancer stem cells have been extensively implicated in disease recurrence (J.-K. Kim, Jeon, and Kim 2015) and putative cancer stem cells in some contexts have exhibited gene expression profiles which promote resistance to chemotherapy and radiation (J.-K. Kim, Jeon, and Kim 2015).

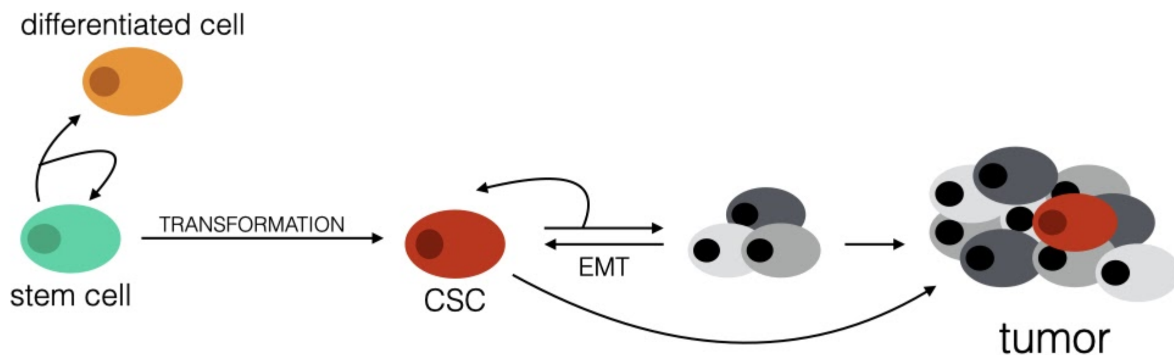


Figure 3. Adapted from (Bonnet and Dick 1997; A. Singh and Settleman 2010). Non-malignant stem cells (green) that can give rise to differentiated cells (orange) undergo transformation to become a cancer stem cell (CSC, red). Cancer stem cells can reversibly give rise to multiple cell types of a tumor (shades of grey).

Measuring gene expression in individual tumor cells is a widely accepted approach for studying tumor heterogeneity (M. Li et al. 2020) and is one of the most prevalent ways in which functional heterogeneity in tumors has been described. The revolutionary discovery that cancer cells can latently reactivate an embryonic developmental program, the epithelial-to-mesenchymal transition (EMT), generated strong support for the cancer stem cell hypothesis (Mani et al. 2008). This program was canonically described during formation of the primitive streak in chick embryos, a process that is important for embryonic gastrulation (Hay 1995). Cells which have undergone EMT during development have the ability to differentiate into many cell types. Importantly, these cells are known to have better migratory potential as a consequence of change in polarization and subsequently in overall morphology (A. Singh and Settleman 2010).

Across many tumor types, particularly those which arise in the breast and nervous system, there is extensive evidence for the existence of cancer stem cells and the importance of their roles in tumor progression (Bjerkvig et al. 2005). Amongst these kinds of tumors, questions still remain about the origin of cancer stem cells. Many tumors are thought to arise from tissues that are terminally differentiated (Sell 2010). Widely accepted models of cell differentiation in development postulate that cells which are more differentiated have less potential to give rise to different cell types than those which are less differentiated. This forms the basis for the Waddington model of cell differentiation (J. Wang et al. 2011). As such, the notion that a multipotent stem cell could evolve from a terminally differentiated somatic cell challenged the Waddington model. With the discovery that terminally differentiated somatic cells can be artificially reprogrammed into pluripotent stem cells (Kyttälä et al. 2016), it became more widely appreciated that it is possible for terminally differentiated cells to become reprogrammed and afforded greater plausibility to the cancer stem cell model.

While this model has been extremely useful in understanding the progression of several cancer types, it is considered controversial. Many criticisms of the cancer stem cell model are based upon the expected prevalence of stem cells in a tumor, and in most studies, is reported to be a rare subpopulation of a tumor (Shackleton et al. 2009). If cancer stem cells are the only populations within a tumor that are able to proliferate, then observed tumor growth kinetics are inconsistent with the cancer stem cell model (Shackleton et al. 2009). At the same time, cases of resistance to cancer therapy have provided strong supporting evidence for the model. Frequently, a less controversial

stance is taken to the cancer stem cell model and it is often reported to be well supported in some cancer types, but not others (Shackleton et al. 2009).

In any case, studies concerning stem cells in cancer created extensive discussion about the differentiation state of cancer cells, with particular importance on the developmental history of a cell. Striking evidence arose implicating transcription factors known to play essential roles during embryonic development, even outside of the context of the EMT, in cancer progression (Mani et al. 2008). Later studies that drew associations between gene expression programs and tumor progression status, particularly the capacity to seed metastases, further characterized how differentiation status in solid tumors becomes dysregulated in tumor evolution (Winslow et al. 2011). In many cases, cells not only progressively lose their original identities, but also simultaneously adopt hallmarks of different tissues (Snyder et al. 2013). Because cell differentiation is thought to be driven by progressive epigenetic modification, profiling chromatin accessibility in cancer cells has more recently become a focal point of cancer research.

1.4.3 Tumor Heterogeneity and Metastasis

Metastasis is the primary cause of complications which lead to death in cancer patients (Christofori 2006). The end-state of cancer progression has conventionally been distant metastatic spread; when cancer has progressed to this point, therapeutic intervention is generally considered non-curative (Croker and Allan 2008; L. Dong et al. 2019; Révész et al. 2017). Tumor heterogeneity is a likely mechanism through which cancer cells metastasize because it generates a diverse array of tumor subclones with varying functions, phenotypes, and capabilities. Some fraction of these subclones either

may promote metastasis or may themselves be capable of metastatic spread (Révész et al. 2017).

A widely accepted model of metastatic spread in cancer is frequently referred to as the Metastatic Cascade (Hapach et al. 2019). In this model (J. Yang et al. 2004) (**Figure 4**), epithelial lesions (carcinomas) develop at the primary site and then begin to disrupt the tissue. Invasive lesions arise by invading through the basement membrane, which defines the boundaries of epithelial tissues, in a process known as cell migration. Thereafter, these lesions become invasive carcinomas and are considered to be malignant. Malignant cells can continue to expand and proliferate in the primary site, and can also eventually intravasate into the surrounding vasculature and lymphatic vessels, after which they become circulating tumor cells. This process is thought to occur through multiple mechanisms; the first of which involves invasion and circulation of an individual cell. The second model for this process involves multiple cells that facilitate different processes associated with intravasation and the remaining steps of the metastatic cascade, and is termed collective cell migration (Yang Yang et al. 2019). After gaining access to the bloodstream, if these circulating tumor cells survive, some fraction of them will extravasate from the circulatory or lymphatic system and infiltrate organs that can be close to the primary site (local metastasis) or far away (distant metastasis). In the final step of the metastatic process, some of these infiltrating cells colonize these sites,resulting in micro-, and eventually, macro-metastases.

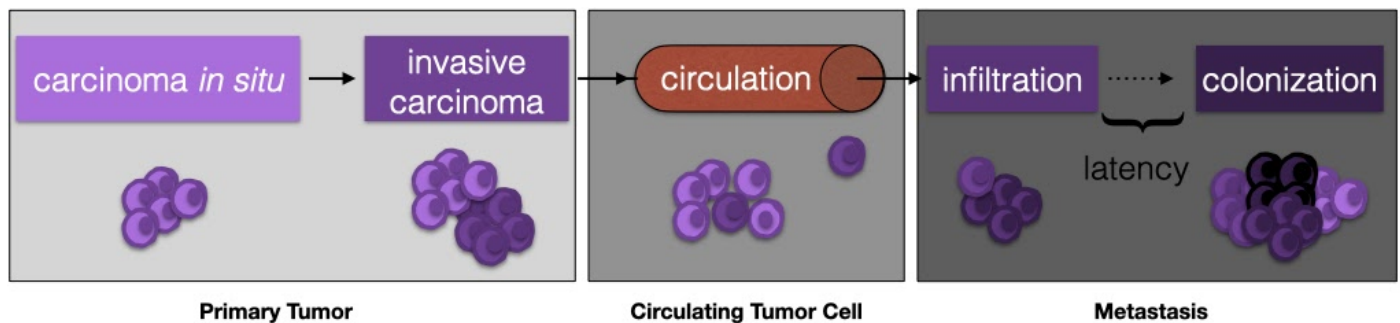


Figure 4. The Metastatic Cascade. Adopted from (J. Yang et al. 2004).

The functional requirements for completion of the metastatic cascade are numerous, and as a consequence, metastasis is believed to be an inherently inefficient process (Hapach et al. 2019). These inefficiencies are compounded by diverse selective pressures that act upon cancer cells in the different environments associated with the metastatic cascade, particularly in the circulatory system, where cells are exposed to circulating lymphocytes. A number of mechanisms have been proposed to address how circulating tumor cells survive these selective pressures (Q. Liu, Liao, and Zhao 2016; Raimondi et al. 2017; Lo et al. 2020). In general, circulating tumor cells are thought to be short lived (Krog and Henry 2018) and must overcome selective pressures in short time scales. Consequently, negative selection during cancer cell dissemination is thought to create a strong bottle-neck and very few cells are thought to survive this process. Furthermore, additional selective pressures in the tissues that circulating tumor cells eventually infiltrate cause an additional bottleneck in metastasis (Massagué and Obenauf 2016). Many cells which reach distant organs may lie dormant for many years, not actively proliferating (Massagué and Obenauf 2016). In some cases, these

infiltrating tumor cells are unable to seed a macro-metastatic lesion (Massagué and Obenauf 2016).

Ultimately, tumor heterogeneity is implicated in metastasis because it generates phenotypic diversity which can allow cancer cells to overcome selective pressures of the metastatic cascade. Although inefficient, cells disseminate from the primary tumor frequently (Celià-Terrassa and Kang 2016). The rate at which these cells survive and complete the metastatic cascade may be low, but over time, the low probabilities of survival are effectively cumulative (Szczurek et al. 2020). The statistical likelihood that an individual cancer cell can generate distant or local metastasis is low, but with respect to overt disease progression, the likelihood that a disseminating tumor cell will successfully seed a metastatic lesion effectively increases as more cells migrate away from the primary tumor over time (Szczurek et al. 2020). In addition, changes in the composition of the primary tumor, which are also largely dependent on time, can increase the likelihood of metastasis by creating a microenvironment that generates tumor subclones which can overcome selective pressures of metastasis (Szczurek et al. 2020).

The majority of the work presented throughout this thesis is primarily concerned with time-dependent changes in the primary tumor, as these changes are major determinants of productive metastasis. Importantly, understanding dynamic changes of primary tumors can still lead to meaningful insight on metastatic processes. Some studies conducted on primary tumors have proposed that certain cancer cells may become poised to metastasize to particular regions of the body while still located in the primary tumor, which is often described as formation of a pre-metastatic niche (Peinado

et al. 2017). While this body of work focuses on the primary tumor, it is important to note that later steps of the metastatic cascade are equally important in driving overall disease progression.

1.5 Emerging Evidence for the Role of the Epigenome in Tumor Heterogeneity

Gene expression is a product of chemical modifications of the genome that control accessibility of DNA without changing its sequences, termed epigenetics (Holliday 1987). Epigenetics is important in biology primarily because it facilitates cell fate commitment and maintenance of cell identity (Morris and Daley 2013). The aberrant epigenomic landscape of cancer cells has been described for many years in cancer research (Jian Cao and Yan 2020). Epigenetic dysregulation is not surprising, given that many tumors have hallmarks of increasingly more aberrant differentiation as they progress. Further, dysregulation of chromatin structure can lead to further regulatory instability (Reinberg and Vales 2018) and may be a mechanism of cancer cell plasticity. Chromatin accessibility is thought to be a determinant of gene expression, and has therefore been an important topic of interest in biology.

Chromatin accessibility is regulated through multiple forms of epigenetic modifications. DNA itself can be directly modified through methylation (Feinberg and Vogelstein 1983) of regulatory sequences frequently found at the 5’ end of genes, many of which exhibit tissue specific expression (Esteller 2008). Additionally, chemical modification of histones, proteins that package DNA, also affect chromatin accessibility (Cavalli and Heard 2019). There is also evidence that chromatin structure can be regulated through other mechanisms, including through non-coding RNAs, to influence ‘higher-order’ chromatin structure. Together, epigenetic modifications orchestrate

complex, 3-dimensional chromatin topologies which are heritable and are believed to ultimately be responsible for regulation of gene expression and cell identity (Cavalli and Heard 2019). Our understanding of how DNA sequences define, regulate, and/or control biological processes (E. T. Liu 2008) is largely the result of advancements in functional genomics. In particular, chromatin accessibility at transcription factor binding sites, which varies across different tissues, is one of the ways cell identity is thought to be regulated (Deplancke, Alpern, and Gardeux 2016). Functionally, it is well established that epigenetic state regulates gene expression (Gibney and Nolan 2010). As such, associations between gene expression and epigenetic features have led to recent computational advancements that aim to identify regions of the genome that are regulated to control cellular identity (Shim et al. 2020).

Chromatin structure and topology are extraordinarily complex and highly regulated (Esteller 2008). Our mechanistic understanding of epigenetics and how chromatin topology or accessibility act to regulate gene expression networks has been largely restricted by technical limitations of measuring epigenetic modifications or chromatin structure. Until recently, the vast majority of epigenetic profiling was conducted through identification of sites in the genome where a particular epigenetic modification is located, which is measured in material isolated from a large number of cells (Buenrostro et al. 2015). While the effects of many epigenetic modifications are well characterized in isolation, they act in a combinational fashion to regulate chromatin accessibility. As such, it is not always possible to predict chromatin accessibility by mapping particular epigenetic modifications (Kouzarides 2007). For this reason, the development of assays which allow epigenomic profiling that are not dependent on

interpreting epigenetic modifications themselves have been crucial in epigenetics. Advancements in these assays led to the development of ATAC-seq, Assay for Transposase-Accessible Chromatin with high-throughput sequencing (Buenrostro et al. 2015).

As our mechanistic understanding of oncogenic mutations has improved, it has become increasingly more widely accepted that cancer progression is driven by mutations in a limited number of genes and is dependent on epigenetic changes (Klein and Klein 1985; Vogelstein et al. 2013; Martínez-Jiménez et al. 2020). Early studies of DNA modifications in cancer found that hypermethylation of tumor suppressor gene promoters that results in gene silencing could be a mechanism by which tumor suppressor genes are inactivated and has otherwise been implicated in destabilizing the genome and cell state (Esteller 2008). Conversely, loss of these modifications in other areas of the genome has also been implicated in genomic instability as well as aberrant expression of embryonic and tissue restricted genes (Esteller 2008).

Perhaps the most compelling evidence for epigenetic dysregulation in cancer is that many genes which modify chromatin accessibility are highly mutated across a variety of cancer types (C. Plass et al. 2013). The epigenomic landscape of cancer is highly dysregulated (Jones and Baylin 2007), both in terms of DNA methylation and histone modification. Furthermore, the precisely regulated epigenetic landscape in pluripotent embryonic stem cells (Viré et al. 2006; S. Sharma, Kelly, and Jones 2010) is largely responsible for their plasticity and differentiation potential. Indeed, epigenetic reprogramming is the primary mechanism by which terminally differentiated cells can be made into induced pluripotent stem cells (Okita, Ichisaka, and Yamanaka 2007). This

lends support for studies of epigenetic dysregulation in cancer stem cells (Toh, Lim, and Chow 2017), but also may be a mechanism by which differentiation status becomes dysregulated in tumor types which do not conform to the cancer stem cell model.

Notably, some experimental cancer models exist that reliably recapitulate clinical disease but are generated through experimentally defined mutations. In many of these models, mutational burden of tumors can differ from those observed in clinical disease (Westcott et al. 2015). Other features of disease progression, such as tumor histology, are quite similar to human disease. This suggests that other changes in tumor cells, and chromatin accessibility in particular, may be partially responsible for the observed changes in gene expression between normal and cancerous tissues. This, in part, created a precedent for the importance of epigenetic dysregulation in cancer progression.

1.6 An Evolutionary Approach to Characterize Tumor Heterogeneity

Although tumor heterogeneity makes tumors difficult to treat and to study, it can be exploited by analytical approaches which have led to many insights about how tumors evolve and progress. Distinct subclones with different functionality within a tumor are thought to arise in response to dynamic selective pressures that, by extension, constrain and influence tumor evolution. Estimation of variant allele frequencies of tumor-specific mutations have permitted more rigorous testing of theoretical models for evolution of mutational heterogeneity, discussed in earlier subsections. Because mutational heterogeneity influences overall evolution of a tumor by affecting gene expression and chromatin accessibility, models of mutational heterogeneity may inform generalized models of tumor evolution. Despite widespread recognition that epigenetic

dysregulation is equally as important as DNA mutations in cancer, few studies exist to describe how this dysregulation can be incorporated into and refine existing models of tumor evolution.

Several studies provide evidence that tumor cell epigenomes become dysregulated in a heterogeneous and stochastic manner (M. Guo et al. 2019; LaFave et al. 2020). Epigenetic dysregulation is likely to drive and itself become a manifestation of tumor heterogeneity. As stated in the previous section, there is substantial evidence for epigenetic dysregulation in settings where cancer stem cells have been identified and described. However, the majority of *quantitative* models of mutational tumor evolution are consistent with, but do not explicitly describe, the cancer stem cell model. In turn, this creates a disconnect between models of tumor evolution which explicitly incorporate epigenetic dysregulation with those which cannot. This is due, in part, to experimental limitations in profiling cancer cell epigenomes. As ATACseq becomes more prevalent, it is likely that enough data will be generated such that quantitative modeling will become possible.

Perhaps the biggest confounding limitation in studies of tumor heterogeneity and evolution is the variation in time over which disease progression occurs. In clinical settings, cancer progression occurs with varying kinetics across different patients. For some of these cases, this variation can at least be partially attributed to variation in responses to treatment, both in efficacy and duration of response (Mitsudomi et al. 1996). However, even amongst patients which do not elect to receive treatment, variation in disease kinetics is still observed (Crispen et al. 2009). Some explanations have proposed this variation is a result of stochastic mutational processes (Hao, Wang,

and Di 2016), but this does not explain variation in cancer progression kinetics across *all* kinds of cancer. This could be partially related to the fact that some cancers appear to be driven by epigenetic, rather than genomic, dysregulation.

An additional confounding limitation is the reliability through which patients and their tumors are stratified into subtypes of cancer. Data-driven analyses and molecular profiling have led to dramatic improvements in how patients can be stratified according to subtype and clinical behavior (McVeigh et al. 2014), and these data can also be used to predict therapeutic outcomes amongst other patients. In spite of these advancements, many additional questions remain about both the underlying causes of variation between patients, as well as the temporal behavior of disease progression. Thus, the ability to accommodate variation in tumor profiling data that is caused by underlying variation in disease kinetics or disease type is relatively limited.

Although this variation remains, many efforts are being made to generate more data on tumor cells using cutting-edge technologies such as ATAC sequencing. This is important because evidence is still being collected to assess the nature of epigenetic dysregulation that occurs across many kinds of cancer. As these data accumulate, they allow refinements in models for epigenetic and transcriptional trajectories in tumor progression and may ideally be used to describe deterministic changes in cell state. Such a model must account for the fact that chromatin accessibility is a product of the cumulative and combinatorial effects of many epigenetic regulators that can be influenced through multiple ways. While the role of epigenomic and transcriptional dysregulation is widely accepted as being at least partially responsible for tumor progression, the complex nature through which chromatin accessibility is regulated and

the dynamic processes that drive changes in epigenetic landscapes are quite poorly understood. In many cases, it is not possible to reliably predict the functional consequences of an individual, or combinatorial, epigenetic mark in terms of effective chromatin accessibility (Santos-Rosa et al. 2002). As a consequence, this has driven efforts to get improved resolution of transcriptional and epigenomic states to create a basis for mathematical modeling that will inform how cell state changes occur during cancer progression.

Efforts to profile the effective phase space of cancer cells have been further complicated by epigenetic and transcriptional heterogeneity in tumors. One important distinction between interpretation of mutational heterogeneity and epigenetic or transcriptional heterogeneity is the degree of independence between mutational and transcriptional or epigenetic changes. Mutations are mostly thought to occur at random loci, barring some notable exceptions (de la Chapelle 2003). In contrast, epigenetic marks are believed to be deposited through the concerted efforts of large, multi-protein chromatin modifying complexes which are directed through a series of networked interactions and activities (L. Y. Wang et al. 2011). Transcription is also regulated in dynamic networks with topologies that are thought to be highly tissue specific (He and Tan 2016). Therefore, a single regulatory change can manifest through a cascade of genes that are connected in a transcriptional (or regulatory) network. Additionally, changes in transcription of an individual gene cannot be assumed to be independent from changes in another. Consequently, evolution of tumor cell transcriptional state occurs in a coordinated manner that is dependent on underlying genetic and regulatory networks. The constraints imposed by these networks are latent determinants of the

topologies of accessible transcriptional states of a tumor cell. Because these networks are strongly influenced by cell identity (Shining Ma, Jiang, and Jiang 2018), which becomes dysregulated over time, the influence of these network constraints is also dynamic over tumor evolutionary processes.

Mutational heterogeneity can be effectively modeled in tumors through the assumption that mutations are stochastic and arise independently. However, these assumptions cannot necessarily be extended to models of transcriptional and epigenetic heterogeneity. Changes of cell identity, which may occur stochastically, can cause effective rewiring of intracellular circuits that allow cells to respond to changes in their environment (Irish et al. 2004). To understand how changes occur in the epigenetic and transcriptional states of tumor cells over time, the constraints of these networks need to be defined (Bandara et al. 2009). However, these constraints change over disease progression and are themselves subject to the variation in disease kinetics observed across patients.

While these sources of variation have restricted models of tumor evolution to be primarily descriptive, rather than quantitative, the inherent heterogeneity of tumors creates analytical advantages. Variation in disease kinetics makes it essentially impossible to reproducibly sample tumor evolutionary processes at a precise point in time. This can be exploited when looking across individual cells, because an individual tumor contains heterogeneous cells that occupy a mosaic of observable accessible states which can be loosely, but not precisely, associated with time. In turn, this decreases the number of time points over which tumor progression must be sampled in order to generate complete coverage of cell states which arise during tumor evolution,

because the subclones of a heterogeneous tumor can reflect its evolutionary history, present, and future. Thus, profiling a single tumor at a single point in time can produce rich amounts of data about the cells it contains that together reflect a set of cell states which reflect more than just a single point in time during tumor progression. Still, tumor evolution is affected by multiple systems which exert selective pressures on a tumor. An explicit model of tumor evolutionary processes with respect to cell state must incorporate higher order interactions between systems that exist within and between cells.

2. Tumor immunology

2.1 Inflammation and cancer

Immune cell infiltration into solid tumors has long been recognized as being positively associated with increased overall survival, most notably in human melanoma (Clemente et al. 1996; Galon et al. 2006; L. Zhang et al. 2003; Fridman et al. 2012). This association is likely related to the fact that the immune system is responsible for surveillance, detection, and elimination of malignant cells. The immune system works through coordinated activity of many different kinds of cells which each play diverse, but defined, roles to generate an immune response. In order for a tumor to develop, it must either evade or adapt to overcome the tumor suppressive actions of the immune system (Hanahan and Weinberg 2011). There are multiple mechanisms by which the immune system can detect and suppress cancer; reciprocally, there are thought to be multiple mechanisms through which tumors suppress and evade these responses (Hanahan and Weinberg 2011). Although it is well accepted that the immune system can be tumor suppressive, some of the early observations about how immune responses are related

to cancer progression suggested that inflammation can promote tumor formation (F. Balkwill and Mantovani 2001). Still, other early observations also associated inflammation with tumor suppressive activity (Hoption Cann et al. 2002).

The relationship between cancer and inflammation remained largely ambiguous until it was discovered that mice which had previously been able to eradicate a transplanted tumor could more effectively eliminate tumors that were transplanted after the first tumor had regressed. The behavior of these responses were remarkably consistent with that of an adaptive immune response and further suggested that tumors must continuously evade the adaptive immune system in order to develop and progress (G. P. Dunn, Old, and Schreiber 2004; Gross 1943; Silverstein 2001). In addition, inflammation creates a microenvironment that stimulates cell proliferation and remodeling of the surrounding extracellular matrix and vasculature in response to a wound. This led to connections that were made about the similarity between the wound healing response and many hallmarks of cancer, and further clarified how the immune system can also promote cancer. As such, it was proposed that tumors exploit these programs and sustain this microenvironment to grow and progress (Flier, Underhill, and Dvorak 1986; Coussens and Werb 2002).

One of the other widely accepted mechanisms of tumor-immune evasion causes immune cells to become dysfunctional in response to tumors that have established an immunosuppressive microenvironment (A. A. Wu et al. 2015). This occurs by hijacking homeostatic programs of immunosuppression that canonically exist to prevent autoimmunity and other harmful immune reactions (Liston and Gray 2014). Notably, immunosuppressive microenvironments can be produced through a variety of different

mechanisms and can affect multiple types of immune cells that coordinate the overall immune response to a tumor (D. Wang and DuBois 2015).

2.2 Influence of solid tumor biology on immune response

Recent evidence has emerged to suggest that tumor regions with higher mutational heterogeneity may drive greater immune responses relative to tumor regions with lower mutational heterogeneity (AbdulJabbar et al. 2020), suggesting tumor mutational heterogeneity may have a meaningful influence on the overall tumor-immune response. This is not surprising, because mutations create abnormal protein products of mutant proteins that can be presented on the surface of tumor cells to immune cells as tumor-specific neoantigens.

Evasion of the immune system can occur in tumors through multiple ways. One of the most widely recognized ways this occurs is through dysregulation of the antigen presentation machinery, but this does not occur across all tumors (Hanahan and Weinberg 2011). The majority of tumor-specific neoantigens arise from passenger mutations which are dispensable in tumor progression (Jhunjhunwala, Hammer, and Delamarre 2021). As such, passenger mutations which create immune reactive neoantigens can undergo immunoediting, wherein these mutations are deleted or silenced in response to selective pressure from immune cells responding to their associated neoantigens (R. D. Schreiber, Old, and Smyth 2011). Antigens are loaded and presented on the surface of tumor cells or immune antigen presenting cells after being processed in the immunoproteasome (Jhunjhunwala, Hammer, and Delamarre 2021), which is induced by inflammatory signals of the microenvironment (G. H. Wong and Clark-Lewis 1983).

While many kinds of immune cells play important roles in response to tumors, the majority of the work presented in this thesis will focus on CD8 T cells, which are discussed in the following sections.

2.3 Characterization of cytotoxic T cell dysfunction

2.3.1 Differentiation of cytotoxic T cells

T cells are lymphocytes that play a major role in cell mediated responses to antigens, or molecules that elicit an immune response (Institute and National Cancer Institute 2020). They differentiate from thymocytes, which are T cell precursors that have differentiated from hematopoietic stem cells in the bone marrow. They are defined by expression of T Cell Receptor (TCR) proteins, which are created as T cells mature in the thymus through recombination of Variable, Diversity, and Joining gene fragments to produce TCRs that can recognize a wide diversity of antigens (Schatz and Ji 2011) on the surface of antigen presenting cells. When TCRs are engaged with their cognate antigens, intracellular signaling networks orchestrate T cell expansion and differentiation into T cells that mediate inflammation (T-helper), immune suppression (T-regs), memory (T-memory) and cytotoxicity (T-effector) (Viola and Lanzavecchia 1996; Youngblood et al. 2017; Speiser, Ho, and Verdeil 2016).

Cytotoxic T cells, marked by expression of CD8, will proliferate when activated to generate a clonally expanded pool of effector cells that have the ability to secrete cytokines and proteins which can induce cell death. They are most often characterized for mediating control over pathogen infection and cancer cells (Zhou et al. 2010; Waldman, Fritz, and Lenardo 2020). Priming or activation of cytotoxic T cells is a tightly regulated process and is thought to have multiple requirements to become fully

“licensed” (Thaiss et al. 2011) for effector and memory function. These include engagement of the TCRs of CD8 cells with antigens presented on MHC-I, the binding of costimulatory molecules, cytokine signaling, and chemokine gradients (Thaiss et al. 2011). Collectively, these signals are processed through networks of signaling transduction pathways to produce transcriptional outputs that mediate functional differentiation (Viola and Lanzavecchia 1996; Smith-Garvin, Koretzky, and Jordan 2009). These networks are also influenced by co-inhibitory receptors which primarily serve to dampen immune responses and prevent auto-immunity (Chen and Flies 2013).

The complexity of the signals and the components of these networks permit precise, multifaceted control over signaling outcomes which ultimately influence the fate of functionally differentiating cytotoxic T cells (Chen and Flies 2013). Naive CD8 T cells have not been exposed to the antigen their TCRs are specific for. After antigen exposure and activation, CD8 T cells functionally differentiate into different subpopulations. Profiling of tumor-reactive CD8 T cells revealed these cells can occupy a number of differentiation states along a continuum of differentiation states, including naive, effector, memory, and exhausted states (van der Leun, Thommen, and Schumacher 2020).

After naive T cells develop in the thymus, they circulate around the body and infiltrate different organ systems in search of immune cells which have presented their cognate antigen on MHC molecules (Mandl et al. 2012; Smith-Garvin, Koretzky, and Jordan 2009). Once a TCR binds an antigen-MHC complex, the process of T cell activation begins (Smith-Garvin, Koretzky, and Jordan 2009). However, engagement of a cell’s TCR without additional activation signals cannot fully license a T cell to undergo

functional differentiation; instead, cells which receive a TCR activation signal in the absence of other required activating signals become anergic, and are unable to proliferate or secrete inflammatory signals in response to antigen. Importantly, this anergic state can be reversed through exposure of inflammatory signals (Appleman and Boussiotis 2003).

The additional signals required for activation can come from and be influenced by a variety of other immune cells. One of the most well described mechanisms of activation occurs through engagement of CD28 on the surface of T cells through binding of costimulatory ligands on antigen presenting cells (Harding et al. 1992). This process can be further facilitated by Helper T cells CD4+ T cells in a manner that permits precise tuning of functional differentiation to promote immune memory that is characteristic of adaptive immunity (S. Zhang, Zhang, and Zhao 2009). The output of these systems can also be modulated by inflammatory signals of the microenvironment that can be produced by a variety of cell types that also facilitate functional activities of differentiated cells (J. M. Curtsinger et al. 1999). Importantly, some of the other signals which affect T cell activation create negative feedback to control activation signals and are induced as a consequence of T cell activation. A notable example of these feedback signals are ‘checkpoint molecules’, such as CTLA-4 and PD-1, both of which negatively regulate TCR signaling.

Activated T cells will proliferate and expand until their targets are cleared from the system and will subsequently undergo programmed cell death. The actions of these cells is one of the primary ways that infection is cleared or malignant cells are eliminated (Kaech and Cui 2012). A smaller subset of CD8 T cells will differentiate into

memory cells that persist after cytotoxic T cells have cleared their targets and then mediate efficient immune responses to their cognate antigen following subsequent exposure. The progenitor CD8 cells that canonically differentiate into memory subsets are marked by expression of TCF-1, the protein product of the gene Tcf7 (Jiaxue Zhang et al. 2021), (Zhou et al. 2010).

Importantly, CD8 T cells can also functionally differentiate into a variety of subsets which have been implicated in mediating tissue specific immune compartments, context-specific viral responses, and more (Mittrücker, Visekruna, and Huber 2014). For the most part, these T cell subsets are functionally distinguished by a specific permutation of interleukin and cytokine receptors as well as transcription factors which are responsible for changes in gene expression that mediate different functional roles of each subset. These T cell subsets can be transcriptionally quite similar and have subtle, but important, differences in functionality.

One example of these subsets, termed Tc17 cells, are induced by TGF-Beta and IL-6 or IL-21 signaling. They are also distinguished from other T cell subsets by expression of ROR-gammaT and ROR-alpha (Hamada et al. 2009). Tc17 cells are similar to conventional cytotoxic T cells because they also produce cytotoxic molecules, but do so at a reduced level and therefore are known to have lower relative cytotoxic activity (Hamada et al. 2009). Tc17 cells are functionally important for driving inflammation and mediating control over diseases in which T cells are persistently exposed to antigens, such as in chronic viral infection or in tumor development (Mittrücker, Visekruna, and Huber 2014).

2.3.2 Hallmarks of T cell dysfunction and exhaustion

T cell exhaustion is defined by reduction in cytokine production and increased expression of co-inhibitory receptors (Welten et al. 2020) that ultimately cause depletion of resources that can be utilized to sustain an immune response (Thangavelu, Smolarchuk, and Anderson 2010). It is conventionally described in the context of chronic viral infection and is characterized by loss of effector function and reduction in clonal expansion (Thommen and Schumacher 2018). There are many parallels between immune responses to tumors (Baitsch et al. 2011) and chronic viral infection (Wherry et al. 2007). Persistent antigen exposure, in particular, is largely responsible for similarities between exhausted T cells in both settings because it induces sustained expression of co-inhibitory proteins (Thommen and Schumacher 2018).

For a T cell to become exhausted, it must have sustained expression of multiple co-inhibitory receptors (Chen and Flies 2013). Importantly, expression of co-inhibitory proteins is a programmed output of interactions with cells which serve to dampen immune responses, such as Tregs, as well as signaling associated with functional differentiation and stimulation of T cells. Expression of co-inhibitory receptors and ligands serve as mechanisms of intra- and inter-cellular negative feedback (Thangavelu, Smolarchuk, and Anderson 2010). As such, transient expression of co-inhibitory molecules also occurs as a consequence of signaling associated with activation (Thangavelu, Smolarchuk, and Anderson 2010). Sustained activation signals therefore cause accumulated expression of co-inhibitory molecules which, in turn, act to suppress the immune response. In some cases, this can lead to abnormalities in T cell

differentiation that produce dysfunctional T cells that phenotypically resemble exhausted T cells (Schietinger and Greenberg 2014; Wherry et al. 2007).

In the context of tumor immunology, T cells of the tumor microenvironment that phenotypically resemble exhausted T cells are described as dysfunctional because T cell responses can sometimes be invigorated with therapeutic treatment (Wei, Duffy, and Allison 2018) and thus do not completely fulfill requirements of exhaustion because dysfunction can be rescued. Co-inhibitory receptors perform non-redundant roles in immune suppression (Thangavelu, Smolarchuk, and Anderson 2010) and are expressed in a context-dependent manner. As such, it is not surprising that dysfunctional tumor associated T cells show similar, but distinct, gene expression signatures relative to exhausted T cells in the context of chronic infection (Thommen and Schumacher 2018). Ultimately, these co-inhibitory molecules regulate signaling networks associated with T cell activation in a manner that can influence expression of transcription factors which dictate functional differentiation of T cells.

Differentiation of Tc17 cells and many other CD8 T cell subsets have been extensively studied in the context of chronic viral infection (Paley et al. 2012; Intlekofer et al. 2008; Wherry and John Wherry 2011). Despite the differences between this setting and that of tumor progression, there are many transcriptional similarities between T cell states of these settings that have formed the basis for current understandings of how T cell differentiation is impaired in cancer (Wherry, John Wherry, and Kurachi 2015). Importantly, it has been shown that CD8 T cells are responsible for controlling chronic viral infections, and abnormalities in CD8 T cell differentiation can result in collapse of immune responses (Paley et al. 2012). This is thought to occur invariably in clinically

detectable tumors; as such, many more recent studies have made comparisons between chronic viral infection and tumor-mediated T cell dysfunction in order to identify distinguishing features of T cell responses to tumors that may be responsible for immune suppression and evasion in cancer (Paley et al. 2012; Z. Zhang et al. 2020). Functional differentiation of T cells has been specifically implicated as the primary mechanism of T cell dysfunction because it was found that healthy donor T cells can respond to specific antigens that do not generate responses from dysfunctional T cells (Strønen et al. 2016). In short, the primary mechanism through which T cells are thought to become dysfunctional in tumors occurs when progenitor CD8 T cell populations undergo abnormal functional differentiation.

2.4 A Brief History of Immunotherapy

One of the earliest described treatments to invigorate immune cell responses to cancer was administration of Coley toxin, which contained mixtures of live and inactivated pathogens that had been previously associated with spontaneous tumor regression in patients (Dobosz and Dzieciątkowski 2019; Decker et al. 2017; Coley 2014; McCarthy 2006). However, skepticism regarding the risk and mechanism associated with Coley toxin, compounded by skepticism of the scientist who developed it, caused these discoveries to become dormant for many years (Dobosz and Dzieciątkowski 2019).

Later, other forms of immunotherapy were developed such as therapeutic utilization of monoclonal antibodies specific for genes that are upregulated in cancer cells. This revolutionized treatment of immune cell malignancies in particular. Eventually, this therapeutic strategy was also extended to malignancies that did not

arise from immune cells, most notably with the development of trastuzumab, which targets a receptor that is upregulated on the surface of breast cancer cells, ERBB2 (HER2) (Dean and Kane 2015). Still, even in cancers which respond well to monoclonal antibody therapy, many tumors become recalcitrant to these therapies. These therapeutic limitations drove development of other forms of immunotherapy.

As dysfunction of immune cells became more widely implicated in cancer progression, adoptive T cell therapies were developed which sought to address deficiencies in endogenous immune responses to cancer by providing functional immune cells that were produced or activated through artificial means (June 2007). Some of these therapeutic strategies specifically targeted antigens which are present across many different patients (van der Bruggen et al. 1991), and clinical indications for these therapies have become increasingly more refined as mutation status in tumors has become more prevalent in molecular diagnostics of cancer (Rosenberg and Restifo 2015). In addition, there have been many efforts to produce therapeutic vaccines to facilitate activation and functional differentiation of productive tumor-specific immune cell subsets (Bowen et al. 2018). Some of these vaccines are developed in a highly personalized manner using material that is isolated from individual patients (Shemesh et al. 2021; Timmerman and Levy 2000), while others seek to target highly recurrent tumor neoantigens (Xu et al. 2014).

One of the biggest limitations in the development of these therapies has been the associated costs (June 2007; Bowen et al. 2018), which for many patients is prohibitive. Furthermore, in many settings it has been difficult to identify neoantigens which can elicit responses without prohibitive toxicity in patients, and some responses to

immunogenic epitopes can cause auto-immune reactions in healthy tissues (J. C. Yang 2015). Neoantigens for vaccination can be selected by their binding affinity for MHC molecules (McMahan et al. 2006), but this approach largely relies upon computational predictions of peptide-MHC binding affinities, which in many cases can be unreliable (Phloyphisut et al. 2019). In addition, the properties that constitute immunogenicity of an epitope are poorly understood (McGranahan and Swanton 2019).

2.5 Immune Checkpoint Blockade

Many other efforts to develop immunotherapies for cancers were based upon the idea that functional differentiation of T cells could be altered to boost productive anti-cancer immune responses through therapeutic intervention. The earliest versions of these therapies attempted to do this through administration of systemic cytokines that were otherwise implicated in T cell differentiation as ‘immune checkpoints’, but they were highly toxic and largely ineffective (Robert 2020). A revolutionizing breakthrough was later made with monoclonal antibodies that could be used to block interactions of co-inhibitory receptors on T cells and ultimately divert functional T cell differentiation to generate durable and essentially curative responses in some cancer patients (Ledford, Else, and Warren 2018).

To date, the most successful checkpoint blockade therapies are monoclonal antibodies targeting CTLA-4, PD-1 or its ligand PD-L1 and they have been shown to reduce dysfunction in cytotoxic CD8 T cells (Kumagai et al. 2020). While these have revolutionary benefits in some patients, many patients have transient responses or do not respond at all, even when receiving a combination of these treatments (Kumagai et al. 2020). In addition, many questions remain about the mechanism of action that

underlies these differentiation outcomes. It is reasonable to expect that improving the understanding of these processes may lead to improvements in therapies that can create more durable responses and benefit a wider range of patients.

2.5.1 Predicting Response to Checkpoint Blockade

Many efforts have been made to identify clinically observable measurements that are predictive of response to checkpoint blockade therapy (Carbone et al. 2017; Rizvi et al. 2015) in order to better identify candidates that will respond well to it. Some initial studies showed that expression of PD-1 on tumor cells was partially, but not fully, predictive of checkpoint response (Carbone et al. 2017). For many years, it was thought that the relationships between tumor mutational burden and the subsequent diversity of the tumor-specific neoantigen repertoire could be used to predict patient response to immunotherapy. In light of recent evidence suggesting tumor mutational burden fails to predict patient response to immunotherapy (McGrail et al. 2021), the relevance of mutational burdens in tumors in predicting immune responses have become controversial (Strickler, Hanks, and Khasraw 2021).

In parallel, more observable features of tumors that are predictive of checkpoint response are being identified, which include dysregulation of antigen presentation machinery (Montesion et al. 2021), gene expression programs (Z. Wang, Li, and Xu 2021), metabolic activity (van Wilpe et al. 2021), inflammation (Kauffmann-Guerrero et al. 2021), and more. However, in nearly all cases, the predictive power of these features are limited. Further studies to identify what specific changes occur in response to checkpoint blockade may illuminate other features of T cells and tumors that predict productive responses to checkpoint blockade.

2.5.2 CD8 T Cell Progenitors are the Target of Immune Checkpoint Blockade Therapy

Currently, checkpoint blockade therapy is thought to act by altering differentiation of CD8 T cells (B. C. Miller et al. 2019a). A subset of progenitor CD8 T cells express moderate levels of PD-1 and other co-inhibitory markers that are then upregulated to high levels of expression as cells differentiate, termed ‘progenitor exhausted’ and ‘terminally exhausted’ CD8 T cells, respectively (Utzschneider et al. 2016; Im et al. 2016). Because these co-inhibitory markers are expressed in progenitor subsets and their respective heterogeneous functionally differentiated subsets, it is still unclear which CD8 populations mediate response to checkpoint blockade, although progenitor CD8 T cells which express TCF1 expand in response to checkpoint blockade (B. C. Miller et al. 2019a); however, this has been further complicated by emerging evidence for multiple TCF1+ progenitor CD8 T cell subsets (Beltra et al. 2020).

CHAPTER 1, PART 2

Non-Small Cell Lung Cancer

1. Clinical Characteristics of Non-Small Cell Lung Cancer

1.1 Epidemiology

Lung cancer is the leading cause of cancer related deaths amongst men and women in both the United States and abroad (Duma, Santana-Davila, and Molina 2019; Fitzmaurice et al. 2015). Approximately 85-90% of these cases are thought to be related to primary or secondary exposure to cigarette smoke (Alberg and Samet 2003), which has prompted lung cancer screenings amongst smokers (Moyer and U.S. Preventive Services Task Force 2014). Still, there are many cases of lung cancer that are not thought to be associated with smoking and do not carry its associated mutational signature (Alexandrov et al. 2016).

Non-small cell lung cancer (NSCLC) accounts for 85% of all lung cancer cases and is a broad classification that encompasses adenocarcinomas, squamous cell carcinoma, and large cell carcinomas (Duma, Santana-Davila, and Molina 2019). In patients with European ancestry, NSCLC is most prevalent amongst smoking men. In contrast, in patients with East Asian ancestry, NSCLC is most common amongst nonsmoking women (Jain et al. 2015). Notably, however, lung cancer incidence and mortality is the lowest amongst Asian-American, Pacific Islander, and Hispanic women, but is highest amongst Black men in the US (Siegel, Miller, and Jemal 2018). Generally, the earlier NSCLC is diagnosed, the better a patient's prognosis (**Table 1**). Overall, the

average 5-year survival rate of patients diagnosed with NSCLC is 25% for NSCLC and is 19% for all lung cancers (National Comprehensive Cancer Network 2021).

Table 1. Diagnostic frequency and survival of NSCLC across different disease stages. Adapted from (National Comprehensive Cancer Network 2021).

Disease State	Diagnostic Frequency (%)	5 year survival (%)
Localized Disease	19%	61.4%
Regionally Invasive Disease	24%	34.5%
Metastatic	55%	6.1%
Unknown	2%	14.6%

According to the World Health Organization, Lung adenocarcinoma is the predominant form of NSCLC and accounts for approximately 40% of all NSCLC cases (Travis, Brambilla, Burke, et al. 2015). The work that is presented in this thesis exclusively studies lung adenocarcinoma, but most clinical guidelines and observations for NSCLC, briefly discussed below, are not necessarily specific for lung adenocarcinoma and are generalized across NSCLC. It should be noted, however, that different subtypes of NSCLC have distinct survival rates (Reck et al. 2019; Ramalingam et al. 2020; Pacheco et al. 2019; Shaw et al. 2019).

1.2 Disease Characteristics and Staging

NSCLC staging follows TNM guidelines set by American Joint Committee on Cancer and considers the primary tumor (T), involvement of lymph nodes (N), and metastasis (M) (Lancia, Merizzoli, and Filippi 2019), wherein each disease stage has a characteristic patient prognosis that decreases with disease progression. NSCLC is

considered to be relatively early stage when patients are diagnosed with stage II disease and lower. When tumors have invaded the mediastinal lymph node, which is the closest lymph node to the lung, patients are considered stage III.

Histological features of patient biopsies are used to further classify lung tumors according to guidelines established by the World Health Organization in 2015 (Travis, Brambilla, Nicholson, et al. 2015). In general, poor histological differentiation is correlated with higher overall risk of death (National Comprehensive Cancer Network 2021; Travis, Brambilla, Nicholson, et al. 2015). Amongst adenocarcinomas alone, many histological subtypes exist that are distinguished from one another by morphological patterns. Adenocarcinomas are frequently distinguished from other subtypes of lung cancer by positive staining for TTF-1, known in research settings as NKX2-1 (Travis, Brambilla, Nicholson, et al. 2015). Approximately 70-90% of lung adenocarcinomas stain positive for NKX2-1 (National Comprehensive Cancer Network 2021).

In patients that present with advanced lung adenocarcinomas, molecular subtyping is frequently utilized to identify therapeutically actionable mutations that can inform therapeutic strategies (National Comprehensive Cancer Network 2021). The majority of these mutations have been previously identified as oncogenic driver mutations in NSCLC (Martínez-Jiménez et al. 2020) (**Table 2**). Of note, the mutational frequency of oncogenic drivers is most often reported from data collected from patients of European descent. However, the prevalence of oncogenic driver mutations may differ in patients with different geographic locations, ethnic backgrounds, age of diagnosis, and smoking status. For example, in one study of patients less than 20 years of age

diagnosed with lung adenocarcinoma, EGFR mutations were more common amongst East Asian patients (Jain et al. 2015; Shi et al. 2014) and less common in patients from India (Shi et al. 2014). In this study, EGFR mutation frequency was found to be associated with ethnic group and smoking status (Shi et al. 2014), and Alk rearrangements may be more prevalent in nonsmoking Chinese patients (W. Luo et al. 2018). Also, **Table 2** does not include mutations in other genes with meaningful roles in lung adenocarcinoma progression, such as TP53, PIK3CA, MET, KEAP1, NF1, RB1, CDKN2A, ARID1A, SMARCA4, RBM10, U2AF1, MGA, and MYC (Cancer Genome Atlas Research Network 2014), whose status is not currently evaluated in diagnostic settings.

Table 2. Therapeutically Actionable Mutations in NSCLC. Biomarkers considered in clinical evaluation of NSCLC, as reported by the National Comprehensive Cancer Network (NCCN). Adopted from (1) (National Comprehensive Cancer Network 2021) ;(2) (Cancer Genome Atlas Research Network 2014; C. Wang et al. 2018); (4) (Kwak et al. 2010); (5) (Chuang et al. 2017) (6) (Hong et al. 2020)

Biomarker	**Biomarker Type[1]**	**Frequency**	**Preferred Therapies[1]**
KRAS	Activating Point Mutation at Codon 12, 13, and 60	~8-33% of adenocarcinomas[2,3]	Sotorasib[6]
EGFR	Exon 19 deletion and Insertion	~14-52% of adenocarcinomas[2,3]	Erlotinib Gefitinib Afatinib Osimertinib dacomitinib
EGFR	Activating Point Mutation in Exon 21		
ALK	Rearrangement / Fusions	2-7% of NSCLC cases[4]	Alectinib Brigatinib Crizotinib ceritinib

ROS1	Rearrangement / Fusions	~ 1-2% of NSCLC cases[2]	Crizotinib Ceritinib Enterectinib
BRAF	V600E Activating Point Mutation	~10% of NSCLC cases[2]	Dabrafenib + trametinib
MET	Exon 14 Skipping	7% of lung adenocarcinomas[2]	Crizotinib Capmatinib
NTRK	Fusions	0.2% of NSCLC[1]	Larotrectinib Entrectinib
ERBB2	Amplification	1-2% of lung adenocarcinomas[5]	Afatinib Trastuzumab
RET	Fusions	1-2% of NSCLCs	Cabozantinib Pralsetinib Selpercatinib Vandetanib

The two oncogenic driver mutations most frequently observed in lung adenocarcinoma are the epidermal growth factor receptor (EGFR), which is often treated with tyrosine kinase inhibitors, and KRAS, which currently only has one form of targeted therapy that targets a specific point mutation (G12C) in KRAS (Hong et al. 2020). In general, patients with KRAS mutations have a poorer prognosis than those without (National Comprehensive Cancer Network 2021). KRAS mutational status can be determined by sequencing, high-resolution melting analysis, PCR, allele-specific hybridization, RT-PCR and can be assessed at codon 12, 13, or 61 (Cagle et al. 2014). Overall, many oncogenic driver mutations, including EGFR and KRAS, occur in a mutually exclusive manner (Cancer Genome Atlas Research Network 2014; C. Wang et al. 2018).

Today, biomarkers that inform therapeutic strategies encompass more than genetic mutations. Expression of PD-1 (Brody et al. 2017) can be used to predict and is a clinical indication for response to immune checkpoint blockade. Some studies have suggested that PD-1 status does not correlate with a particular oncogenic driver mutation (Brody et al. 2017), when looking specifically at East Asian patients, it was found that a particular mutational signature (MS3) with a high frequency of EGFR mutations was associated with B lymphocyte infiltration (C. Wang et al. 2018) and observed most frequently in female Chinese never-smoker patients. Discoveries such as this highlight the importance of precise epidemiological stratification in molecular profiling of lung adenocarcinomas, which may uncover other mutational signatures associated with immune infiltration and tumor-specific immune response.

1.3 Treatment

In general, NSCLC is treated through different permutations of chemotherapies, radiotherapy approaches, surgical resection, targeted therapy, image-guided thermal ablation therapy, and immunotherapy (National Comprehensive Cancer Network 2021). Today, standard-of-care for NSCLC patients that present with different subtypes, stages of disease progression, and pathologies is highly stratified. The guidelines for cancer treatment are continuously updated and re-evaluated to optimize patient outcomes.

Most patients with NSCLC are not symptomatic until tumors are extensively progressed. In patients that are diagnosed with earlier stages of disease progression, which is approximately 19% of patients, surgical resection of lung tissue that contains the primary tumor is preferred (National Comprehensive Cancer Network 2021). In settings of more advanced disease, the clinical benefits of surgical resection are

controversial (Duma, Santana-Davila, and Molina 2019; Martins et al. 2012). Another method of local treatment for tumors is image-guided thermal ablation therapy, which can be performed through radiofrequency ablation, microwave ablation, and cryoablation (National Comprehensive Cancer Network 2021). It is often performed in patients who, for a variety of reasons, are not good candidates for surgical resection but is considered to be a local therapy that is often only indicated for patients presenting with earlier stages of disease progression.

In contrast, radiotherapy is clinically indicated for many patients across all stages of NSCLC and is most often used to maintain control over tumor growth in a manner that minimizes associated toxicity (National Comprehensive Cancer Network 2021). Radiotherapy can be administered preoperatively, postoperatively, alone, or in combination with chemotherapy or targeted therapy. Similarly, platinum-based chemotherapy is utilized across stages of NSCLC under specific clinical indications (National Comprehensive Cancer Network 2021). It is often delivered as a secondary form of treatment, either before the primary tumor is surgically removed or in conjunction with chemotherapy. Cisplatin and carboplatin are frequently used in combination with other chemotherapies (National Comprehensive Cancer Network 2021). Patients with advanced disease that initially respond well to chemotherapy are also eligible to receive anti-PD-1 checkpoint therapy, durvalumab, concurrently with chemoradiation (Antonia et al. 2018)

In the absence of therapeutically actionable mutations, patients positive for PD-1 will typically receive chemotherapy alongside single-agent or dual-agent checkpoint blockade therapy. Some forms of chemotherapy are thought to be immunogenic by

creating secondary mutations in response to therapeutic DNA damage, such as cyclophosphamide, and are believed to cause cell death through an immunogenic cell death pathway that can then elicit a productive immune response (J. Wu and Waxman 2018). The impact of immunogenic chemotherapy has yet to be re-evaluated inlight of recent evidence that has questioned how reliably predictive tumor mutational burden is for response to immune checkpoint blockade (McGrail et al. 2021).

2. Lung Development

2.1 Physiology

The distal lung epithelium is the primary site of gas exchange in the lung, and is composed of alveolar air sacs that contain epithelial cells with distinct functional roles (Rawlins et al. 2009). Alveolar type I (AT1) cells, otherwise known as Type I pneumocytes, are terminally differentiated epithelial cells that mediate gas and ion exchange with characteristically high surface area (M. C. Williams 2003). Alveolar type II (AT2) cells, also known as Type II pneumocytes, are responsible for production of surfactant proteins. Surfactants are composed of a combination of lipids and phospholipids, which reduce surface tension, as well as surfactant apoproteins, which facilitate molecule adherence (adsorption). In the lung alveolus, this mixture acts to prevent lung collapse during periodic biophysical changes associated with breathing (Ingenito et al. 1999).

AT1 cells make up the majority of the lung alveolus (Mason 2006), and as a result, are the primary target of lung tissue damage. During development, both AT1 and AT2 cells are thought to differentiate from common progenitors (D. B. Frank et al. 2019). After development, AT2 cells, which comprise only a minority of the lung alveolus,

mediate regenerative stability of the lung alveolus and replenish AT1 cells (Mason 2006; Evans et al. 1975), particularly in the context of lung injury. For this reason, AT2 cells are believed to be the primary source of AT1 cells (Desai, Brownfield, and Krasnow 2014) in adult animals. More recently, a transcriptionally distinct stem cells that exist between the bronchus and alveoli, termed bronchioalveolar stem cells (BASCs), have also been described and are believed to be responsible for regeneration of AT2 and club cells (Salwig et al. 2019). Specifically, AT2 cells have been described as the primary source of AT1 cells for homeostatic regeneration, whereas BASC cells are thought to contribute to distal airway renewal following severe lung injury, but also have a minor contribution to homeostatic renewal (Salwig et al. 2019).

2.2 Anatomy

To generate the complex branching pattern of the lung, development of lung airways occurs through three dimensional branching processes (Metzger et al. 2008). The bronchial tree is formed through initial branching events from the bronchus (primary branching) followed by repeated, smaller branching events (secondary branching events) that extend the airways to the surrounding mesenchyme (Warburton et al. 2010).

Multiple patterning events govern three dimensional branch formation that manifest through three branching modes. The first, domain branching, generates branches that are arranged as rows and arise as a result of signaling circuitry that dictate branching periodicity and domain positioning (Metzger et al. 2008). The two remaining branching modes are distinguished from domain branching because branching occurs at the tip. Of these two modes, planar bifurcation occurs through

sequential branching of lung buds that are formed after primary branching events (Metzger et al. 2008). This process creates proximal-distal patterning of the lung bud and serves to create branches that exist in the same plane. Similarly, orthogonal bifurcation occurs towards the end of branching processes but generates branches that are normal (perpendicular) to one another, and is largely responsible for generating three-dimensional branching patterns (Metzger et al. 2008).

2.3 Genetics of development

The lung airway is composed of multiple tissue types, most notably epithelial cells, which are responsible for gas exchange and surfactant production, as well as endothelial cells which form the vasculature required for gas exchange (Cardoso and Whitsett 2008). Other cell types of the lung, such as neuroendocrine cells, have distinct developmental origins from those of epithelial cells (Perl et al. 2002). However, most cell types of the lung airway develop from common progenitors during development (Perl et al. 2002). Notably, the final steps of lung maturation, during which surfactant production begins, occurs very late in embryonic development and continues postnatally (Warburton et al. 2010).

During embryonic development, lung tissue is specified during differentiation of the endoderm. Co-expression of *Nkx2-1*, a transcription factor that is otherwise known as *Ttf-1* or *Titf-1*, and the forkhead transcription factor, *Foxa2* (Maeda, Davé, and Whitsett 2007) distinguishes lung tissue from other fates of the gut tube, which include thyroid, liver, and pancreas. Importantly, both of these transcription factors have extensive interactions with other transcription factors that mediate progressive differentiation of different tissue types of the lung and also play important roles in

differentiation of other gut tube derived tissues (C. M.-C. Li et al. 2015; Warren et al. 2020; C. S. Lee et al. 2005).

Differentiation of tissues that make up the lung airway are highly dependent on proximal-distal patterning of the developing lung. When these patterns are established, cells fated to become part of the airway proliferate in an undifferentiated state in response to Wnt induced effectors (Okubo et al. 2005; Shu et al. 2005). During tissue patterning and branching, processes that lead to differentiation of cells fated to become various cell types of lung alveoli are highly interdependent. Together, these interactions orchestrate a network of signaling pathways and transcription factors (Maeda, Davé, and Whitsett 2007) which acts across multiple tissues that dictate their development.

In particular, Fgf10 is a morphogen expressed in the mesenchyme adjacent to the differentiating endoderm that directs lung bud formation, outgrowth, and proliferation (Bellusci et al. 1997; Abler, Mansour, and Sun 2009). It primarily mediates signaling through activation of the fibroblast growth factor receptor 2, Fgfr2 (Abler, Mansour, and Sun 2009). Fgf10 expression is largely regulated through interactions with surrounding tissues; in the mesenchyme, for example, it is positively regulated by the growing distal endoderm (Bellusci et al. 1997). Additionally, Fgf10 signaling induces expression of genes and activity of signaling pathways that serve to restrict and dampen further Fgf signaling. This feedback is of particular importance because branching events are thought to occur in the absence of Fgf10 (Abler, Mansour, and Sun 2009). These feedback signals include Sonic Hedgehog (Shh) signaling, bone morphogenic proteins (BMPs) (Hyatt, Shangguan, and Shannon 2004), and Sprouty expression (Mailleux et al. 2001).

Negative feedback for Fgf10 signaling occurs through multiple mechanisms. In the case of Sprouty proteins, which act intracellularly to inhibit FGFR kinase activity, they negatively regulate Ras signaling downstream of FGF signaling (Perl et al. 2003). Mechanistically, they can initiate and regulate branching induced by Fgf10 (Tefft et al. 1999). In contrast, Shh signaling primarily affects the surrounding mesenchyme of the developing lung bud rather than the lung bud itself (L.-A. D. Miller et al. 2004) despite its requirement for proper lung branching and patterning (Pepicelli, Lewis, and McMahon 1998). Shh signaling is canonically activated when a Shh ligand binds and inactivates its receptor, Patched, which ultimately results in translocation of effector Gli transcription factors to the nucleus (Carballo et al. 2018). When overexpressed in the developing lung, Fgf10 becomes downregulated (Bellusci et al. 1997). Further mechanistic studies have demonstrated that Shh acts to restrict Fgf10 expression to the distal end of the lung bud (Kugler et al. 2015).

While Bmp4 has consistently been shown to be a critical regulator of lung development, the precise nature of its behavior has been less clear. In genetic studies, manipulation of Bmp4 causes disruption of lung branching events; overexpression of Bmp4 reduces the number of terminal buds (Lu et al. 2001), while overexpression of a BMP inhibitor causes defects in distal differentiation (Weaver et al. 1999). Together, it is thought that the effects of Bmp4 seem to vary as a function of its effective concentration (Weaver, Dunn, and Hogan 2000; Weaver et al. 1999) and are dependent on interactions with the surrounding mesenchyme (Warburton et al. 2005). Further, Bmp4 expression is tightly regulated by multiple signaling pathways, including Wnt and Fgf (Shu et al. 2005).

These genes and pathways form a genetic circuit that ultimately coordinates branching processes of the distal lung. (Maeda, Davé, and Whitsett 2007) Importantly, because cell type differentiation is exquisitely linked to patterning, this genetic circuitry not only influences lung anatomy, but also differentiation. It should be noted that postnatal development of the lung also occurs and is distinct from prenatal development. For example, RAS expression in the lung occurs mostly in postnatal development (Thrane et al. 1997). This expression pattern is not surprising, given that lung development occurs late during fetal development and continues through the first few days after birth (Warburton et al. 2010).

2.4 Lung Development and Non-Small Cell Lung Cancer

In terms of histopathology, the differentiation state of a tumor has long been recognized to be prognostic of tumor behavior and patient survival (Jögi et al. 2012). Loss of differentiation status intuitively correlates with patient prognosis; the more differentiated a tumor is, the more it resembles the tissue that it arises from. Mechanistically, the relationship between differentiation status and hallmarks of tumor progression has been less clear. The discovery and characterization of the EMT program (Brabletz et al. 2018) in the context of carcinomas offered clarity on this relationship, as EMT was specifically implicated in metastasis. However, loss of differentiation in carcinomas that arise from different tissue types occurs through tissue specific mechanisms that are poorly understood. The importance of the developmental context under which a tumor forms is made evident through comparison of how tumors with different origins behave and respond to therapy. Notably, these associations can be complicated by the fact that histopathology is not always indicative of the tissue a tumor

arises from (Visvader 2011). Overall, the relationship between differentiation status and disease behavior is highly context-specific.

Differentiation status is of particular importance for non-small cell lung cancer. One of the features considered when determining the histological grade of a tumor, which robustly stratifies patient survival, is histological differentiation of the tumor (Yasukawa et al. 2018). Aggressive, high grade tumors are characterized by lack of differentiation. Expression of NKX2-1, a master regulator of lung identity, is often assessed in clinical specimens to identify tumors that originate from the lung. In agreement with the associations made with differentiation status and patient prognosis, expression of NKX2-1 is associated with improved prognosis relative to patients with low expression of NKX2-1 (Moisés et al. 2017). Despite this association, NKX2-1 is frequently amplified in lung cancer (Kwei et al. 2008) and is described in some contexts as a lineage-specific oncogene. However, mechanistic studies have suggested that NKX2-1 suppresses metastatic disease progression (Winslow et al. 2011).

3. Genetically Engineered Mouse Model of Lung Adenocarcinoma

3.1 The KP Model

3.1.1 Kras

The RAS family, which is the most frequently mutated family of genes in cancer, includes HRAS, KRAS, and NRAS. Of the RAS family, KRAS is mutated the most frequently (Waters and Der 2018). WHen bound to GTP, RAS proteins activate a variety of intracellular signaling networks that generally promote cell proliferation (Cox and Der 2010). Point mutations at glycine-12, glycine-13, or glutamine-61 have been implicated as gain-of-function mutations which render KRAS constitutively active in the KRAS-GTP

state, and thereby drive uncontrolled proliferation and transformation (Tchernitsa et al. 2004). The oncogenic effects of point mutant $KRAS^{G12D}$ have been extensively modeled in the mice of lungs, and it was found that spontaneous expression of $Kras^{G12D}$ will invariably lead to the development of lung tumors (Johnson et al. 2001). Histopathologies of the resultant lesions of the lung recapitulate human NSCLC progression, from hyperplasia to carcinoma (Johnson et al. 2001), and as such, experimental introduction of oncogenic $KRAS^{G12D}$ in the lung has become a widespread genetically engineered mouse model of cancer.

Oncogenic $Kras^{G12D}$ has been engineered into an allele that allows expression of oncogenic $KRAS^{G12D}$ following removal of a transcriptional stop element in a Cre-recombinase dependent manner (E. L. Jackson et al. 2001). Lung specific expression of Cre recombinase can be achieved through intranasal or intratracheal administration of an adenovirus containing Cre (E. L. Jackson et al. 2001; DuPage, Dooley, and Jacks 2009) and results in multifocal tumors of the lung. *Kras* is required for embryonic development but is haplosufficient, thus, this allele can only be bred heterozygously into viable animals. Importantly, these tumors do not progress through late histological stages of NSCLC progression, which led to incorporation of other Cre conditional oncogenic alleles.

3.1.2 Trp53

The tumor suppressor P53 is one of the most extensively studied and characterized genes in cancer (Vousden and Lane 2007). It has been an identified regulator of many systems within a cell that become dysregulated in tumor cells, including cell death, proliferation, genomic stability, and senescence (Vousden and Lane

2007). The mechanisms behind P53 functionality occur both through its activity as a transcription factor and its direct signaling pathway activity (Yu and Zhang 2003). It is regulated through modulation of its activity, stability, and localization in response to cell cycle arrest and apoptotic signals (Vousden and Lane 2007). A conditional null allele of TP53 was generated through insertion of LoxP sites, which become recombined by Cre recombinase, in locations flanking exons 2 through 10 of TP53.

In clinical cases of NSCLC, KRAS and TP53 mutations frequently occur together, and as such, these Cre-conditional *Kras* and *Trp53* alleles were combined. This produced *Kras*$^{LSL\text{-}G12D/+}$;*Trp53*$^{fl/fl}$ mice that produce tumors similar to those of *Kras*$^{LSL\text{-}G12D}$ alone (K only) but better reflected more progressed histological stages of NSCLC with reduced tumor latencies (Jackson et al. 2005). This model has been well characterized and is hereafter referred to as the KP model.

3.1.1 Behavior of the KP model

Both the K only and KP model of lung cancer have variable progression kinetics across mice that are infected with the same viral dose (DuPage, Dooley, and Jacks 2009). This has been observed with respect to histological tumor progression, metastasis, and survival of KP mice (DuPage, Dooley, and Jacks 2009), and with multiple forms of viral delivery. For example, at 6 weeks post-infection with adenoviral Cre, approximately 50% of tumors are grade 1, 40% of tumors are grade 2, ~17% of cells grade 3, and the remaining fraction grade 4. 20 weeks later, almost half of tumors are grade 3, more than 20% grade 4 or grade 2, and a small minority of grade 1 or grade 5 tumors (Jackson et al. 2005). Interestingly, KP tumor progression is not

currently believed to be driven by acquisition of additional mutations (DuPage et al. 2011).

While lentiviruses used to generate tumors can be flexibly generated in-house and have viral genomes that can be experimentally modified with ease, they infect a wide variety of cell types, including macrophages (Buckley et al. 2008). Importantly, there is no evidence that *Kras* and *Tp53* alleles cause transformation of tissues except those of the lung epithelium. While tissue specific promoters can be used to drive expression of Cre recombinase, many of these promoters are too large to be encapsulated by lentiviral packaging proteins. Instead, Cre expression in alveolar type II cells can be achieved by utilizing the promoter for surfactant protein C (*Sftpc*) to drive expression of Cre recombinase with adenoviruses (DuPage, Dooley, and Jacks 2009; Tippimanchai et al. 2018).

3.1.2 Cell of origin in the KP model

Multiple epithelial cells of the lung can give rise to KP lung adenocarcinomas. KP lung adenocarcinomas can be generated with viruses that contain Cre driven by promoters that are specific for multiple cell types of the lung, including AT2 cells and Clara cells (Sutherland et al. 2014). Although both tissues can give rise to adenocarcinomas, the tumors differ in kinetics and expression (Sutherland et al. 2014). Some rare cells of the lung that express Clara Cell markers and AT2 markers simultaneously, termed Bronchioalveolar Stem Cells (BASCs), can give rise to *Kras* driven tumors as well (C. F. B. Kim et al. 2005). However, it is now appreciated that tumors that arise from these cell types differ from those of AT2-derived tumors (C. F. B. Kim et al. 2005). The tumors that will be described in the forthcoming sections will focus

on AT2-derived tumors, which are believed to be the predominant tumors that result from intratracheal lentiviral Cre.

3.2 Developments in experimental tools for studying the KP model

The flexibility afforded by lentiviral mediated delivery of Cre recombinase permits integration of sequences which, beyond controlling expression of Cre recombinase, can mediate alterations in other genes or drive ectopic expression of a transgene. In particular, incorporation of CRISPR machinery to create double stranded DNA breaks at target loci in the genome has made introduction of additional mutations at the onset of tumor initiation very efficient (Sánchez-Rivera et al. 2014). Additionally, lentiviruses can be incredibly powerful tools to study the immune response to a strong neoantigen by inducing ectopic antigen expression in a tumor-specific fashion (DuPage et al. 2011). By expressing LucOS, a fusion of luciferase to two peptides, T cells that recognize the antigens of LucOS can be isolated and characterized via tetramer staining (DuPage et al. 2011). By staining for T cells that specifically recognize LucOS antigens, longitudinal responses to a tumor-specific antigen can be measured.

Furthermore, through introduction of additional Cre responsive alleles to the *Rosa26* locus of KP mice, even further flexibility is introduced. This is particularly advantageous in settings where lentiviruses are not preferred, and allows for expression of Cas9, fluorophores, and more (Ng et al. 2020).

CHAPTER 1, PART 3

High-Dimensional Biological Data

1. Technological Advancements in Sequencing Technology

The development of next-generation sequencing technologies, which have made DNA sequencing increasingly more cost and labor efficient, has undoubtedly changed the nature of biological research. Widespread utilization of this technology has led to unprecedented production of biological data at a scale that is difficult to make meaningful conclusions from, in even a single biological setting. Until recently, the materials sequenced through these technologies were isolated from pools of cells. Advancements in droplet-based, microfluidic, and barcoding technologies have now made it possible to sequence material isolated from individual cells (Zheng et al. 2017; Macosko et al. 2015; Amini et al. 2014).

As technologies have advanced, it has become more efficient to sequence a larger number of cells in a manner that produces higher quality data at a lower per-cell cost. Subsequently, large-scale efforts have sponsored utilization of single-cell technologies to characterize cells of different tissues, in various biological contexts, across many organisms (Rozenblatt-Rosen et al. 2017; Regev et al. 2017, 2018; Consortium et al. 2018; Z.-J. Cao et al. 2020; Packer et al. 2019). These data generated from these experiments have been aggregated and made publicly available in a way that has brought biological research into a new era, wherein biological processes are characterized and studied at single-cell resolution.

Single-cell technologies have allowed measurement of RNA and protein expression in individual cells. single-cell DNA sequencing has also become more prevalent; however, it is difficult to perform this technique in a manner that yields sufficient coverage and resolution of a cell's genome. As such, many of the more widely adopted single-cell technologies provide better relative coverage of sequenced material through targeted capture of RNA and DNA (Yong Wang and Navin 2015). While some single-cell technologies designed to assay protein expression utilize DNA sequencing based readouts, all are dependent on, and limited by, the availability of validated antibodies (Stoeckius et al. 2017; Han et al. 2018).

Single-cell RNA sequencing (scRNA-seq) is a widely utilized single-cell-omic technology, primarily due to widespread adoption of the 10X Genomics platform (Zheng et al. 2017), which allows for efficient and simultaneous measurement of 3' and 5' mRNA in up to thousands of cells. It has now incorporated other technologies to allow for immune-cell related readouts, spatial measurement, and chromatin accessibility readout simultaneously with gene expression. Additionally, chromatin accessibility, which is measured through assay for transposase-accessible chromatin using sequencing (ATAC-seq), has also become increasingly more popular and can be performed with the 10X Platform or by single-cell combinatorial indexing (sciATAC-seq) (Cusanovich et al. 2015). Transcriptional and chromatin accessibility states serve as reliable proxies for overall cell state because they are collectively responsible for determining cell behavior. Thus, these technologies have now permitted generation of high-dimensional biological data that describes biological state at single-cell resolution.

2. Structures of High-Dimensional Biological Data

2.1 Advancements in discovery of complex data structures

When biological data is produced from assays with single-cell resolution, each cell is considered a “variable” and each entity measured (e.g. each mRNA in scRNA-seq) is considered an “observation”. Each variable is measured for each observation or cell, and this produces high-dimensional biological data. Analytically, many challenges exist when interpreting this kind of data, which are typically discussed as ‘the curse of dimensionality’ or the ‘dimensionality problem’ (Donoho and Others 2000). The curse of dimensionality refers to the inability to approximate the underlying structure of data that is created with complete accuracy. In other words, the data cannot be interpreted in its true form.

To overcome this challenge, implementations of multivariate statistics and probability theory have been utilized in computational algorithms to approximate data structures on a manifold, which is sometimes described as a topological space (Martin 2002). In single-cell analysis, projections of these manifolds are often made in low-dimensional (usually 2-dimensional) space to visualize latent data structures. In each approximation approach, assumptions are made or artifacts can be created which can lead to obstruction of true underlying structures and even misrepresent some relationships. The optimal analytic approach is usually determined on a case-by-case basis and is largely dependent on the assumptions made about underlying data structures (Luecken and Theis 2019). In addition, hyperparameter selection can have a big impact on the performance of various analytical approaches.

2.2 Relationships between data structures and underlying noise

Some single-cell analysis techniques require handling that can lead to substantial technical and batch related artifacts, which was particularly true for early forms of scRNA-seq (Xiliang Wang et al. 2021). In some cases, these effects can be strong enough to entirely obstruct underlying structures of data. In many instances, metadata for plate or batch origin is recorded for each cell, which can be used to approximate the extent to which batch identity can explain variation observed in data. This can be easily performed through comparison of expected and observed genes that show the most variation across a dataset (Xiliang Wang et al. 2021). When comparing datasets empirically known to contain functionally distinct cell types, where subtle differences in observations are expected, it can be difficult to distinguish batch effects from true underlying biology. This is compounded by variation produced from ‘drop-out’, which occurs when the utilized technology fails to capture expression of a gene and typically occurs most frequently amongst lowly expressed genes (Kharchenko, Silberstein, and Scadden 2014). Dropout is usually most apparent when two highly similar cells show moderate to high expression of a particular gene in one cell, but not of another. In single-cell data, dropout requires analytical approaches of single-cell data to be able to tolerate noise generated from resulting sparse data. This is the primary reason why batch correction techniques utilized in bulk RNA-seq analysis, which are not plagued by sparsity, have seldom been extended to the single-cell setting (H. T. N. Tran et al. 2020).

Batch correction of single-cell data can be approached through multiple ways. Some of the most popular forms of batch correction depend on construction of nearest

neighbor graphs (Lun 2019; Jialin Liu et al. 2020), while others 'align' data by drawing correlations between them (Hardoon, Szedmak, and Shawe-Taylor 2004). All of these algorithms depend on similarities that exist between cells across batches that are being corrected for. If similarity is expected between some, but not all, cells across different batches, this can lead to loss of true underlying biological variation. Some techniques explicitly state that at least one cell type must be shared amongst batches (H. T. N. Tran et al. 2020). Other important statistical assumptions are also made for some of these algorithms, particularly about the behavior of underlying variation in the data.

In many cases, it must be determined whether or not batch correction is appropriate and, if indicated, which algorithm is optimal. Some batch correction techniques also employ methodologies that can be extended to integrate high-dimensional data structures that are generated through different techniques or modalities (Jialin Liu et al. 2020), but batch correction is still primarily used in the setting of a singular kind of data (e.g. mRNA, chromatin accessibility). While some techniques tend to perform better than others across many forms of data, there is currently no "one-size-meets-all" solution for batch correction (H. T. N. Tran et al. 2020).

3. Dimensionality Reductions and Interpretability

3.1 Modern approaches to dimensionality reduction in biology

The general premise of dimensionality reduction in biological data is to effectively reduce the number of dimensions, or variables of gene expression, across which variation can be studied. Like batch correction algorithms, the techniques utilized in dimensionality reduction generally perform optimally in some, but not all, settings and each have associated statistical assumptions (Luecken and Theis 2019) which are not

appropriate for all biological settings. Most conventional approaches to single-cell analysis currently utilize multiple forms of dimensionality reduction to increase computational and processing speed, which are briefly outlined in the subsections below.

Importantly, construction of these transcriptional or epignomic phase spaces are independent from and do not explicitly consider external biological metadata such as time, genetic context, genotype, gender, etc, except when these covariates are empirically known and explicitly controlled for through batch correction. These identities create variation in biological data, and the extent of their influence has been demonstrated in settings where this metadata is explicitly recorded and defined (Peng et al. 2021). While this can pose challenges of interpretability, this can also be extremely useful. The identities that are not considered in your dimensionality reduction can be used as “ground truths” to benchmark the reliability of your analysis for your specific use case against empirically known relationships and patterns of variation.

3.1.1 Linear Dimensionality Reduction

Principal Component Analysis (PCA), entails linear transformation of data in a manner that maximizes differences between features of the data (Turk and Pentland 1991). Its applications for analysis of biological data have allowed biologists to make meaningful observations of relatively low dimensional data, such as those from bulk RNA sequencing experiments (Yeung and Ruzzo 2001; Shuangge Ma and Dai 2011). Identified principal components, which can be interpreted as describing different forms of variation in data, are evaluated by genes which vary the most across a given component. The use of PCA set a precedent in biology for application of mathematical

approaches for analysis of high-dimensional data, which became particularly advantageous for analysis of single-cell RNA sequencing data that produces datasets with dimensionality that can be multiple orders of magnitude larger than bulk RNA-seq datasets.

When applied to high-dimensional data, the limitations of PCA and similar linear transformations are more pronounced (Sun et al. 2019). In many cases, the first two principal components capture a fraction, but not all, of the variation in a dataset. Additionally, PCA does not preserve local structures of data (van der Maaten 2008). Such limitations have led analysis approaches to instead utilize algorithms that are designed to better preserve both the local and global relationships between data (van der Maaten 2008). Still, PCA is widely recognized as a reliable method for estimation of global data structures. It is often used to preprocess data prior to nonlinear dimensionality reduction (Zappia, Phipson, and Oshlack 2018) and is also useful for reducing computational resources required for downstream analysis (Tsuyuzaki et al. 2020). In addition, it can be performed only on genes that are highly variable across the data through feature selection that occurs prior to PCA (Xiliang Wang et al. 2021).

Dimensionality techniques that rely on linear transformation of data inherently assume that the underlying ‘subspace’ (or gene expression space) is linear. However, most biological data have nonlinear data structures (Y. Cheng and Newell 2016; Schulte-Schrepping et al. 2020) and, as a consequence, most linear dimensionality techniques perform poorly on high-dimensional data.

3.2.2 Nonlinear Dimensionality Reduction

Functionally, nonlinear dimensionality reductions create an intangible "space" that is unique to the form of data, i.e. topologies transcriptional space or chromatin accessibility space, through a generalized approach of manifold learning (X. Huang, Wu, and Ye 2019).

Nearest neighbor graphs generate data structures that identify positions for data points that are most similar to another given data point, and are often created to data that has been transformed into a Euclidean subspace (Andoni and Indyk 2006). Because PCA produces approximations of distance between data in Euclidean space, many algorithms construct nearest neighbor graphs in principal component space. Nearest neighbor graphs can be constructed in multidimensional space, often on a fraction of selected principal components produced during preprocessing (Tsuyuzaki et al. 2020) that collectively capture most of the variation observed across a dataset.

The t-stochastic neighbor embedding (tSNE) projection is one of the most widely utilized embeddings for single-cell analysis. It works by approximating nearest neighbor graphs through construction of probability distributions from Euclidean distances between data (van der Maaten 2008). It acts to reduce dimensionality by then finding an arrangement for data in lower (usually 2 or 3) dimensional space. Distances between data arranged in lower dimensional space can be used to generate probability distributions (specifically, student t-distributions) and compared to those generated in high-dimensional space.

This is then performed iteratively to find and evaluate an arrangement for the data in lower dimensional space that generates probabilities most similar to those

constructed on data in high-dimensional space. A given arrangement in lower dimensional space is considered to be optimal when Kullback-Leibler divergence across all datapoints is minimized. A tSNE projection will specifically preserve local distances between data points by prioritizing optimal arrangements between data points that are close together at the expense of those which are far apart when evaluating lower dimensional arrangements of data points. Notably, in datasets where most data points are similarly distanced from one another, this will create artifacts that will manifest as crowding of data points and distortion of underlying data structures. Additionally, this will cause poor performance in datasets with highly variable underlying manifolds (van der Maaten 2008).

Uniform Manifold Approximation and Projection (UMAP) has become a widely adopted method of visualizing high-dimensional data and is often a preferred alternative to tSNE embeddings (McInnes et al. 2018). Mechanistically, it is similar to that of tSNE, but takes different approaches to calculate neighbor graphs and approximate them in lower dimensional space (McInnes, Healy, and Melville 2018). When a UMAP dimensionality reduction is performed, a k-nearest-neighbor graph for each individual cell is constructed by approximating distance between neighboring cells with a Reimannian metric. The resulting graphs for all cells in the data are then iteratively merged or “patched” together to form a global space, represented by a k-nearest neighbor graph, that approximates the topology of the data on a uniform manifold. In a manner that is conceptually similar, but mechanistically distinct from tSNE embeddings, a low dimensional (2D or 3D) UMAP embedding is generated by finding an optimal arrangement for cells data in lower dimensional space that generates probabilities most

similar to those constructed on data in high-dimensional space. This differs from the approach utilized in tSNE embeddings in the features of the probability distributions that are compared and optimized upon. The cost function used to find an optimal 2 or 3 dimensional embedding, cross entropy, is also distinct from that of tSNE (McInnes, Healy, and Melville 2018). UMAP is favored in most circumstances over tSNE because it better approximates global and local structures of data (McInnes et al. 2018).

3.2 Limitations of dimensionality reductions

Although dimensionality reduction is a powerful and necessary tool in interpretation of high-dimensional data, it always comes at a cost: information must be lost (Donoho and Others 2000). Optimally, dimensionality reduction will result in preservation of meaningful forms of variation across data in low dimensional space. In most use cases, the exact nature of the variation in a dataset is not necessarily known ahead of time, and this can confound interpretations of dimensionality projections. Further, it can impact the reproducibility and robustness of data interpretation across parameters and analytical approaches. In many cases, multiple approaches to dimensionality are taken when performing an initial analysis of single-cell data.

Across published studies of scRNA-seq and other forms of single-cell analysis, there is no single broadly accepted approach to data analysis because selection of analytical methodologies is entirely dependent on the objectives of the study and the nature of the underlying structure of the high-dimensional data that is produced. In most experimental settings, this underlying cannot be accurately determined, so, the optimal methodology for dimensionality reduction cannot be selected in a straightforward or standardized manner (Luecken and Theis 2019). A low dimensional embedding is an

approximation, and the degree to which an embedding reliably depicts the variation in data that is of interest is dependent on the use case. For example, distinguishing one cell type from another may be more easily or robustly achieved than distinguishing highly similar cells with nuanced differences. Ultimately, multiple approaches are frequently utilized to determine an optional dimensionality reduction and embedding that are dependent on the objectives of the study and the underlying structure of the data.

4. Clustering and Classification of Cell Types

One of the most widely used analytical approaches used to make interpretations of single-cell data has been to cluster cells based on their transcriptional or chromatin accessibility profiles in an unsupervised manner (Kiselev, Andrews, and Hemberg 2019). In experiments that produce data structures where distinct clusters of cells are expected, many of which have profiled transcriptional differences between distantly related cell types (Rozenblatt-Rosen et al. 2017), these clustering algorithms perform well. Many clustering algorithms have been developed, which have been iteratively improved by benchmarking their ability to correctly assign cells to empirically defined groups.

Most clustering algorithms are dependent on dimensionality reduction and are sensitive to transformations of data during preprocessing steps, including normalization and batch correction (Krzak et al. 2019). While many different algorithms to cluster high-dimensional data have been developed, the majority of which were developed for applications outside of biology (Krzak et al. 2019). In recent scRNA-seq studies, cell clustering is predominantly performed through Louvain clustering (Meo et al. 2011). Louvain clustering is a network modularity algorithm that generates communities, or

clusters, using information from global and local network information (J. H. Levine et al. 2015).

As cell types of individual tissues have become better characterized, it has become widely appreciated that transcriptional heterogeneity amongst cells of a particular tissue is common across multiple biological settings (Z. Wu and Wu 2020). Because cell clustering is conventionally performed based on differences in gene expression alone, there is no way to predict the boundaries between groups of cells that functionally distinguish heterogeneous populations of a tissue. As such, it is impossible to distinguish clusters that are driven by functional or phenotypic differences between cells from clusters that are driven by variation in gene expression data that reflects biological noise. This is particularly relevant when analyzing data that is highly continuous. In these situations, one popular way that this has been overcome is through identification of gene modules or topics that appear to be co-regulated, which are based upon algorithms designed for natural language processing (X. Wu, Wu, and Wu 2021). Importantly, the ‘topics’ that are formed from these kinds of analysis can incorporate genes that are represented in many other ‘topics’ and perform well on data that cannot be reliably or robustly clustered.

Gene expression modules and cell clusters are functionally treated as equal and independent entities. However, differentiation of cell types occur in a hierarchical manner, and gene expression programs change in a manner that is constrained by gene regulatory networks (S. Huang 2012; Z. Wu and Wu 2020). In situations of development, this may be particularly of importance because cell differentiation occurs in a semi-hierarchical manner (Packer et al. 2019; Fincher et al. 2018; M. Plass et al.

2018). The data produced in many of these biological settings can be both clustered and continuous, which may be reflective of differentiation programs that connect different populations of cells, but may also suggest that clustering methods may perform well even in settings that contain continuous data. Although many clustering algorithms have been compared through benchmarking studies, there are very few ways to evaluate the clustering schematic of a dataset in the absence of reference labels (Z. Wu and Wu 2020). Most clusters are annotated through referencing and correlation to previously described gene expression signatures (Stuart et al. 2019).

In practice, clustering schematics and annotations are evaluated through manual interpretation of genes differentially expressed in a cluster and through comparison to gene signatures described in other biological settings. While denoising is inherent in many transformations of high-dimensional single-cell data (van der Maaten 2008; McInnes et al. 2018; Smolander et al. 2021), even after data has been denoised, many latent variables can still create variation across measurements. Specifically, when clustering based methods generate groups of cells with gene expression patterns that have not been previously described, or in situations where a model does not exist for a biological process, it is extremely difficult to determine if the resulting clusters from data are driven by technical or biological noise. As such, conclusions made from clustering schemas of single-cell data are often considered unreliable without empirically derived data to support them.

5. Biological Variation

5.1 Biological Noise

Much of the data produced in biological experiments cannot be explained by technical handling or explicitly defined experimental variables, even in the most simplified systems (Elowitz et al. 2002; Eling, Morgan, and Marioni 2019). Stochastic processes govern many systems that are relevant in biology, ranging from those that dictate molecular thermodynamics to those that dictate gene expression (Tsallis 1988; Elowitz et al. 2002) and beyond.

In animal models, many attempts are made to control for experimental variables that can produce noise in experiments. Through careful selection of animals with a particular genetic background, age, and gender, these variables can be experimentally defined. Other sources of biological variation such as diet (Y. Yang et al. 2014), circadian rhythm (Wager-Smith and Kay 2000), and pathogen exposure (León et al. 2009) can be controlled through adoption of standardized care and housing of experimental animals (Voelkl et al. 2020). Additional variation that is inherent for aging-related processes ("Know Thy Mouse: Variability in Aged Mice" n.d.; Bahar et al. 2006), hormone fluctuations (T. Liu et al. 2020), and genetic circuitry (Kaern et al. 2005) also produce biological noise that can be partially, but not completely, controlled for.

In some tissue contexts, and in the lung in particular, there are periodic changes associated with normal organ function, such as breathing, which can also drive biological noise. Breathing rhythms directly cause periodic changes in pressure and biomechanical force within the thoracic cavity (Ferris and Pollard 1960; Zamprogno et al. 2021). Further, the phase and amplitude of breathing rate also varies such that

equilibrium can be achieved in response to dynamic changes in pH and oxygen or carbon dioxide concentrations of the blood (Serna et al. 2018).

Although all of these processes occur in a regular manner, the time scales over which they act vary with biological context. Systems that exhibit periodic behavior can undergo phase shifts, changes in amplitude or frequency. Even amongst stochastic processes, such as aging, characteristic reproducibility is observed (J. W. Curtsinger et al. 1995). In many cases, these influences can create variations in biological data that are ultimately observed as noise.

5.2 Biological Metadata

As previously discussed, biological identities, which are treated as metadata for a tumor or cell, often serve as biological ground truths to ensure we are interpreting the data with respect to the things that matter to us. This method is reliable because metadata is typically information that is collected about a cell or tumor independently from the assay used for analysis.

In the context of KP tumors, these identities can reflect the mouse a tumor develops in, the time point a sample was harvested at, genotype, or an explicitly defined experimental variable. Likewise, in studies of T cell responses to antigens, this can be the cognate antigen of the cell. In most cases, even when a study has a clearly defined objective and underlying hypothesis, the ultimate purpose of that study is to either refute or support a theory that explains a biological process. Often, we seek to understand how biological metadata influences a biological process.

In tumor evolution, the most pertinent biological metadata is the time point a sample was harvested at, with the ultimate goal of understanding changes in the tumor

microenvironment, changes within a tumor, and changes within a tumor cell as cancer progresses. Tumor progression is a time-dependent process that occurs reproducibly, but with some variation in kinetics (Jackson et al. 2005). Although this particular feature of the model makes an absolute time point less reliable, an estimation can still be made about the distribution of tumor states at a given time point. This combination of reproducibility and kinetic variation is also observed amongst immune cells in these tumor models (DuPage et al. 2011).

5.3 Biological Pseudotime

Absolute time is an intangible dimension (Hoefer and Ray 1992) that is entirely independent from tumor progression. It is symmetric and is invariably defined by intervals of a constant size. Stated plainly, absolute time is what is usually referred to in time, and is measured by seconds, minutes, etc. The kinetics of biological processes vary, and as such, the changes which occur during those processes can also be thought to progress on an axis of time that is rescaled according to the dynamics of an individual process. Mathematically, this is often termed biological “pseudotime”, which is treated as an intangible dimension and largely follows the philosophical frameworks proposed by Substantivalism (Sklar 1976), in which relative and gradual biological changes occur on an axis that is an ontology of absolute time, or rescaled time. In other words, pseudotime is an axis that is defined by sequential changes that occur during some sort of biological process, rather than the time required for those changes to occur.

Units on this axis are defined by distance in an informational space that governs an overt biological phenotype, rather than absolute time. For example, a cell may undergo rapid change in transcriptional state upon transformation with respect to

absolute time. However, subsequent transcriptional changes or progression through a transcriptional space during tumor progression may occur more slowly. In this example, the axis of pseudotime would be defined by the specific transcriptional changes between a cell before transformation and cells from well-progressed tumors. If the amount of transcriptional change observed in early transformation is equivalent to the amount of change observed between cells at later stages of tumor progression, the pseudotime distance between these cells may be equal, even though the time that is required for these changes may be entirely different.

Although there have been many proposed ways of calculating pseudotime in single-cell data, they all operate on a similar principle of pseudotime theory. Statistically, pseudotime can be modeled by gaussian mixture models (Lönnberg et al. 2017), differentiation potential (S. Jin et al. 2018), optimal transport (Schiebinger et al. 2019), and minimal spanning trees (Trapnell et al. 2014). All of these algorithms exploit the continuity of cell state, transcriptional or otherwise, to generate these models of differentiation trajectories. Settings where cells undergo gradual transcriptional changes during differentiation produce a continuum of cell states, and in single-cell data, can produce highly continuous data structures (M. Plass et al. 2018; Fawkner-Corbett et al. 2021). Often, subsets of data produced from biological settings where continuous developmental cell states are observed can also produce highly clustered data structures. As such, many of these algorithms are relatively limited to situations where continuity in cell state is observed. Importantly, many of these algorithms are extremely sensitive to artifacts of dimensionality reduction, batch correction, and in some cases clustering schemas. Some software provides the option to determine the ancestral

populations computationally, although these predictions are rarely robust; in most cases, these algorithms perform most optimally when ancestral states are empirically known or defined.

Many biological processes associated with disease are not programmed. In contrast, developmental processes and differentiation are highly conserved programs that have evolved as a consequence of Darwinian selection. This reasoning can be extended to all processes designed to protect evolutionary fitness and are tightly regulated ("Evolutionary Bioscience as Regulatory Systems Biology" 2011). However, cancer and many diseases occur as a consequence from loss of regulatory processes that result from somatic mutation, epigenomic, and/or genomic instability. Intriguingly, many of these diseases progress with a surprising degree of reproducibility across individuals (Hanahan and Weinberg 2011). This reproducibility forms the basis for patient stratification in modern medicine.

Cancer progression is largely dependent on stochastic and unprogrammed changes, which is why its progression is not entirely deterministic ("A Stochastic Model in Tumor Growth" 2006). The probability that these stochastic changes will occur is a function of many factors, including number of cell divisions, replication error, and stability of cell state. Because the likelihood of these events accumulates with time, this forms the basis through which time plays a causative role in cancer progression. The influence of time on the progression of a tumor is thought to be probabilistic in nature ("Permanence and Extinction of a Stochastic Hybrid Model for Tumor Growth" 2019). As such, metrics of time may be largely unreliable when studying disease across organisms. Instead, characterization of tumor evolution with respect to pseudotime

eliminates the variations observed at a given time point, because it occurs on an axis that is unaffected by the asynchronous nature of tumor progression (Trapnell et al. 2014).

Several precedents exist to support the fact that the developmental history of a cell is closely influenced by its transcriptional state (Packer et al. 2019; S. Huang 2012). Every trajectory an entity (e.g. a tumor or cell) may have through an informational space (e.g. transcriptional space) will progress along a path that reflects changes of cell state and is *specific* to the biological process. The states that are connected by these trajectories are defined by the starting and ending populations (Trapnell et al. 2014) of the biological process. In the context of tumors, these trajectories can be thought of as adaptations to selective pressures of the tumor microenvironment.

Cell states are heterogeneous for both tumor cells and cells of the tumor microenvironment; the global trajectory a tumor and its related cells will take is the aggregate product of the individual trajectories associated with each cell state change in the tumor and its microenvironment (“Tumor Functional Heterogeneity Unraveled by scRNA-Seq Technologies” 2020). This is complicated by the fact that tumors are thought to arise from a single cell (Peter C. Nowell 1978), so all tumor cells are thought to have some sort of ancestral relationship. As a consequence, it is likely that tumor cells progress through many trajectories that ultimately give rise to the variation in phenotypes observed in a heterogeneous tumor (Peter C. Nowell 1978; P. C. Nowell 1976). The degree of heterogeneity observed in tumor cells themselves are not constant throughout tumor evolution. Cells of the tumor microenvironment, in contrast, do not undergo changes during tumor progression in a clonal manner, as many of these

tissues are heterogeneous under normal homeostatic conditions. Still, non-tumor cells of the microenvironment undergo phenotypic changes during tumor progression. In these situations, ancestral cell states are not empirically known and cannot be assumed, and limits the utility of pseudotime based approaches.

CHAPTER 2

ETS Transcription Factors in KP Lung Tumor Evolution

Amanda M. Cruz, Lindsay LaFave, Christina Cabana, Isabella DelPriore, Emma Dawson, Michelle Yin, Toni-Ann Nelson, Santiago Naranjo, Nemanja Marjanovic, David Canner, Matan Hofree, Tuomas Tammela, Aviv Regev, Tyler Jacks

AUTHOR ACKNOWLEDGMENTS

A.M.C., L.L., and T.J. designed the study. A.M.C, I.D, S.N., C.C., T.N., M.Y., and E.D. performed all of the mouse experiments in the lab of T.J.; A.M.C. performed the computational analyses associated with the scRNA-seq datasets.

Figure 1, Figure S1, Figure 2, Figure 3, and Figure 4 were produced using computational analyses performed entirely by A.M.C. with data originally published in Marjanovic et al. 2020. Figure 5 and Figure 6 contain data from experiments entirely performed and designed by A. M. C. with technical assistance for intratracheal delivery of lentivirus by Megan Burger, except Figure 5E, which contains data from experiments designed and executed by Amanda Cruz, Christina Cabana, Isabella DelPriore, Emma Dawson, and Lindsay LaFave.

ABSTRACT

In lung adenocarcinoma (LUAD), aberrant expression of genes canonically expressed during embryonic development identifies highly metastatic and poorly differentiated primary lung tumors of genetically engineered mouse models (GEMMS) and of human LUAD patients. Through analysis of scRNA-seq data generated from longitudinally sampled individual cells of lung adenocarcinomas and adenomas, we determined that tumor evolution occurs reproducibly across mice and tumors. Pseudotime based analysis identified that the *Pea3* transcription factor family becomes aberrantly expressed during lung tumor evolution. *Etv4*, an embryonic-restricted transcription factor, is silenced in the normal lung epithelium but becomes latently expressed in tumors. *Etv5*, a master regulator of alveolar type II cell (AT2) identity, is co-expressed with *Etv4* during embryonic lung morphogenesis and its expression is maintained through adulthood. Etv4 and Etv5 play essential roles during stages of early transformation. Manipulation of both genes at the onset of transformation events drastically diminish the frequency of tumorigenesis, but are likely dispensable in later stages of tumor evolution.

INTRODUCTION

Solid tumors are ecosystems of cells that share a common ancestor and evolutionarily diverge from transformation of a single cell (Greaves and Maley 2012). As these cells evolve during tumor progression, they will phenotypically and functionally diverge to form multiple subclones within a tumor with differing levels of evolutionary fitness that can affect both the subclone alone or the tumor as a whole. Importantly, this

heterogeneity is believed to be the primary mechanism of tumor plasticity and ultimately makes tumors difficult to treat or entirely eliminate (Lawson et al. 2018).

Lung adenocarcinoma (LUAD), which is the most common form of non-small cell lung cancer, is believed to predominantly arise from a cell type of the distal lung epithelium that produces surfactants, known as alveolar type II (AT2) cells (Mainardi 2013; Sutherland et al. 2014). Two of the most common oncogenic mutations observed in clinical LUAD are activating gain-of-function mutations in KRAS and loss of function mutations in the tumor suppressor TP53 (Sutherland et al. 2014; Jackson et al. 2005; E. L. Jackson et al. 2001). In mice, expression of *Kras*^{G12D} is sufficient to drive transformation of cells in the lung epithelium, frequently producing lesions which histologically recapitulate hyperplasias and adenomas (E. L. Jackson et al. 2001). By introducing an additional oncogenic hit during transformation with homozygous loss of *Tp53*, lung lesions will progress to histological adenocarcinomas with the capacity to metastasize (Jackson et al. 2005), which occurs rarely with *Kras*^{G12D} alone. These adenocarcinomas, both in mice and humans, are distinguished from adenomas by characteristic morphological features that are indicative of highly dysregulated cell identity (DuPage, Dooley, and Jacks 2009). Therefore, LUAD can be faithfully modeled in immune competent genetically engineered mice in an autochthonous manner.

Multi focal tumors can be selectively initiated in the lungs of mice harboring Cre recombinase conditional alleles that permit expression of heterozygous *Kras*$^{LSL\text{-}G12D}$ (K), driven by its endogenous promoter, and homozygous deletion of *Trp53*$^{fl/fl}$ (P) through intratracheal delivery of viral Cre recombinase (DuPage, Dooley, and Jacks 2009). Importantly, although oncogenic mutations in *Kras* and *Trp53* (KP) occur simultaneously,

tumor progression kinetics and disease latency are somewhat variable (Jackson et al. 2005). The cells that are believed to give rise to lung adenocarcinomas and adenomas are alveolar type II (AT2) cells (Sutherland et al. 2014), which are found in the distal lung epithelium and are primarily responsible for surfactant production. The other epithelial cell type of the distal lung is known as an alveolar type I cell (AT1) and is responsible for gas exchange in the alveolus (Little et al. 2021). Despite their functional differences, these two cell types have intertwined developmental relationships; both cell types arise from a common progenitor, and AT2 cells can regeneratively differentiate into AT1 cells upon injury (Barkauskas et al. 2013).

Previously, metastasis of lung tumors in KP mice has been characterized through bulk gene expression studies, which led to the discovery that *Nkx2-1*, a well-characterized transcriptional regulator of lung identity (Yuan et al. 2000), is frequently downregulated in poorly differentiated high grade primary lung adenocarcinomas and their metastases (Winslow et al. 2011; C. M.-C. Li et al. 2015). Loss of *Nkx2-1* will cause tumor cells to lose their lung identity and instead adopt a gastric-like phenotype that is partially, but not fully, responsible for driving gene expression associated with late stage and metastatic tumors (Winslow et al. 2011; C. M.-C. Li et al. 2015; Snyder et al. 2013). Until recently, developmental *Nkx2-1* expression was believed to be selectively retained in AT2 cells compared to AT1 cells once lung development is complete, but it has since been found to be required for AT1 transcriptional and morphological identity as well as their quiescent behavior (Little et al. 2019). When deleted, *Nkx2-1* null embryos display drastic defects of tracheal and lung

morphogenesis, and specifically do not generate surfactant producing AT2 cells (Minoo et al. 1999).

The Pea3 subfamily, which consists of *Etv4*, *Etv5*, and *Etv1*, plays a crucial role during lung morphogenesis by creating primary and secondary branching events in the developing lung bud. Importantly, these branching events ultimately establish proximal and distal patterning in the lung (Cardoso and Lü 2006). *Etv4* and *Etv5* in particular are co-expressed at the distal end of the developing lung bud, while *Etv1* is expressed in the surrounding mesenchyme. Although co-expressed, *Etv4* and *Etv5* have non-redundant roles during development. The PEA3 family is expressed in the mouse both during development and in adult cells (Hollenhorst, Jones, and Graves 2004; Chotteau-Lelievre et al. 2001; Y. Liu et al. 2003).

In the early stages of murine development, *Pea3* transcription factors are ubiquitously expressed. As development progresses, their expression becomes restricted to tissues whose differentiation relies upon extensive interactions between the epithelium and mesenchyme (Cardoso and Lü 2006). Their expression is largely associated with cell migration, proliferation, and remodeling of the surrounding extracellular matrix (Chotteau-Lelièvre et al. 1997). Upon maturation of the distal lung bud, expression of *Etv4* is silenced, while *Etv5* expression persists (Y. Liu et al. 2003; Chotteau-Lelièvre et al. 1997). *Etv4* and *Etv5* are believed to create lung bud branches through regulating periodic interactions between the Fibroblast growth factor (Fgf) and Sonic Hedgehog signaling pathways. *Etv4* and *Etv5* are nuclear targets of Fgf10 and induce Fibroblast growth factor signaling, and promote Sonic Hedgehog signaling (Herriges et al. 2015). In turn, a branching event occurs when Shh signaling creates a

temporal delay in Fgf signaling (Herriges et al. 2015). In the context of oncogenic *Kras*^{G12D}, *Etv5* is required to maintain normal AT2 identity; deletion of *Etv5* disrupts lung tissue regeneration following bleomycin-induced injury and inhibits oncogenic transformation by *Kras*^{G12D} in the lung (Z. Zhang et al. 2017). Importantly, in lesions with heterozygous deletion of *Etv5*, *Etv4* is upregulated, suggesting they are mechanistically redundant for early stages of transformation (Z. Zhang et al. 2017).

Therefore, aberrant expression patterns of *Etv5* and *Nkx2-1*, which have been implicated in both early and late stages of *Kras*^{G12D} transformation, respectively, demonstrate the functional implications of dysregulation of cell identity throughout *Kras* mutant lung tumor progression. In order to understand how cell identity becomes dysregulated in Kras mutant tumors as they progress, a single-cell RNA-sequencing (scRNA-seq) study (Marjanovic et al. 2020) was conducted on longitudinally sampled Kras mutant lung tumor cells, and a single-cell combinatorial indexing ATAC-sequencing study was conducted on late stage tumors and their metastases (LaFave et al. 2020). Using these data, I identified putative transcriptional regulators that are known to affect cell identity.

RESULTS

KP Tumor progression occurs reproducibly over time across mice and tumors

Through introduction of a *Cre*-conditional fluorescent reporter allele to the *Rosa26* safe harbor locus, *Rosa26*$^{\text{LSL-tdTomato}}$, the cells that are transduced by viral *Cre* and give rise to Kras mutant lung lesions can be isolated from other lung tissue through fluorescence-activated cell sorting. Additionally, further specificity over Cre expression

can be achieved through intratracheal delivery of adenovirus, which drives expression of Cre using an AT2 specific promoter, *Sftpc*, a surfactant gene (Sutherland et al. 2014). With this methodology, to landscape transcriptional evolution of Kras mutant lung tumor cells throughout disease progression, adenoviral Sftpc-Cre (AdSPC-Cre) was used to infect the lungs of $Kras^{LSL\text{-}G12D/+}$,$Rosa26^{LSL\text{-}tdTomato/+}$;$Trp53^{fl/fl}$ (KP) and $Kras^{LSL\text{-}G12D/+}$,$Rosa26^{LSL\text{-}tdTomato/+}$;$Trp53^{+/+}$ (K) mice and generate transformed cells that were longitudinally sampled after tumor initiation that reflect hyperplasias, adenomas, and adenocarcinomas (Jackson et al. 2005). As a control, *$Kras^{+/+}$,$Rosa26^{LSL\text{-}tdTomato/+}$;$Trp53^{+/+}$* (T) mice, which do not harbor tumor initiating alleles, were included in these experiments to empirically characterize the cells that are infected by AdSPC-Cre and give rise to tumors in K and KP mice. $TdTomato^{+}$ cells were then isolated from the lungs of these animals at various time points (0, 2, 4, 12, 18, 20, and 30 weeks) after tumor initiation, sorted by flow cytometry to specifically isolate non-immune cells, and then assayed by SMART-Seq2, a plate-based scRNA sequencing technology (Marjanovic et al. 2020).

The following exploration of this scRNA-seq experiment will utilize data that was originally produced and described by (Marjanovic et al. 2020), which has been reanalyzed with slightly different methodologies (see Methods) than those originally described, for the following discussions. For visual inspection of transcriptional profiles, the structure of data from scRNA seq was approximated on a uniform manifold with Uniform Manifold Approximation and Projection (UMAP) (McInnes et al. 2018; Becht et al. 2018) and on a hyperbolic manifold with scPHERE (Ding and Regev 2021). In both embeddings, the transcriptional state of cells isolated from control T mice and from early

time points after tumor initiation (2 and 4 weeks) appear to have minimal transcriptional heterogeneity, but continually diverge by 12 weeks and beyond, as expected (**Figure 1A, 1B**). Across these timepoints, very few genes selectively distinguish cells harvested at one time point from another (**Figure 1C**), and most marker genes appear to have similar, but variable, expression across time points. These data suggest that, when aggregating cells at an individual time point, the primary feature that distinguishes cells isolated at different time points is the degree of transcriptional heterogeneity observed, rather than global transcriptional changes. Still, there are some transcriptional states observed, albeit variably, in cells isolated from late timepoints (18, 20, and 30 weeks after tumor initiation) that are rarely, if ever, reflected in cells isolated at earlier time points. With respect to transcriptional state, despite variation observed at individual time points, evolution of Kras mutant tumor cells over time have characteristic, but heterogeneous, features that can generally distinguish between tumor cells isolated from early and late time points. Importantly, these observations largely agree with and have also been discussed by (Marjanovic et al. 2020).

Tumor cells isolated from K mice are assumed to predominantly reflect cell states of histological hyperplasias and adenomas, while cells isolated from KP mice are assumed to reflect adenomas and adenocarcinomas (Jackson et al. 2005). To visualize transcriptional states that distinguish these genotypes, and by proxy their associated histological stages, we compared positions of cells isolated from T, K, and KP mice (**Figure 1D, 1E**) in the same embeddings described in **Figure 1AB**. There was a clear overlap of transcriptional states that appear to be most similar to control AT2 cells of T mice, which are reflected in cells of early, late, and intermediate time points (**Figure**

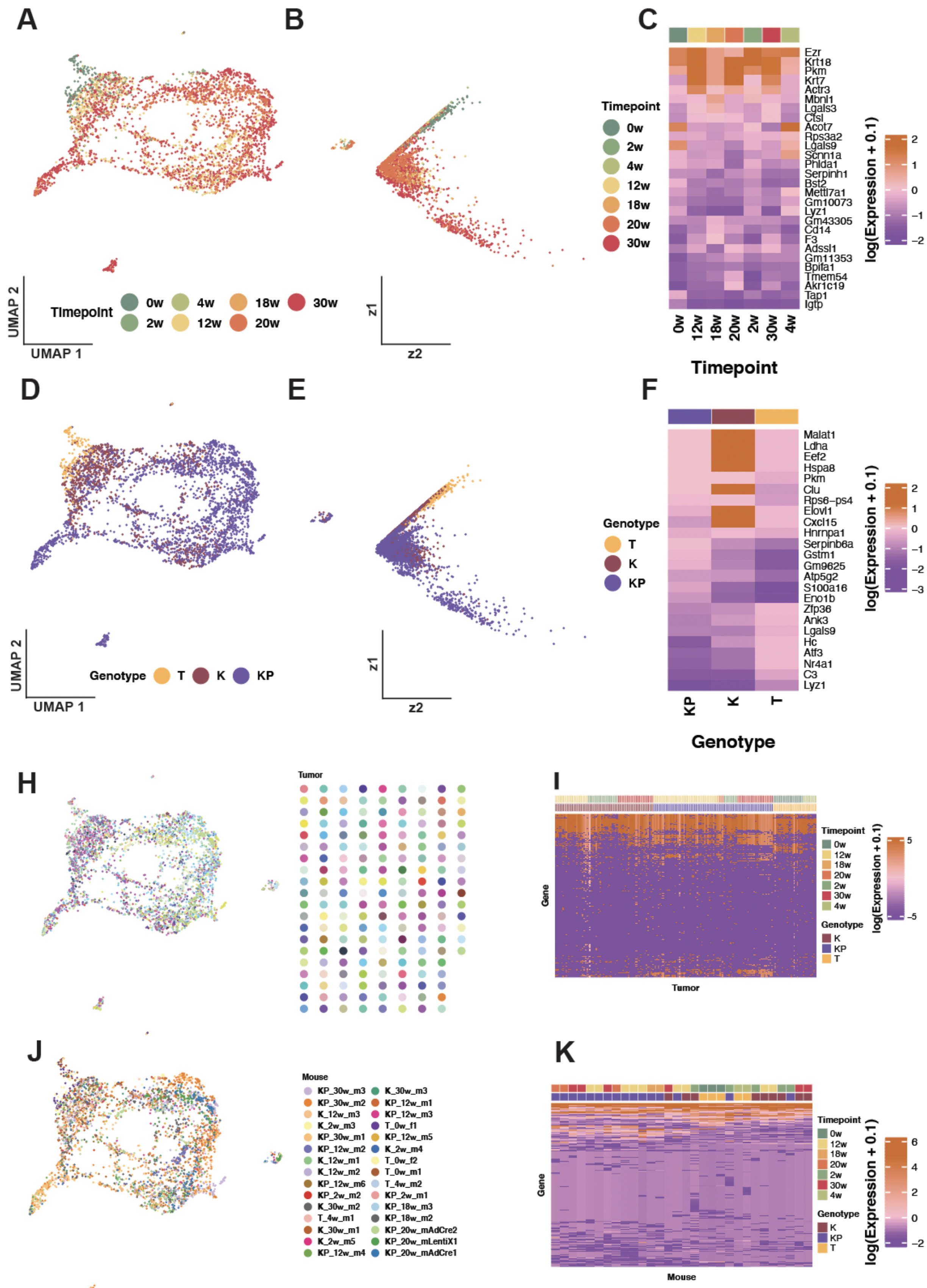
1AB). However, transcriptional states which are predominantly observed at late time points in KP tumor cells, but less frequently in K tumor cells, reflect a clear divergence of evolutionary trajectories between these cells with K and KP genotypes, as is reflected most clearly in **Figure 1E**.

The genes that distinguish these transcriptional trajectories may include genes that drive transitions to histological adenocarcinomas, and as such, I assessed bulk expression of marker genes that are characteristic of K, KP, and T cells and visualized them across these genotypes (**Figure 1F, Figure S1A-D**). As expected, many of the identified marker genes robustly distinguish transformed (K & KP) tumor cells from control AT2 cells (T), and other genes distinguish KP from K and T cells, consistent with the transcriptional overlap and divergence observed in low dimensional embeddings. These transcriptional states are observed recurrently and reproducibly across mice within the same genotype, and provide additional evidence for the meaningful impact oncogenic driver mutations can have on transcriptional trajectories of tumor cells over disease progression.

Despite the fact that tumor progression is generally believed to be a stochastic process (M. Guo et al. 2019), both with respect to mutations and epigenetic state, there is a striking degree of reproducibility observed in transcriptional state of KP tumor cells. Nearly all of the states observed during tumor progression are observed in multiple tumors (**Figure 1G**). Similarly, the genes with the greatest variable expression across the entire dataset have highly similar bulk expression between individual tumors or lung samples of a given genotype (**Figure 1H**).

In agreement with the general behavior of transcriptional evolution during KP lung tumor progression with respect to genotype and time point, unbiased hierarchical clustering with Canberra distances between expression of highly variable genes across individual tumors roughly arranged individual tumors in a manner that reflects variations and similarities of tumor progression kinetics observed across time points and genotypes (**Figure 1H**). Particularly, in agreement with **Figure 1D-F,** K tumors are generally most similar to cells of T mice and early to intermediate time point KP tumors. The divergence in overall tumor state was maximal between KP tumors and T samples.

In addition, transcriptional states of individual cells are also highly reproducible across mice (**Figure 1I**). The variation in expression of highly variable genes in the dataset observed across tumors in **Figure 1H** are similar to those observed across aggregate gene expression of cells derived from an individual mouse (**Figure 1J**); the variation observed can be primarily attributed to genotype of the mouse and the time point the mouse was sacrificed at. Additionally, reproducibility is observed between male and female mice (**Figure S1E-G**) and, following batch correction, across days batches were processed (**Figure S1H**).



A
B
C
UMAP 2
UMAP 1
Timepoint
0w
2w
4w
12w
18w
20w
30w
z1
z2
Ezr
Krt18
Pkm
Krt7
Actr3
Mbnl1
Lgals3
Ctsl
Acot7
Rps3a2
Lgals9
Scnn1a
Phlda1
Serpinh1
Bst2
Mettl7a1
Gm10073
Lyz1
Gm43305
Cd14
F3
Adssl1
Gm11353
Bpifa1
Tmem54
Akr1c19
Tap1
Igtp
log(Expression + 0.1)
D
E
F
Genotype
T
K
KP
Malat1
Ldha
Eef2
Hspa8
Pkm
Clu
Rps6−ps4
Elovl1
Cxcl15
Hnrnpa1
Serpinb6a
Gstm1
Gm9625
Atp5g2
S100a16
Eno1b
Zfp36
Ank3
Lgals9
Hc
Atf3
Nr4a1
C3
Lyz1
H
Tumor
I
Gene
J
Mouse
KP_30w_m3
KP_30w_m2
K_12w_m3
K_2w_m3
KP_30w_m1
KP_12w_m2
K_12w_m1
K_12w_m2
KP_12w_m6
KP_2w_m2
K_30w_m2
T_4w_m1
K_30w_m1
K_2w_m5
KP_12w_m4
K_30w_m3
KP_12w_m1
KP_12w_m3
T_0w_f1
KP_12w_m5
K_2w_m4
T_0w_f2
T_0w_m1
T_4w_m2
KP_2w_m1
KP_18w_m3
KP_18w_m2
KP_20w_mAdCre2
KP_20w_mLentiX1
KP_20w_mAdCre1
K

Figure 1. ***KP Tumor progression occurs reproducibly over time across mice and tumors.***

Data used for these analyses has been previously published (Marjanovic et al. 2020) and were independently analyzed to produce these figures. The rows in all heatmaps of this figure are arranged by complete linkage hierarchical clustering of their euclidean distances.

(**A-B**) (**A)** UMAP or (**B**) hyperbolic scPHERE embedding of scRNA-seq gene expression data cells harvested at 0 weeks (n = 3, 162 cells), 2 weeks (n = 5, 142 cells), 4 weeks (n = 2, 44 cells), 12 weeks (n = 6, 946 cells), 18 weeks (n = 2, 103 cells), 20 weeks (n = 3, 435 cells), and 30 (n = 15, 2,059 cells) weeks post-tumor initiation. Total n = 24 mice, 3,891 cells.

(**C**) Heatmap depicting continuity of mean log-normalized expression of the top 4 marker genes for each time point, ranked by q-value.

(**D-E**) (**D)** UMAP or (**E**) hyperbolic scPHERE embedding of scRNA-seq gene expression data of cells harvested from mice without an oncogenic mutation (“T”, n = 5, 206 cells), with Kras$^{G12D/+}$ (“K”, n = 9, 1,071 cells), and with Kras^{G12D};Tp53$^{\Delta/\Delta}$ (“KP”, n = 16, 2,614 cells), post-tumor initiation. Total n = 36 mice, 3,891 cells.

(**F**) Heatmap depicting continuity of mean expression of the top 8 marker genes for each genotype, ranked by q-value.

(**G**) UMAP embedding of scRNA-seq gene expression data of cells harvested from each sampled tumor. (“T”, 29 samples), with Kras$^{G12D/+}$ (“K”, 66 tumors), and with Kras^{G12D};Tp53$^{\Delta/\Delta}$ (“KP”, 80 tumors), post-tumor initiation. Total n = 146 tumors.

(**H, J**) Heatmap depicting mean log-normalized expression of the top 2000 highly variable genes across (I) tumors or (K) mice. Genotype and Timepoint metadata is annotated above the heatmap. Canberra distances between mouse or tumor samples (columns) are hierarchically clustered by complete linkage.

(**J**) UMAP embedding of scRNA-seq gene expression data of cells harvested from each Mouse.

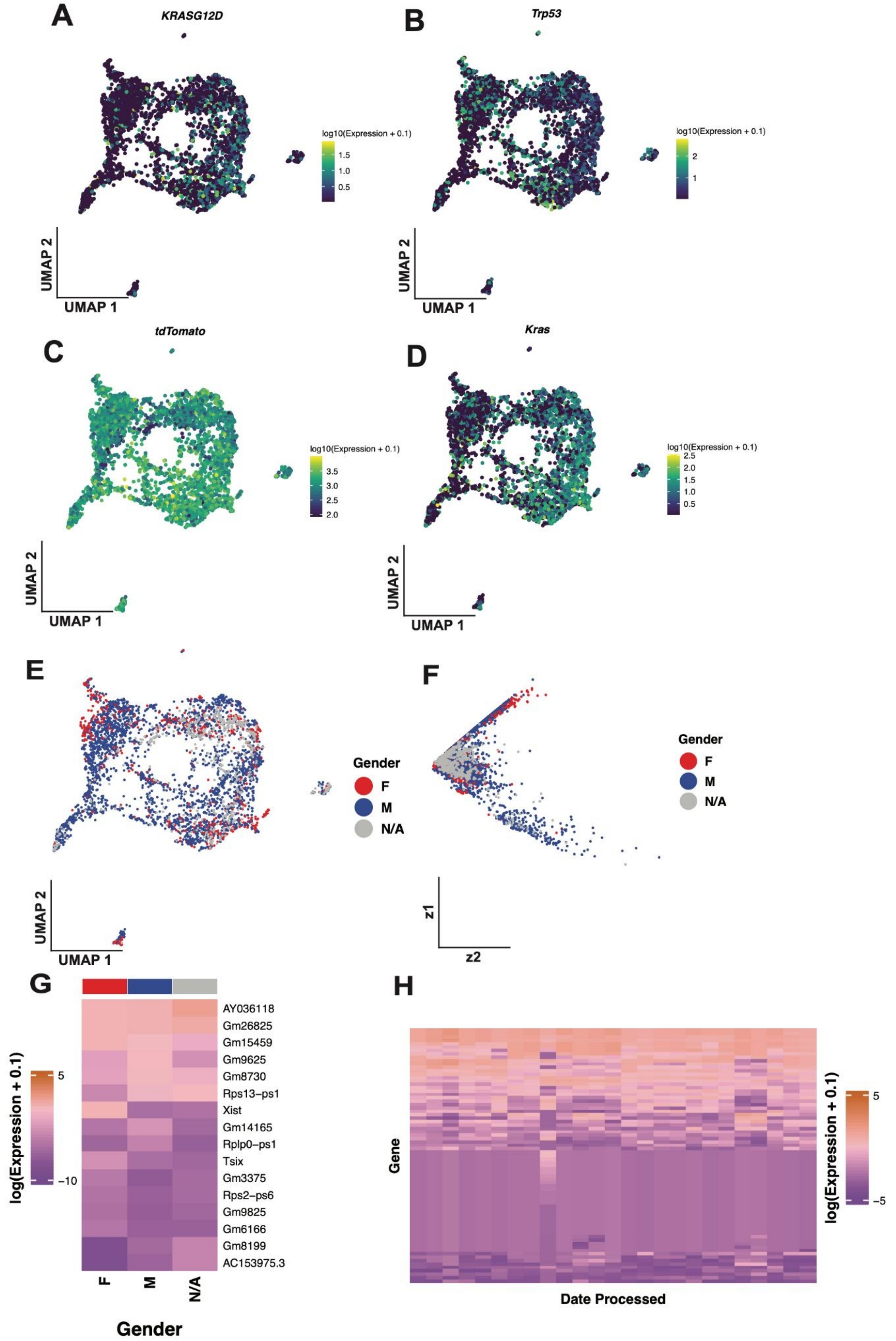

A
KRASG12D
log10(Expression + 0.1)
1.5
1.0
0.5
UMAP 2
UMAP 1
B
Trp53
log10(Expression + 0.1)
2
1
UMAP 2
UMAP 1
C
tdTomato
log10(Expression + 0.1)
3.5
3.0
2.5
2.0
UMAP 2
UMAP 1
D
Kras
log10(Expression + 0.1)
2.5
2.0
1.5
1.0
0.5
UMAP 2
UMAP 1
E
Gender
F
M
N/A
UMAP 2
UMAP 1
F
Gender
F
M
N/A
z1
z2
G
AY036118
Gm26825
Gm15459
Gm9625
Gm8730
Rps13−ps1
Xist
Gm14165
Rplp0−ps1
Tsix
Gm3375
Rps2−ps6
Gm9825
Gm6166
Gm8199
AC153975.3
log(Expression + 0.1)
5
−10
F
M
N/A
Gender
H
Gene
Date Processed
log(Expression + 0.1)
5
−5

Figure S1. ***KP Tumor progression occurs reproducibly over time across mice and tumors.***
Data used for these analyses has been previously published (Marjanovic et al. 2020) and were independently analyzed to produce these figures. The rows in all heatmaps of this figure are arranged by complete linkage hierarchical clustering of their euclidean distances.
(**A-D**) UMAP embedding of scRNA-seq gene expression data depicting expression of (A) Oncogenic *Kras*^{G12D}, (B) *Trp53*, (C) *tdTomato*, and (D) wild-type *Kras*.
(**E-F**) (**E)** UMAP or (**F**) hyperbolic scPHERE embedding of scRNA-seq gene expression data of cells harvested from female mice (n = 5 mice, 650 cells), male mice (n = 20 mice, 2,073 cells) or mice with unassigned genders (n = 5 mice, 538 cells).
(**G**) Heatmap depicting continuity of log-normalized mean expression of the top 8 marker genes for each genotype, ranked by q-value.
(**H**) Heatmap depicting mean log-normalized expression of the top 2000 highly variable genes across dates when cells were processed.

Biological pseudotime approximates gene expression changes in KP lung tumor progression

Transcriptional changes during progression of KP tumors that are thought to have the greatest translational relevance are those that occur between early stage hyperplasias and metastatic tumors. As such, to develop a framework for studying tumor progression in the KP model, we utilized expression of two previously characterized markers of late-stage tumor cells, *Hmga2* (Winslow et al. 2011) and *Runx2 (LaFave et al. 2020)* to identify tumor cells in transcriptional space that are believed to give rise to metastases (**Figure 2A)**. Although the methodologies of the analyses of this thesis are different from those originally used to analyze this scRNA-seq data, the transcriptional similarity observed in the cell clusters identified by (Marjanovic et al. 2020) generally agree with those observed in this analysis (**Figure 2B-C**). One notable exception to this agreement are the relative positions and distances of cells assigned to Marjanovic cluster 9 in low-dimensional space in comparison to cells of other clusters, which vary dramatically across different embeddings and projections (**Figure 2B-C**, (Marjanovic et al. 2020). Thus, it is entirely likely that the bonafide transcriptional similarities between cells of cluster 9 and other cells in the dataset are distorted by dimensionality reduction.

The clusters identified by (Marjanovic et al. 2020) were named in the order they were believed to arise during tumor evolution and “begin” at Cluster 1, which contains untransformed AT2 cells from T mice. The positions of “early” clusters (1-4) in UMAP and hyperbolic embeddings largely agree with the dynamics originally proposed by (Marjanovic et al. 2020). For “late” clusters, the ordering of these clusters appears less

robust. Still, the cells assigned to "late" clusters are highly connected, supporting that tumor cells progress into and between these states. As such, I reasoned that we might better be able to order these states with greater resolution and confidence through computationally predicting transcriptional trajectories in nearest-neighbor space created by UMAP, which is well suited for continuous data. While there are many suggested methodologies to computationally infer the ancestral or starting population of an evolutionary process, during tumor progression, tumor cells become progressively less differentiated and have transcriptional profiles that resemble stem populations observed during development. A prominent example of this behavior is *Hmga2* expression in met-like cells, as *Hmga2* is otherwise silenced in the mature lung epithelium and restricted to embryonic and early postnatal development (I. Singh et al. 2014). Consequently, many pseudotime algorithms fail to reproducibly and robustly infer tumor cell trajectories in these data.

This concern is alleviated by the experimental design that produced this scRNA-seq dataset, which included cells derived from T mice to empirically identify the cell-of-origin for this particular setting of lung tumor evolution. As such, I elected to construct transcriptional trajectories and generate approximations of biological pseudotime using *Monocle3* (Qiu et al. 2017; Trapnell et al. 2014). This software was also intentionally chosen because it was able to robustly predict a trajectory that connects AT2 cells with metastatic-like *Hmga2*$^{+}$ *Runx2*$^{+}$ cells (**Figure 1D**) and that is consistent with other studies characterizing KP lung tumor evolution (Winslow et al. 2011; C. M.-C. Li et al. 2015). By specifying cells derived from T only mice as the ancestral population of KP tumor cells, otherwise termed as the "root node" (**Figure**

1D), this “anchored” these pseudotime predictions in an empirical biological truth. Consequently, the approximated positions of individual cells across pseudotime (**Figure 1E**) are largely consistent with these previously published studies.

The pseudotime approximations for cells assigned to Marjanovic cluster 9 provide further evidence for distortions in the relationships between these cells with other known populations of the dataset; despite having close positions in 2D-UMAP space, the cells which would otherwise appear to “connect” cells of Marjanovic cluster 9 with met-like *Hmga2*$^{+}$ cells are estimated to have very different positions in biological pseudotime (**Figure 1E**). While it is entirely possible that the branched trajectories predicted by Monocle3 that connect Cluster 9 cells and with untransformed AT2 cells are a computational artifact, at minimum, their relative positions in hyperbolic embeddings do provide further support for the evolutionary distance between these populations (Ding and Regev 2021) (**Figure 1E**). This is not surprising, given that multiple dimensionality reduction techniques fail to robustly approximate the position of cells in Cluster 9 in high-dimensional space. As such, for downstream pseudotime-based analyses, calculations were performed on a subset of cells which excluded those along the trajectories connecting Cluster 9 and the starting AT2 population.

Pseudotime-dependent terms estimated by linear regression of gene expression changes as a function of pseudotime with Monocle3 identified many genes that have been previously characterized to be important in KP lung tumor or lung adenocarcinoma progression, including *Hmga2*, *Tigit (Marjanovic et al. 2020)*, *S100a6 (De Petris et al. 2009)*, and *Hnf4a* (Snyder et al. 2013) (**Figure 2F**). Unbiased hierarchical clustering of

the average expression of these pseudotime-dependent genes across individual tumors (**Figure 2F,** left) arranged tumors in a manner that is consistent with observations made about kinetics of tumor progression observed in **Figure 2** and (Jackson et al. 2005); Early time point K tumors are most similar to T only control cells, and late time point K tumors are most similar to KP tumors (**Figure 2F**). While this may appear to suggest that bulk expression of these genes can be used as an estimation of tumor progression, it should be noted that many early stage tumor cells were too small to be individual plucked from lung tissue, and are instead isolated via sorting cells by tdTomato fluorescence (Marjanovic et al. 2020). Hierarchical clustering fails to arrange individual cells in a meaningful order with respect to their genotypes or associated time points (**Figure 2F**), but this is expected given the magnitude of transcriptional heterogeneity between KP tumor cells.

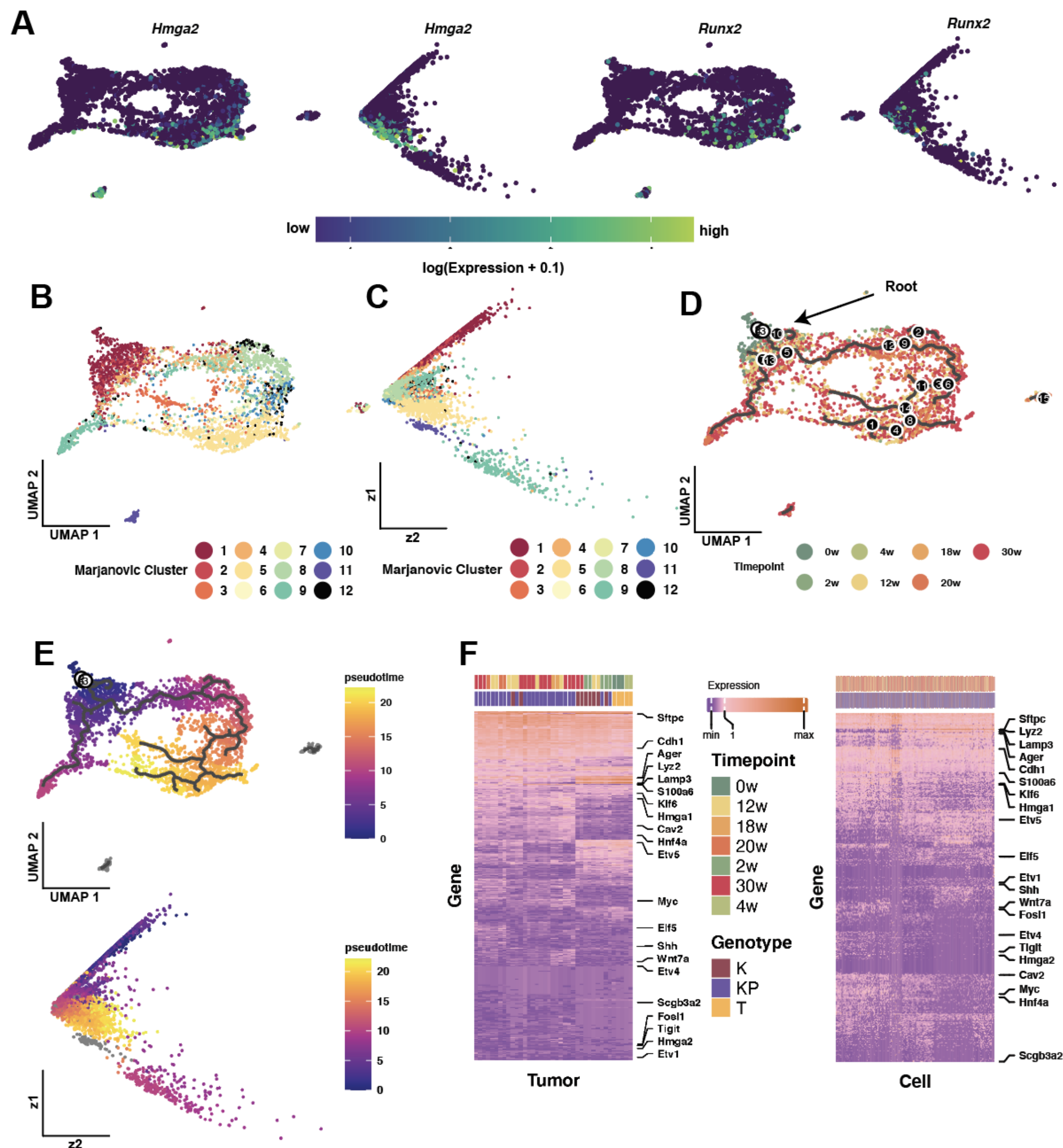


Figure 2. Biological pseudotime approximates gene expression changes in KP lung tumor progression.
Data used for these analyses has been previously published (Marjanovic et al. 2020) and were independently analyzed to produce these figures.
(**A**) UMAP (left) or hyperbolic scPHERE embedding (right) of log-normalized scRNA-seq gene expression of *Hmga2* and *Runx2*.

(**B-C**) (**B)** UMAP or (**C**) hyperbolic scPHERE embedding of scRNA-seq data with cells colored by cluster identities originally published and characterized by (Marjanovic et al. 2020).
(**D**) UMAP embedding of scRNA-seq data depicting trajectories inferred by Monocle3 (black line), with cells colored by their respective time points. The root node for pseudotime analyses is shown with an arrow.
(**E**) UMAP (top) or hyperbolic scPHERE embedding (bottom) of scRNA-seq data with cells colored by predicted pseudotime values generated from (**D**).
(**F**) Heatmaps depicting expression of the top-ranked ($q < 1E\text{-}180$) pseuodotime-dependent terms across individual tumors (left) and across individual cells (right). Rows of the heatmaps are arranged by complete linkage hierarchical clustering of their euclidean distances. Cells and tumors are annotated (top) with their respective genotypes and associated time points.

Sequential changes in cell identity occur throughout KP lung tumor cell evolution and progression.

The objectives of this study were to identify genes that drive evolution of cell state in tumor cells. Because tumor cells are known to have dysregulated cell identity that is also indicative of disease progression and prognosis in clinical settings (Yasukawa et al. 2018), our approach towards interpretation of pseudotime-dependent terms was conducted with a particular emphasis on transcription factors that are hallmarks of cell identity in lung and foregut-derived tissues. Consistent with observations made in the past about KP lung tumor progression (DuPage, Dooley, and Jacks 2009), tumor cells gradually lose hallmarks of AT2 cell identity, including *Etv5*, *Sftpc*, and then *Nkx2-1* over pseudotime (**Figure 3A-C**).

Although cells from Cluster 9 are excluded from these analyses, it is interesting to note that these cells upregulate *Sftpc,* but not *Etv5*, even though they are conventionally expressed in the same context and *Etv5* is known to bind the promoters of surfactant genes and regulate their expression (Z. Zhang et al. 2017). As such,

expression of *Sftpc* may be driven by different mechanisms and provides further insight to the ambiguous cell states observed in Cluster 9. Expression of *Sftpc* and other epithelial markers is surprising amongst these cells (**Supplemental Figure 3**), because they predominantly arise from late time points (**Figure 1A**). Furthermore, these cells are well separated from early time point cells in the hyperbolic embedding, and also express markers associated with AT1 identity, such as *S100a6* ("Website" n.d.; Beauchemin 2016) (**Figure 3D**). Thus, the transcription profiles of cells in Cluster 9 are similar, but distinct compared to those observed at early time points. These similarities likely underlie inconsistencies observed with Cluster 9 across dimensionality reduction techniques.

Expression of *S100a6* and other hallmark genes of AT1 identity, including *Ager*, appears to occur simultaneously with downregulation of AT2 genes (**Figure 3D**). Interestingly, expression of these genes seems to be "turned on" at positions which resemble a "saddle point" that manifests in both UMAP and hyperbolic embeddings (**Figure 3D**). Further, beyond this "saddle point", very few K tumor cells are observed, and the remaining transcriptional states beyond this point are predominantly derived from KP tumor cells (**Figure 1D-E, Figure 3D).** Functionally, S100A6 is known to bind P53 to coordinate cell division and apoptosis (A. J. Levine 1997) and has been extensively described as a tumor suppressor in lung adenocarcinoma (P. Li et al. 2019; T. Wang, Han, and Du 2021; Orre et al. 2007; De Petris et al. 2009). Together, this may suggest that expression of *S100a6* reflects oncogenic stress induced by loss of *Trp53* function. After this point, cells appear to transcriptionally diverge, and have lost many hallmarks of lung identity. Instead, they adopt a gastric-like state, marked by expression

of *Hnf4a,* which has been previously identified as a transcriptional consequence of *Nkx2-1* loss (Snyder et al. 2013) (**Figure 3E**). Thereafter, cells upregulate *Hmga2* and other markers associated with metastasis in lung adenocarcinoma and KP tumors (**Figure 3F**).

An important assumption made with these analyses is that transcriptional changes are continuous (Trapnell et al. 2014), which is generally not observed in settings of selective pressure or terminal differentiation. Although there are several branched trajectories predicted by Monocle3 (**Figure 2E**), unlike the primary trajectory that proceeds continuously from untransformed AT2 cells to metastatic-like *Hmga2*$^{+}$ cells, these branches are not predicted robustly and are quite sensitive to the parameters used for graph construction. Still, along this primary trajectory, cells are assumed to arise from cells that precede them. The stochastic nature of tumor progression would otherwise suggest that evolution along this trajectory for any given individual tumor is an inherently inefficient process, which likely underlies the variation in kinetics observed in the model. However, the reproducibility observed across many tumors in this dataset (**Figure 1G-H**) suggests that although there is transcriptional noise which is consistent with this stochastic behavior, these evolutionary trajectories macroscopically reflect deterministic behavior of tumor evolution. As such, the aggregate behavior produced from selection that would otherwise manifest as some form of a critical point is “smoothed” out across many individual tumors, and is assumed to satisfy this requirement for continuity.

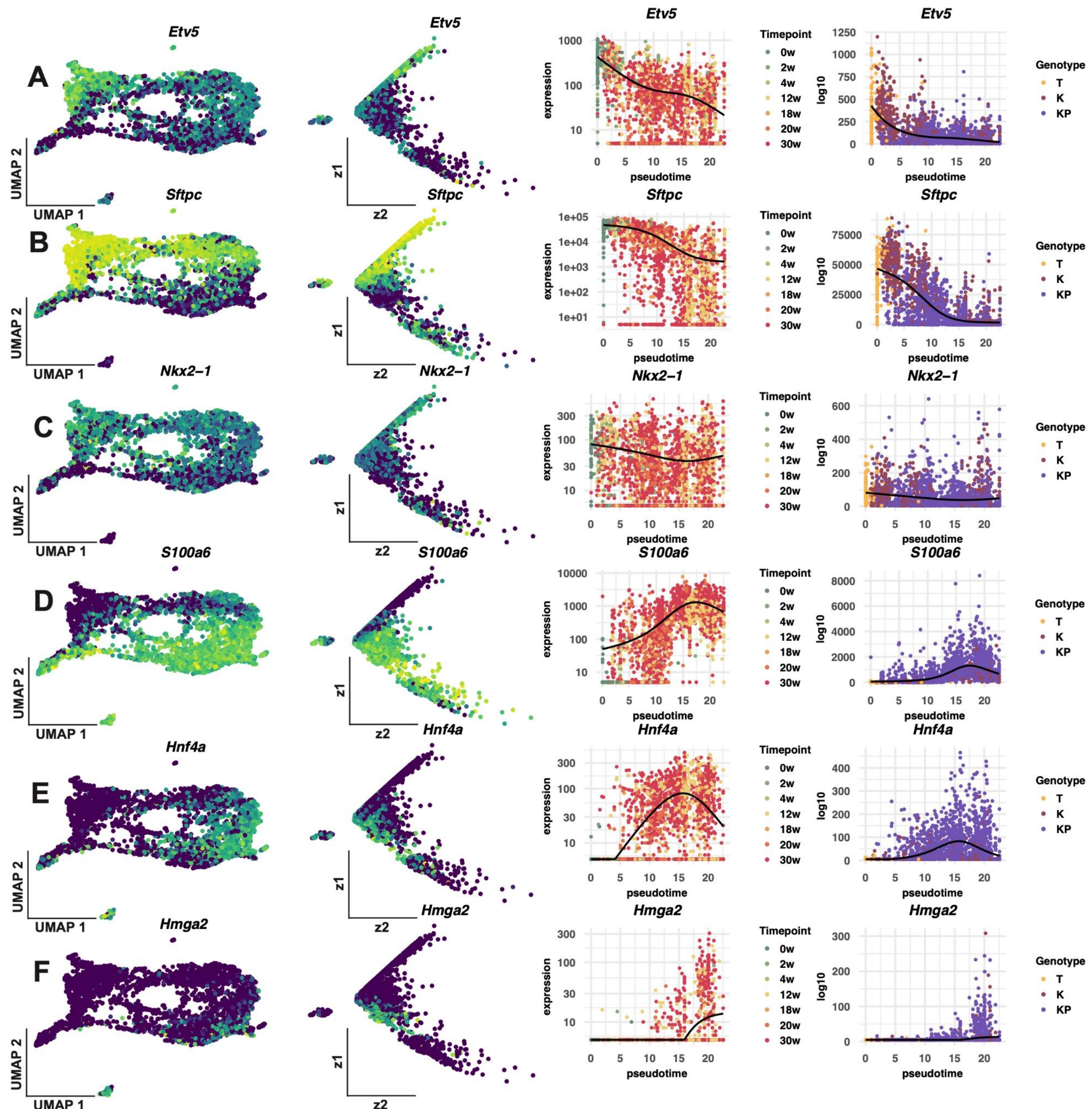
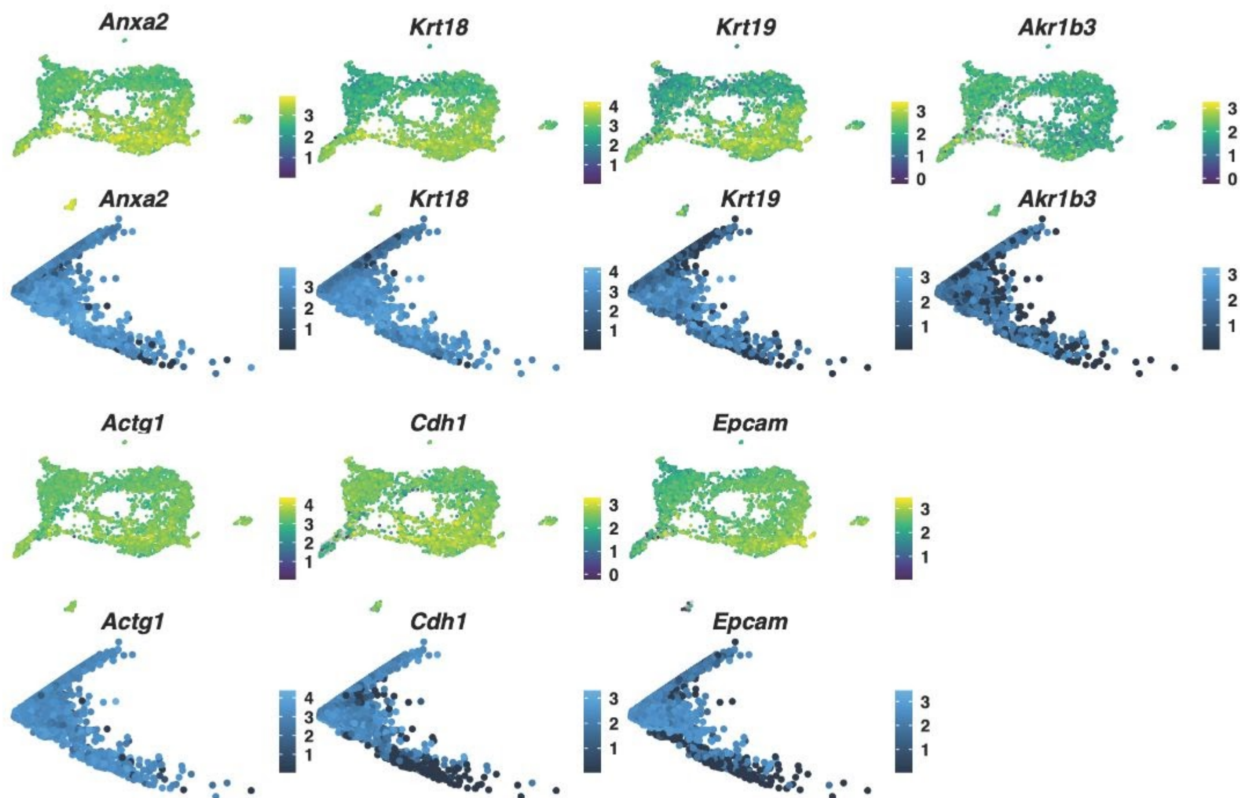


Figure 3. Sequential changes in cell identity occur throughout KP lung tumor cell evolution and progression.

Data used for these analyses has been previously published (Marjanovic et al. 2020) and were independently analyzed to produce these figures.

UMAP (**Column 1**, left) or hyperbolic scPHERE embedding (**Column 2**, second from the left) of log-normalized scRNA-seq gene expression of cell identity markers identified by Monocle3. Log-normalized expression is plotted against pseudotime across individual cells (**Column 3, Column 4**, right), with each point representing a cell that is colored by its respective time point (**Column 3**) or genotype (**Column 4**).

(**A**) *Etv5*, a regulator of AT2 identity maintenance.
(**B**) *Sftpc*, a marker of AT2 identity and functionality.
(**C**) *Nkx2-1*, a marker of lung identity that marks both AT2 and AT1 cells.
(**D**) *S100a6*, a regulatory partner of p53 associated with early AT1 identity. (A. J. Levine 1997; Beauchemin 2016)
(**E**) *Hnf4a*, a marker of gastric identity.
(**F**) *Hmga2*, a marker of metastatic-like KP tumor cells.



Supplemental Figure 3. Expression of epithelial markers associated with metastasis in non-small cell lung cancer.
Data used for these analyses has been previously published (Marjanovic et al. 2020) and were independently analyzed to produce these figures.
Log normalized expression is depicted in UMAP (Top) or hyperbolic (Bottom) embeddings. Genes are identified by (Jiewei Liu et al. 2012; Ruan et al. 2020).

Evaluation of Pea3 transcription factors in KP lung tumor progression.

In human lung adenocarcinoma, ETV4 transcript is expressed in NSCLC tumors, but not in the normal lung (Hiroumi et al. 2001), which is consistent with the patterns of *Etv4* expression observed in KP tumors (**Figure 4C**). In addition, ETV4 has been implicated in activation of the Rho pathway, which contributes to the metastatic potential of NSCLC (Hakuma et al. 2005). In contrast, *Etv5* is required to maintain alveolar type II cell identity in the context of oncogenic *Kras*. Loss of *Etv5* is also known to promote change in chromatin accessibility of AT2 identity genes, and AT2 cells deficient for *Etv5* partially adopt alveolar type I cell identity (Z. Zhang et al. 2017). The role of Etv1 in lung cancer is less well described, but has been shown to promote stromal expansion and tumor cell metastasis in pancreatic cancer ("ETS-Transcription Factor ETV1 Regulates Stromal Expansion and Metastasis in Pancreatic Cancer" 2016). Additionally, Etv1 has been shown to be regulated by miRNAs that are frequently dysregulated in lung cancer (X. Jin et al. 2017).

In consideration of their extensively described roles during lung morphogenesis and development, we sought to further investigate the role of the *Pea3* transcription factors in KP lung tumor progression. CRISPR/Cas9 technology has somewhat recently been extended to genetically engineered mouse models and to the KP model in particular (Sánchez-Rivera et al. 2014), which allows for generation of rapid genetic knockout of targeted genes by introducing sgRNAs alongside *Cre* recombinase and Cas9 to initiate KP lung tumors. In later generations of this technology, *Cas9* is expressed in the lungs of KP mice via a Cre-conditional Cas9 transgene into the Rosa26 locus (Platt et al. 2014) to allow for targeted knockout of genes at the onset of

tumor initiation through viral delivery of an sgRNA and *Cre* recombinase alone (pUSEC, **Figure 4D**) (Romero et al. 2020). An important limitation to this system is that it is largely dependent on utilization of lentiviruses, which transduce a wide range of cell types, but cannot accommodate AT2 specific promoters for *Cre* recombinase. As such, because *Etv1* is expressed in many cell types of the normal lung, is not restricted to alveolar cells, and is known to drive meaningful changes in surrounding stromal cells that can affect tumor progression (Thul et al. 2017), we have chosen to narrow our focus to *Etv4* and *Etv5*.

To model the impact of *Etv4* and *Etv5* in KP lung tumor progression, I utilized $Kras^{LSL\text{-}G12D/+}$,$Rosa26^{LSL\text{-}tdTomato/LSL\text{-}Cas9}$;$Trp53^{fl/fl}$ ($KP^{Cas9/Tomato}$) mice to knockout *Etv4* and *Etv5* when initiating tumors. In this schematic, expression of *Cas9* and *tdTomato* are induced by Cre recombination. Importantly, because the LSL-Cas9 allele is in the germline of KP mice, introducing Cas9 in this manner is not believed to be antigenic, which addresses a vital limitation of the earlier generation *in vivo* CRISPR technologies (Crudele and Chamberlain 2018). Expression of tdTomato further aids in isolation of individual tumors under the dissecting microscope for further analysis. Multiple sgRNAs were used to target *Etv4* (**Figure 4E**) and *Etv5* (**Figure 4F**), primarily in the DNA binding ETS domain, to produce loss-of-function frameshift mutations through non-homologous end joining and error-prone double stranded DNA damage repair. The selected guides were validated to produce efficient knockout of *Etv4* and *Etv5 in vitro* using KP tumor cell lines and assessed with TIDE via Sanger Sequencing (Brinkman et al. 2014).

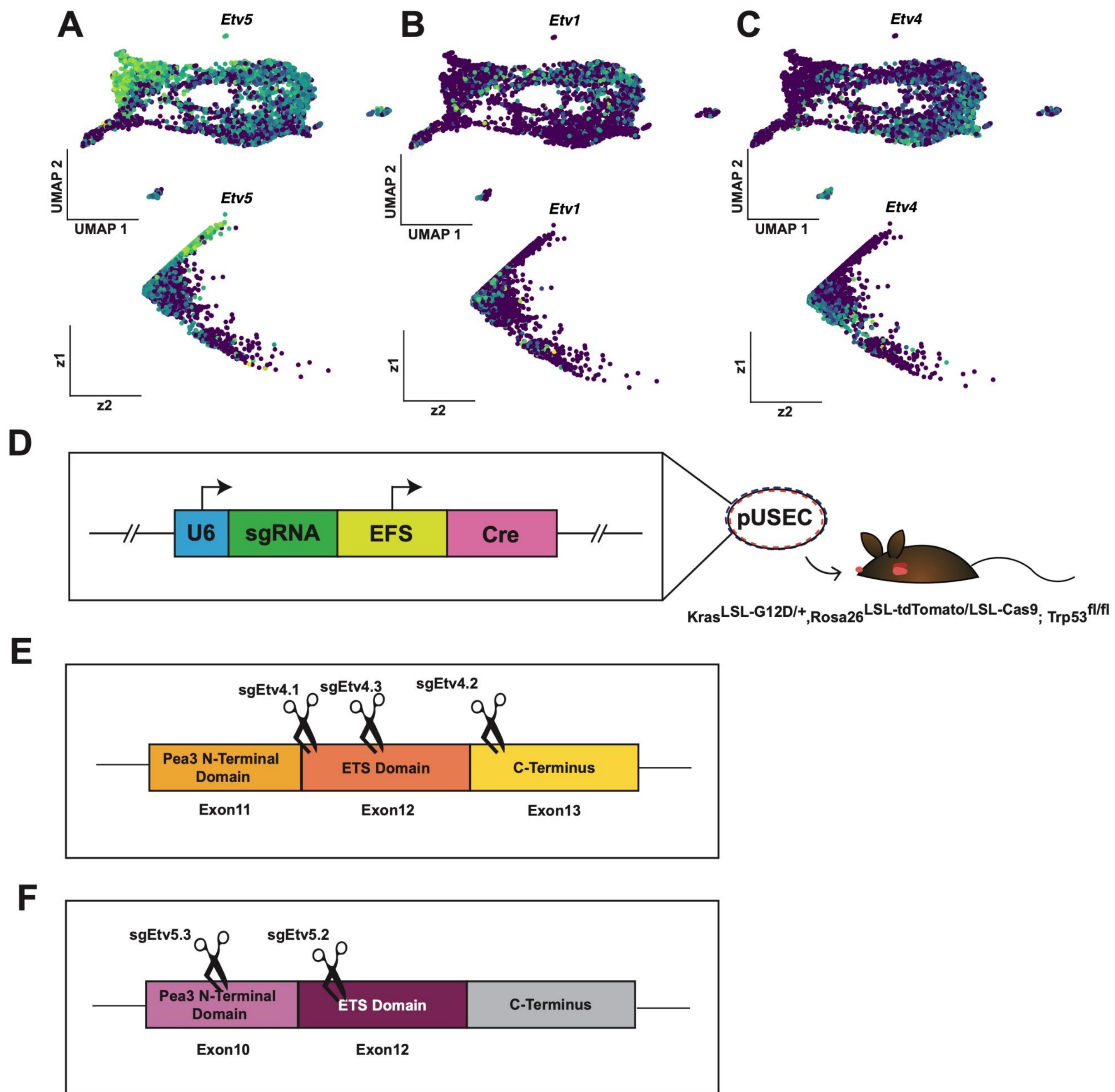


Figure 4. Evaluation of Pea3 transcription factors in KP lung tumor progression. Data used for these analyses has been previously published (Marjanovic et al. 2020) and were independently analyzed to produce these figures.
(**A**) UMAP (top) or hyperbolic scPHERE embedding (bottom) of log-normalized scRNA-seq gene expression of Pea3 transcription factors, *Etv5*, *Etv4*, and *Etv1*.
(**B**) Lentiviral CRISPR construct, pUSEC, designed to deliver Cre and sgRNAs to the distal lung epithelium of mice harboring a Cre conditional Cas9 transgene ($Kras^{G12D/+}$, $Rosa26^{LSL\text{-}tdTomato/LSL\text{-}Cas9}$; $Trp53^{fl/fl}$). Previously described by (Romero et al. 2020).
(**E-F)** Visualization of sgRNA targeting locations for (**E**) *Etv4* and (**F**) *Etv5*.

***Etv5* is required for KP lung tumor initiation *in vivo* but not *in vitro*.**

We attempted to generate *Etv5* knockout (KO) tumors using the pUSEC system in KP$^{Cas9/Tomato}$ mice (**Figure 4D**), utilizing an sgRNA targeting Olfr102 that was previously shown to behave neutrally in KP tumor cells (Romero et al. 2020). In the first cohort of mice used for this study, tumors were harvested at 13 and 17 weeks post infection. The loci targeted by *Etv5* sgRNAs was amplified from genomic DNA isolated from plucked tumors by polymerase chain reaction (PCR), and then further analyzed by Sanger sequencing and TIDE analysis. Although tumors did arise in animals that received virus containing sgRNAs against *Etv5*, invariably, all plucked tumors were wild-type for *Etv5* (data not shown).

In a second cohort, mice were sacrificed uniformly at 26 weeks post-infection, but again, no editing was observed at the *Etv5* locus in any of the resulting lesions. We reasoned that the selection bias imparted through plucking tumors, namely selection of tumors that are large enough to pluck, may explain why no Etv5 edited tumors were observed. To determine ETV5 status in lesions too small to pluck, we validated that ETV5 can be selectively stained for via immunohistochemistry (IHC) in KP lung tumor tissue (**Figure 5A**). In KP tumors, ETV5 staining is nuclear, but variable in intensity across tumor cells, reflecting heterogenous expression, which given findings in **Figure 4A**, was expected. This was observed across all mice, irrespective of the sgRNA delivered in pUSEC. No tumors were observed that clearly and uniformly did not express ETV5. We reasoned that it could be possible loss of ETV5 produces small lesions that cannot be plucked, and larger lesions produced from Etv5 targeting lentiviruses may be a consequence of tumor cells escaping CRISPR editing of ETV5. To

further investigate this possibility, we sought to determine whether a relationship exists between the size of a lesion produced from a lentivirus targeting ETV5 and the percentage of cells within it that express ETV5. If *Etv5* edited lesions grew more slowly than control lesions, this would suggest that the absence of CRISPR-edited *Etv5* sequences with DNA sequencing is attributable to selection bias of lesions that can be plucked. We quantified relative tumor areas by determining the ratio of total area with normal lung tissue to that of tumor tissue, and further quantified the number of cells in each tumor that had positive nuclear staining of ETV5 (**Figure 5B**). However, no correlations were observed between relative tumor area and ETV5 staining, and there appeared to be no difference in the distribution of tumor sizes across lesions produced from *Etv5*-targeting lentiviruses compared to control lesions. To further examine whether the fraction of cells in each tumor expressing ETV5 varied as a consequence of transduction with pUSEC-sgEtv5, the percentage of tumor cells expressing ETV5 was compared across viruses. While a minor difference was observed between control animals and those receiving sgEtv5, unexpectedly, slightly fewer ETV5+ cells were observed in control animals than sgEtv5 (**Figure 5C**).

In KP tumors, *in vivo* CRISPR editing efficiency is known to be sensitive to selective pressures (Rogers et al. 2018). As such, we hypothesized that selection against *Etv5* loss occurred in the pUSEC system. To circumvent these issues, we bred a Cre conditional *Etv5* floxed allele to KP mice (Z. Zhang et al. 2009) and initiated tumors in KP *Etv5*$^{fl/fl}$, KP *Etv5*$^{fl/+}$, and KP *Etv5*$^{+/+}$ animals with AdSPC-Cre to determine the effect of *Etv5* KO. Mice were sacrificed at 18 weeks post infection. In *Etv5*$^{fl/fl}$ mice, across nearly all mice, very few lesions were found in the lung (**Figure 5D**). In contrast,

many tumors were observed in *Etv5*$^{+/+}$ animals. Mice heterozygous for Etv5 showed an intermediate tumor burden in comparison to *Etv5*$^{+/+}$ animals, which demonstrates a clear dose-dependent phenotype that results from loss of *Etv5* (**Figure 5D**). Knockout of Etv5 affected both the observed tumor number and tumor burden, which suggests that Etv5 is required for transformation of KP tumors *in vivo*.

To further characterize the nature of the selective pressure that acts against *Etv5* KO cells, we utilized a recently developed alveolar organoid culture model (Naranjo and Cabana, Manuscript in preparation) that can be used to grow untransformed AT2 cells in 3D culture that can then be transformed *in vitro* to study early stages of KP transformation. With this system, we expanded untransformed AT2 organoids from KP mice that did not harbor any additional Cre-conditional alleles. In this regime, CRISPR/Cas9 technology can be transiently introduced into cells via adenoviral transduction. A modification of the pUSEC construct (**Figure 4D**) was made that replaces Cre recombinase with EGFP, pUSEG and was used to make lentivirus to deliver sgRNAs to organoids that become stably expressed alongside EGFP. Staggered transduction of adenoviral Cre, adenoviral Cas9, and Lentiviral pUSEG thereby permits the ability to precisely control transformation and Cas9 editing events.

When untransformed KP AT2 cells were simultaneously transduced with Adenoviral Cas9 and pUSEG-sgEtv5.2, the genomic locus targeted by sgEtv5.2 was unedited across multiple passages, despite retaining expression (>90%) of pUSEG, as assessed by FACS analysis of EGFP+ cells. In contrast, when KP AT2 cells were simultaneously transformed with adenoviral Cre, adenoviral Cas9, and lentiviral pUSEG-sgEtv5.2, a dominant in-frame deletion (-9) was observed at an estimated

frequency of 69.2% after two passages. When transformation was decoupled from Cas9 editing by first transforming KP AT2 cells with adenoviral Cre and then subsequently transducing them with lentiviral pUSEG-sgEtv5.2 and adenoviral Cas9, a +1 frameshift mutation was observed after the first passage at an estimated frequency of 51.1%, and after two passages, 24.8% (**Figure 5E**). Together, these results suggest that *Etv5* can be efficiently knocked out during early KP transformation *in vitro* but not *in vivo*. Notably, widespread loss of ETV5 does occur in overt KP adenocarcinomas during normal KP tumor progression (**Figure 5F**).

The primary differences between settings in which *Etv5* editing can or cannot be conducted efficiently are the environments of KP *Etv5* KO cells. At early stages of transformation *in vivo*, transformed tumor cells are believed to be subject to intense selective pressure from cells of their microenvironment, including immune cells. In contrast, when cultured and transformed *in vitro*, the environment of KP *Etv5* KO cells is experimentally defined, and *Etv5* KO is more efficient (**Figure 5E**). ETV5 is likely to be dispensable in more advanced tumors, as expression of ETV5 is widely lost in many of these tumors (**Figure 5F**). It is possible that the microenvironment produced by higher grade tumors elicit conditions that allow cells to tolerate loss of ETV5. However, the selective pressures that result in elimination of *Etv5* KO cells are unlikely to come entirely from the tumor microenvironment, as loss of *Etv5* is not well tolerated *in vitro* when KP cells are not transformed. Thus, the role of *Etv5* is dependent on the context of the tumor microenvironment and tumor progression.

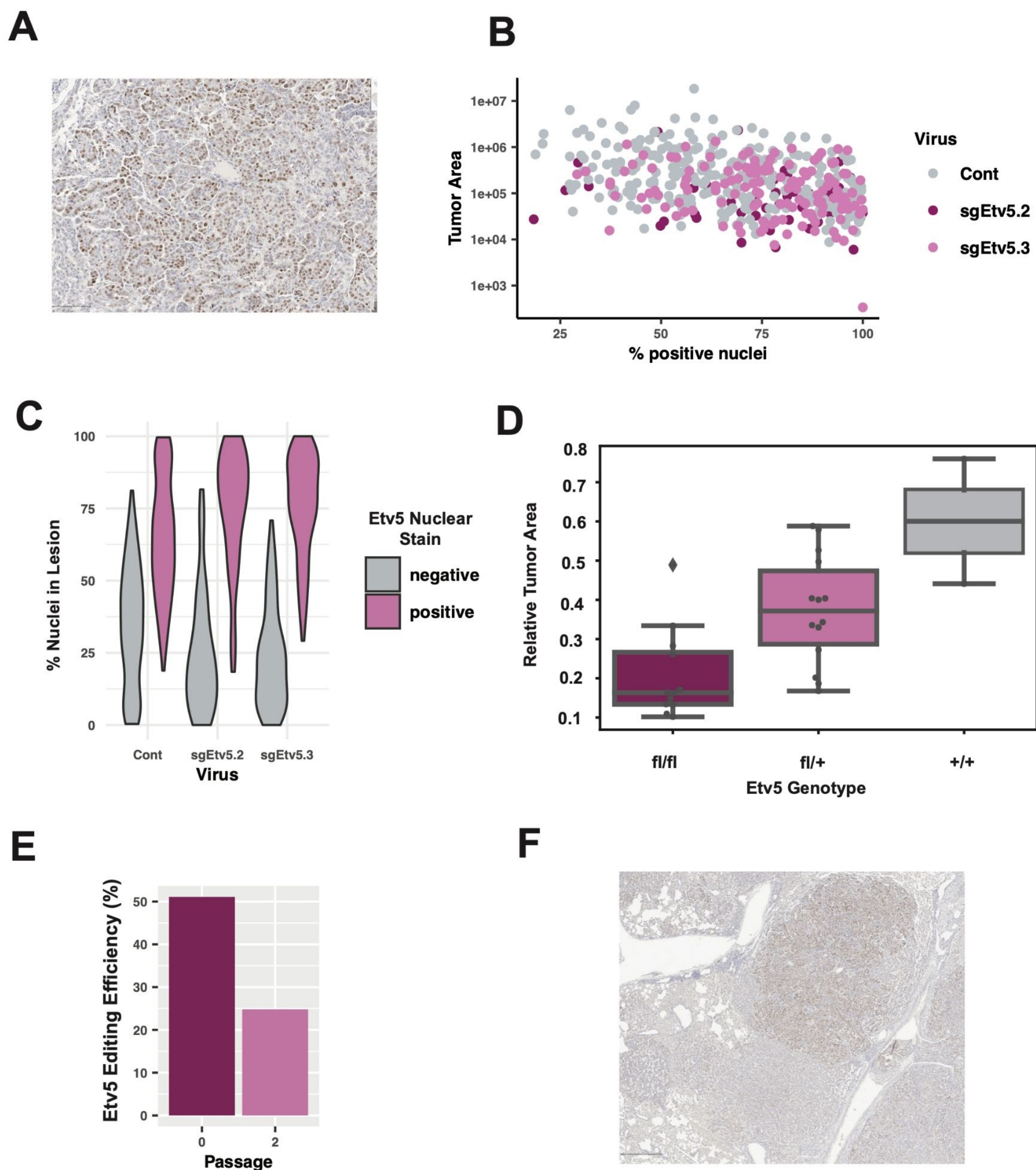


Figure 5. Etv5 is required for KP lung tumor initiation *in vivo* but not *in vitro*.
(**A**) IHC of ETV5 in KP lung tumors validates expression of *Etv5*.
(**B**) Expression of ETV5 in KP lung tumors does not correlate with tumor size. For each individual tumor in sections of KP lung tumors, the percentage of nuclei staining positively for ETV5 are plotted against its tumor area relative to normal lung tissue.

(**C**) CRISPR mediated knockout of *Etv5* does not meaningfully affect expression of ETV5.
(**D**) Cre-recombinase mediated knockout of *Etv5* suggests that *Etv5* is required for formation of KP lung tumors. Relative lung tumor area compared to normal surrounding lung tissue area is plotted for mice with homozygous or heterozygous knockout of *Etv5* compared to control mice wild-type for *Etv5*.
(**E**) Knockout of *Etv5* during transformation of KP lung cells can be achieved *in vitro* via CRISPR-mediated knockout of *Etv5 (*sgEtv5.2). Approximate proportions of edited allele frequency are shown at the beginning of the experiment (passage 0) and at the end (passage 2).
(**F**) Validation that ETV5 expression is lost in normal KP lung tumor progression and is dispensable at later stages of disease progression. IHC of ETV5 in KP lung tumors validates expression of *Etv5*.

The ETS domain of Etv4 is likely required for KP lung tumor initiation in vivo.

Unlike *Etv5*, *Etv4* can be knocked out when sgRNAs targeting *Etv4* are introduced during transformation by pUSEC (**Figure 4D**), again alongside a neutral control guide targeting Olfr102. Mice were sacrificed at 20 weeks after tumor initiation, and the loci targeted by *Etv4* sgRNAs (**Figure 4E**) was PCR amplified from genomic DNA isolated from plucked tumors, sequenced, and then analyzed by TIDE as described above. The overall editing efficiency in each tumor was dependent on the sgRNA. Approximately half of the tumors generated with pUSEC-sgEtv4.1, which targets a region spanning the *Pea3* N-terminal domain and the DNA binding ETS domain, were wild-type at the Etv4.1 locus. The sgEtv4.1 tumors that were edited for *Etv4* had a relatively high overall editing efficiency, and a mixture of mutations were observed across these tumors (**Figure 6A**).

Similarly, the tumors generated with pUSEC-sgEtv4.2, which targets the C-terminus of ETV4, had a higher overall editing efficiency (**Figure 6B**). In some of these edited tumors, a dominant mutation is clearly observed (e.g. G33), but in others, a greater diversity of mutations is observed (e.g. G37). Still, some tumors without mutations at the Etv4.2 locus were still observed.

In contrast to the tumors generated with pUSEC-sgEtv4.1 and sgEtv4.2, out of the nine mice infected with pUSEC-sgEtv4.3, only one lung lesion was observed (**Figure 6C**). As a result, genotyping of transformed KP sgEtv4.3 KO cells was not possible. Importantly, out of the three sgRNAs targeting *Etv4*, sgEtv4.3 most directly targets the DNA binding ETS domain. As such, it is not entirely surprising that the phenotypes generated from sgEtv4.3 have greater penetrance than those of sgEtv4.1

and sgEtv4.2. This manifests most apparently in the relative tumor areas of lungs harvested from these mice (**Figure 6D**), in which mice that received sgEtv4.3 uniformly had the lowest relative tumor area. The relative tumor areas of lungs from mice that received sgEtv4.1 or sgEtv4.2 appear to be bimodally distributed; approximately half of the sgEtv4.1 mice and one quarter of the sgEtv4.2 mice had comparatively low relative tumor areas (< 5%) (**Figure 6D**). The sgEtv4.1 guide targets the sequences connecting the most N-terminal Pea3 domain with the ETS domain, and produced a phenotype seemingly intermediate to that of sgEtv4.2 and sgEtv4.3. Notably, the mutations produced with the sgEtv4.1 guide predominantly produced deletion mutations. The remaining fractions of sgEtv4.2 and sgEtv4.1 mice had very similar relative tumor areas as control sgOlf102 mice.

Interestingly, out of the three methodologies used to assess selective pressures from knockout of Etv5, when these assays were performed with sgEtv4.1, no editing was observed amongst any of the three schematics. As such, further work is required to assess the role of Etv4 in KP lung tumor evolution. It is extremely important to note that no commercially available Etv4 antibodies have been robustly validated. Intriguingly, the antibody utilized in the Human Protein Atlas appears to selectively stain the nucleolar regions of cells, and there is emerging published data that supports this staining pattern using different Etv4 antibodies (Cosi et al. 2020). Still, in our hands, no commercially available antibodies were able to robustly and selectively stain for ETV4 in KP lung tumors. In addition, no commercially available Cre-conditional *Etv4* KO mouse strains exist and further limit the tools available to study the role of *Etv4* in KP lung tumor evolution.

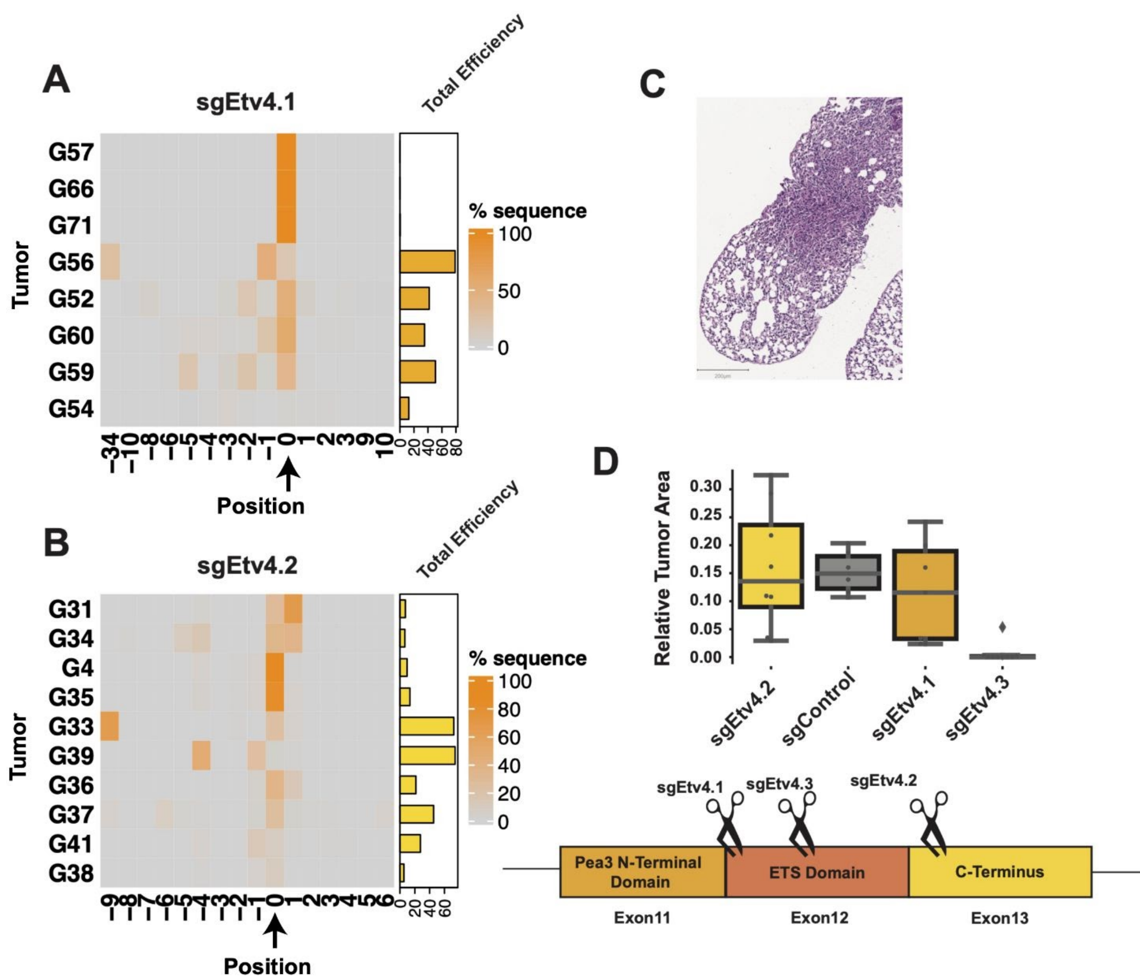


Figure 6. The ETS domain of *Etv4* is required for KP lung tumor initiation *in vivo*.
(**A-B**) Heatmaps depicting estimated allelic distribution of CRISPR mediated Etv4 mutations in plucked KP lung tumors, produced by Sanger Sequencing followed by TIDE analysis. Overall editing efficiency is denoted for each tumor in bar graph annotations on the right of the heatmap. A position of "0" is a wild-type sequence, denoted by the black arrow. (**A**) Mutations produced by sgEtv4.1, targeting the *Pea3* N-terminal domain. (**B**) Mutations produced by sgEtv4.2, targeting the C-terminal region of *Etv4*.
(**C**) Representative images of the rare histological lesions observed in mice infected with virus containing sgEtv4.3. Scale bar is 200 uM.
(**D**) Relative tumor areas of lung sections from mice infected with pUSEC-Etv4 viruses: sgEtv4.1 (n = 8 mice), sgEtv4.2 (n = 8 mice), sgEtv4.3 (n = 9 mice) compared to sgControl (n = 6 mice). (**C**) CRISPR mediated knockout of *Etv5* does not meaningfully affect expression of ETV5.

(**D**) Cre-recombinase mediated knockout of *Etv5* suggests that *Etv5* is required for formation of KP lung tumors. Relative lung tumor area compared to normal surrounding lung tissue area is plotted for mice with homozygous or heterozygous knockout of *Etv5* compared to control mice wild-type for *Etv5*.

DISCUSSION

In the KP model, at a given time point, the tumors of the lung have a spectrum of histological grades that change as time progresses, with lower grade tumors being most abundant at early time points, and higher grade tumors more abundant at late time points (Jackson et al. 2005). This heterogeneity has made it possible to comprehensively profile KP lung tumor evolution across limited time points that span many weeks. As more cells were collected from different mice and tumors, very few transcriptional states emerged that were not previously captured from existing data, which leads us to believe that this dataset captures the majority of transcriptional states that arise in primary KP lung tumor progression. The reproducibility observed in tumor evolution across tumors (**Figure 1G-H**) and mice (**Figure 1I-J**) is striking. If tumor cells truly became entirely dysregulated, it would be more likely to observe stochastic trajectories of tumor cell states as tumors progress over time. Instead, cell states convergently evolve into states associated with metastasis as well as *Hmga2* and *Runx2* expression (**Figure 2A**). However, these interpretations do not explicitly consider terminal trajectories within the confines of the primary tumor, or if it is possible for them to regress to other cell states observed in tumor evolution.

The pseudotime trajectories predicted by Monocle3 in UMAP space as well as connectivity in hyperbolic space (**Figure 2E**) suggest that a small population of cells will begin to express *Nkx2-1* and many other markers of AT2 cell identity after previously

silencing it and expressing *Hmga2* (**Figure 3B**). Importantly, *Nkx2-1* and *Hmga2* still exhibit mutually exclusive patterns of gene expression amongst this group of cells (**Figure 3C, 3F**). This interpretation is supported by the fact that the majority of the cells in this group are derived from tumors harvested at late timepoints (30 weeks) (**Figure 3F**). However, this proposed model of tumor evolution is largely generated by predictions, and further experimentation will be required to generate empirical evidence for these gene expression patterns. If true, many questions come into focus; in particular, it is not clear why it would be beneficial for these tumor cells to re-adopt hallmarks of their original identities in advanced and metastatic tumors. It is possible this occurs as a consequence of selection, but it is not immediately clear why such a selective pressure exists.

Many of the late-stage tumor cells that appear to express *Nkx2-1* after having expressed *Hmga2* that are most apparently distinguished from other tumor cells in hyperbolic space were assigned to Cluster 9 of (Marjanovic et al. 2020). In their work, they annotate this cluster as ‘biosynthetic mixed activity’. However, this annotation seems to be driven by the cells assigned to Cluster 9 that are separated from those that express *Nkx2-1* in hyperbolic space. Interestingly, these cells express many genes that have been proposed to be important in non-small cell lung cancer metastasis (Jiewei Liu et al. 2012; Ruan et al. 2020) (**Supplemental Figure 3**). Amongst these genes, Krt18 expression has been previously proposed to be predictive of lymph node metastasis (H. Zhang et al. 2014). Krt18 has also been found to be predictive of metastasis in colorectal cancer (Jingfeng Zhang, Hu, and Li 2019) and has further been

shown to be predictive of unfavorable clinical outcomes in both lung and pancreatic cancer in the human protein atlas (Uhlén et al. 2015; Thul et al. 2017).

We largely assume that the observed tumor cell states are reflective of the selective pressures that a cell experiences. There is extensive evidence that supports this assumption, but if it were the sole determinant of cell state, we would expect that circulating tumor cells (CTCs) across different kinds of cancer would have similar gene expression profiles because they experience similar selection pressures after extravasating from primary tumors. Empirically, there are some generalized similarities in CTCs across different cancers, but there are also many differences (Ruan et al. 2020). Taken together, these speculations further suggest the developmental context of a tumor cell must be at least partially a determinant of cell states in different microenvironments outside of the primary site.

If cells assigned to Cluster 9 do indeed represent cells that are poised to become circulating tumor cells, this study of KP Lung Tumors supports the observations made in clinically derived non-small cell lung cancer CTCs (Ruan et al. 2020). Subpopulations of tumors with intrinsic propensity to become a CTC has been long theorized and is most often discussed as a “pre-metastatic niche” (Kaplan et al. 2005; Eccles and Welch 2007) in primary tumors. TThe KP tumor cells profiled in this study are derived from primary tumors; as such, these cells are unlikely to represent those that have already entered the bloodstream. Instead, the similarity in the gene expression profiles between these cells and circulating tumor cells can be interpreted to suggest that Cluster 9 cells may represent those poised to intravasate into the bloodstream and become circulating

tumor cells. However, additional experimental evidence that compares transcriptional states of empirically identified CTCs to primary tumor cells are required.

In contrast with the findings published by (Marjanovic et al. 2020), the analyses that are presented in this chapter do not appear to provide supporting evidence for a singular high-plasticity stem cell state, annotated as Marjanovic cluster 5 (**Figure 2B-C**). Although (Marjanovic et al. 2020)suggest that cells that correspond to Cluster 5 and Cluster 9 are observed ‘in every tumor after adenomas’, while it is possible these cells are indeed present at all time points, these states may predominantly arise in late stages of KP tumor progression but may be infrequently found in early stage tumors. This could be attributed to the extensive heterogeneity of tumor progression kinetics observed in K and KP tumors (Jackson et al. 2005; E. L. Jackson et al. 2001). Furthermore, the data structures depicted in **Figure 1** do not necessarily support the idea that Cluster 5 or Cluster 9 can “seed” all of the other tumor cell states (Marjanovic et al. 2020). However, they do support the fact that Cluster 5 is highly connected with many other cell states and clusters.

Many studies have characterized transcriptional networks and signaling pathways in a general sense, particularly in the context of cancer. In development, it is well known that these pathways also regulate differentiation and produce vastly different transcriptional outcomes in different tissues. In cancer, because tumor cell identity often becomes dysregulated, this hinders the ability for us to understand how these pathways contribute to tumor evolution beyond functional evaluation of biological outcomes such as cell proliferation and migration. In particular, although loss of *Nkx2-1* expression has been shown to induce gastric programs in the context of KP lung adenocarcinomas

(Snyder et al. 2013) and strongly correlates with metastatic spread (Winslow et al. 2011), the effects of experimental loss of *Nkx2-1* do not fully recapitulate transcriptional profiles of poorly differentiated, metastatic tumors (Snyder et al. 2013; Winslow et al. 2011). Of the genes that differ in expression between cells that do not metastasize and those that are likely to metastasize, changes in expression of *Etv4* and *Etv5* occur independently of *Nkx2-1* loss, as assessed through shRNA knockdown of Nkx2-1 (Winslow et al. 2011).

Etv4 and *Etv5* belong to the ETS transcription factor family, which is distinguished by a highly conserved DNA binding domain and is also one of the largest evolutionarily conserved transcription factor (TF) families (Sizemore et al. 2017). ETS TFs are expressed across many tissue types but do exhibit some degree of tissue specificity and are generally known to have non-redundant biological roles. Interestingly, nearly all members have the ability to bind the ETS motif, and it is believed that this tissue specific activity is conferred through other functional protein domains that mediate interactions with other proteins rather than the ETS binding domain itself (Findlay et al. 2013). Importantly, the context-specificity of ETS transcription factor activity complicates interpretations of chromatin accessibility at ETS consensus motifs, and dysregulated transcriptional activity of aberrantly expressed ETS transcription factors may not manifest as readily in sciATAC data as has been observed for scRNA-seq data (LaFave et al. 2020). Consensus motifs that appear to be nonredundant ETS sites are more often accompanied by atypical, low affinity ETS binding site sequences that are proximal to other transcription factor binding sites (Wheat et al., 1999, Hollenhorst et al., 2011b). Still, extensive evidence exists that

implicates the ETS family in the development of cell states that promote cancer progression across multiple cancer types, both through changes in expression and mutation (Sizemore et al. 2017).

The overlapping expression patterns of *Etv4* and *Etv5* have historically been discussed in the context of embryonic development and AT2 biology as evidence for functional redundancy of the two genes, which was further supported by their involvement in similar biological processes and phenotypes (Herriges et al. 2015; Z. Zhang et al. 2017). Specifically, when originally characterized, differential expression of the *Pea3* subfamily of the ETS family across the embryos of mice suggested these transcription factors play non-redundant roles (Chotteau-Lelièvre et al. 1997). However, the gene expression patterns suggested by the data presented in this chapter provides evidence that KP tumor progression is an example of a setting in which these genes have distinct expression patterns and distinct biological roles (**Figure 4A**).

In this study, we show that *Etv5* expression is frequently downregulated during KP lung tumor progression and precedes loss of *Nkx2-1* (**Figure 3A**). It has previously been shown that in the context of oncogenic Kras^{G12D}, loss of *Etv5* in AT2 cells has been previously shown to generate a hybrid cell state characteristic of a distal progenitor and AT1 cell (Z. Zhang et al. 2017). In the same study, it was also discovered that *Etv5* is required for Kras^{G12D} mediated transformation and the rare lesions that do appear in *Kras*^{G12D} *Etv5*$^{fl/fl}$ animals have evaded Cre recombination (Z. Zhang et al. 2017). Interestingly, lesions which were heterozygous for *Etv5* upregulated *Etv4*.

Given that *Etv5* is lost during late stages of normal KP tumor progression, we reasoned that loss of *Etv5* would be tolerated in settings with deletion of Trp53 in

Kras^{G12D};Trp53$^{\Delta/\Delta}$ cells. Surprisingly, we discovered that, although *Etv5* is required for lung tumor initiation *in vivo* in KP mice, it can be deleted *in vitro*, but only when the cells are transformed (**Figure 5E**). *In vitro* culture of untransformed AT2 cells is performed in a defined media that is designed to support the growth and differentiation state of AT2 cells (Naranjo and Cabana, Manuscript in Preparation). As such, it is not surprising that loss of Etv5 in untransformed AT2 cells, which has been shown to destabilize AT2 identity, is not tolerated (Z. Zhang et al. 2017). It is particularly interesting to note that, although CRISPR-mediated deletion of Etv5 is possible *in vitro*, there was reasonably strong evidence that cells with Etv5 deletion were outcompeted by cells that have evaded CRISPR-mediated deletion of Etv5 (**Figure 5E**). Although the lung organoids remained transduced by pUSEG, as assessed by EGFP expression, the predicted frequency of *Etv5* frameshift +1 mutations decreased over two passages. An important caveat to these *in vitro* experiments is that they were only repeated one time, and as such, further work is required to characterize these *in vitro* and *in vivo* selective pressures. If these conclusions are validated; it will be interesting to determine whether *Etv5* null lesions are not observed *in vivo* solely due to becoming outcompeted by *Etv5* competent cells, or if there are selective pressures unique to the *in vivo* microenvironment which drive selection against *Etv5* loss.

One of the most unexpected findings in this study was that, despite the fact that *Etv4* is not expressed in normal AT2 cells and does not appear to be expressed in cell states associated with early stages after transformation (**Figure 4A**), CRISPR-mediated loss of *Etv4* through targeting a locus within its ETS domain almost entirely blocks tumor formation (**Figure 6C-D**). Previously, it was shown by the Barbacid group by bulk

RNA-Seq that *Etv4* is upregulated in tumor cells very early after transformation in the *Kras*^{G12V} mouse model of lung adenocarcinoma (Mainardi 2013). Similar to our findings in this study, when one allele of *Etv4* was deleted, the number of overt tumors observed in the lung is strongly diminished (Mainardi 2013). Future studies utilizing lung organoid culture technology may provide further insight about the expression of *Etv4* at early stages of *Kras* mediated transformation in the lung.

Interestingly, in bulk tumors heterozygous for *Etv4*, *Etv5* expression is upregulated (Mainardi 2013), which was previously interpreted as additional evidence that *Etv4* and *Etv5* have at least some functional redundancy. However, this conclusion is not supported by the expression patterns observed for Pea3 transcription factors in single cells during KP lung tumor evolution (**Figure 4A**), in which all three genes have distinct expression patterns. It is likely that the resolution possible by assessing expression at the single-cell level afforded better resolution of the expression patterns of the Pea3 transcription factors during KP lung tumor progression. Furthermore, the selective pressures that act on cells that have lost *Etv4* are likely distinct from those that act on cells that have lost *Etv5*. When performing *in vivo* editing of *Etv5*, many lesions are found in mice receiving pUSEC-sgEtv5, but all of these resulting lesions are wild-type for *Etv5*. This is consistent with selective pressures observed against *Etv5* deleted cells *in vitro* (**Figure 5E**). In contrast, when targeting the ETS binding domain of *Etv4* with CRISPR/Cas9, cells were not able to evade CRISPR editing, and very few lesions were observed in animals receiving pUSEC-sgEtv4.3. In studies conducted by Mainardi et. al, loss of *Etv4* reduced but did not eliminate *Kras* lung tumor initiation (Mainardi 2013). However, the remarkable absence of large lung lesions in animals

receiving sgEtv4.3 provides evidence that loss of *Etv4* may have greater importance in *Kras* transformation than previously described. Still, however, the phenotypes generated with viruses containing other guides targeting *Etv4* at different loci of the gene that either less directly target the ETS domain of *Etv4* (sgEtv4.1) or that do not target the ETS domain (sgEtv4.2) were largely consistent with the conclusions made by Mainardi et. al.

Further work remains to be conducted to assess whether the differences observed in the phenotypes generated with these sgRNAs targeting *Etv4* are a consequence of differences in the functional targeting of *Etv4,* or whether these differences are merely observed due to guide-intrinsic differences in Cas9 targeting efficiency. Additionally, further comparisons between loss of *Etv5* may provide additional insight to the complex behavior of the *Pea3* transcription factors in KP lung tumor evolution. It is important to note that the sgRNAs were tested in an aggressive and metastatic KP lung tumor cell line and deemed to produce efficient knockout of *Etv4* or *Etv5 in vitro*, a setting that is very different from that of early transformed cells, which are modeled using an organoid culture system. Thus, it is possible that deletion of *Etv4* or *Etv5* may be tolerated *ex vivo*, once tumors have progressed and evolved sufficiently to tolerate deletion of these genes. As the tools for perturbing expression of genes in genetically engineered mouse models of cancer become more advanced, it will be worthwhile to determine whether loss of *Etv4* in the stages of tumor progression after early transformation restrict the ability of KP tumors to progress along this now characterized axis of tumor progression.

MATERIALS AND METHODS

scRNA-seq analysis

scRNA-seq data was analyzed primarily in R with Monocle3 (version 0.2.3.3) (Trapnell et al. 2014; Qiu et al. 2017), using integrated features of *Monocle3* to perform mutual nearest neighbor (MNN)-based batch correction using *Batchelor (Haghverdi et al. 2018)* and dimensionality reduction for pre-processing via Principal Component Analysis (PCA) (Haghverdi et al. 2018) and low dimensional embeddings via Uniform Manifold Approximation and Projection (UMAP) (Becht et al. 2018; McInnes et al. 2018), with default parameters. 100 principal components were used for nearest neighbor graph construction, however, it should be noted that UMAP embeddings and pseudotime and trajectory analyses produced highly similar results when less principal components were used. Clustering was performed with default parameters in Monocle3 at a resolution of 1E-5 prior to trajectory analysis.

Hyperbolic embeddings were performed with scPhere (Ding and Regev 2021) using default parameters in *Python 3.8*. Visualizations were then produced in R using ggplot.

Gene expression in single cells was calculated on a by-gene basis prior to visualization and log normalized prior to visualization. A threshold of 0.1 was used as a minimum expression value to filter lowly expressing cells as described in Monocle3.

Markers of cell clusters, genotypes, and timepoints were identified by the top_markers function in Monocle3. All marker genes were ranked by q-value, except gender specific markers, which were ranked by pseudoR2.

Highly variable genes were identified in Seurat using the “vst” method and otherwise default parameters.

Figures and Visualizations

Heatmaps and their annotations were produced in *R* (version 4.0.4) with *ComplexHeatmap* (Gu, Eils, and Schlesner 2016). Rows were hierarchically clustered by complete linkage of distance metrics that were employed in a use-case dependent manner, including euclidean distance and canberra distance.

Mean log gene expression was calculated for heatmaps with a 0.1 pseudocount and scaled using various parameters (below) for each use case. Methodologies are identical to that of the plot_genes_by_group function in Monocle3.

For genotype specific markers, scale_min = -3 and scale_max = 3.

For timepoint specific markers, scale_min = -5 and scale_max = 5.

For tumor highly variable genes, scale_min = -5 and scale_max = 5.

For timepoint highly variable genes, scale_min = -5 and scale_max = 5.

For batch highly variable genes, scale_min = -5 and scale_max = 5.

For pseudotime-dependent terms, scale_min = -8 and scale_max = 10.

Bar plots and box plots were generated with ggplot2 in R or seaborne in Python.

Lentiviral Production

Lentivirus was produced in HEK293T cells cultured maintained in DMEM (Corning #10-013-CV) supplemented with 10% fetal bovine serum. Cells were plated 1 day before transfection and then co-transfected with lentiviral constructs and packaging plasmids psPAX2 and pMD2.G (Addgene #12260 and #12259) with PEI or MirusLTE. Viral supernatant was harvested 48 and 72 hours after transfection, filtered through a

0.45 uM filter, concentrated at 25,000g for 2 hours at 4C, resuspended in optiMEM overnight, then frozen and aliquoted at -80°C.

Lentiviral Titering

Functional titering of lentiviruses was performed using Green-Go cells, which contain a Cre-responsive GFP cassette. Cells were seeded in a 24 well plate and then transduced 24 hours after plating with serial dilutions of concentrated lentivirus. The percentage of GFP+ cells was assessed on the Guava easyCyte BGR 48 hours after transduction, which was then used to calculate viral titers.

Lentiviral Vectors and sgRNA Cloning

pUSEC and pUSEG lentiviral vectors and cloning strategy is previously described (Sánchez-Rivera et al. 2014). sgRNA guide sequences were designed using GUIDES (Meier, Zhang, and Sanjana 2017) and the Broad sgRNA design tool (Doench et al. 2016; Sanson et al. 2018).

CRISPR Guide & Targeting Validation

Using guides targeting Etv4 and Etv5, we tested three sgRNAs per gene by cloning them into LentiCRISPRv2 as previously described (Sanjana, Shalem, and Zhang 2014). Lentivirus was then produced with LentiCRISPRv2 vectors, filtered, and then used to transduce KP1233 cells *in vitro*. 48 hours after transduction, cells were selected with Puromycin at 10 ug/mL. An additional 24 hours later, cells were collected and used to isolate genomic DNA.

Guides targeting Etv4 and Etv5 were assessed by amplifying gDNA with Q5 polymerase, using primers below, sanger sequenced, and analyzed by TIDE (Brinkman et al. 2014) with large decomposition windows and parameters that allow detection of

larger indel events when chromatogram and editing efficiency were inconsistent. Otherwise, if editing efficiency and chromatogram appeared consistent with default parameters, default parameters were used. Importantly, Kapa2G HotStart polymerase was also used for TIDE analysis of gDNA from tumor tissue, as Q5 performed poorly on gDNA from tissue.

gDNA was isolated from cells *in vitro* using the Qiagen puregene core kit A. gDNA was isolated from tumor tissue through homogenization of tissues in DNA extraction buffer ("DNA Isolation Buffer" 2019) with 3 uL of proteinase K per tumor overnight at 55C. Subsequently, phenol-chloroform DNA extraction was performed. gDNA with sufficient quality for further analyses were selected by purity assessed via nanodrop.

sgRNA sequences

Guide	**Target**	**Sequence**
sgEtv5.1	Etv5	GGGCCTCCTTATCAGAG ACG
sgEtv5.2	Etv5	CATAGTAATAGCGGAGA GAG
sgEtv5.3	Etv5	GTACTTTGATGATACTTG CG
sgEtv4.1	Etv4	GCCGGGGTGCCTTACAA CTG
sgEtv4.2	Etv4	GTTTGACCGGCCAGTCA GTG
sgEtv4.3	Etv4	CATAATAGTATCGCAGCG AG
sgOlfr102	Olfr102	CATCTTTGGCAGTGTCA CAG

PCR primers for TIDE analysis

Primer F	Primer R	sgRNA locus
AGTGAGAGGGTTGGCGATGT	AACAACCAGCATCGTACAAAACAA	sgEtv5.1
GTTACACGGCTACCCCAGGT	AGGTGAAACAGGCCTTGGCT	sgEtv5.2
ATGTGCCCTTGAACATGTCGTG	CGCCCGGACTCACCTCTTC	sgEtv5.3
AAGTCAGGACACTCGGGGAAGA	CTCCCTTCTGGTTCTTGTTCACGT	sgEtv4.1
TGTGGCCCATGAGAAGGGGA	ACAGGTGAGCCACAGCGAAC	sgEtv4.2
GCCTAGATTGTCCCCGCACC	TCCGGCTCGCACACAAACTT	sgEtv4.3

Animal Studies

KP mice were infected as previously described (DuPage, Dooley, and Jacks 2009). For pUSEC viruses, mice were infected with 10,000 Tu of virus. Mice infected with AdSPC-Cre, lot Ad4237, given 2.5E8 Tu. Mice were sacrificed at indicated time points after tumor initiation. Tumor tissue and lung tissue was fixed in zinc formalin overnight and then stored in 70% ethanol for a minimum of 24 hours before being processed by the KI histology core. All mice were bred and handled according to IACUC approved protocols.

Antibodies Used

An anti-Etv5 antibody, purchased from Abcam (ab102010) was used for both IHC and western blots. Western blots were performed at dilutions of 1:1000, and IHC was performed with 1:500 dilutions, both overnight at 4C.

Histology

Hematoxylin and eosin stain (H&E stain) was performed with a standard method by the Hope Babette Tang Histology Facility at Koch Institute.

ACKNOWLEDGEMENTS

We would like to acknowledge M.H, N.M., and T. T. for interpretation of scRNA-seq phenotypes; Megan Burger for technical assistance; Koch Institute's Robert A. Swanson Biotechnology Center for computing resources and histology; K. Yee, J. Texeira, K. Anderson, K. Mercer and M. Magendantz for administrative and logistical support.

This work was supported by the Howard Hughes Medical Institute (T.J.); the Ludwig Center at Harvard (S.S.); NIH U54 CA225088 (S.S.); NIH T32HL007627(G.G.); NCI Cancer Center Support Grant P30-CA1405; the Bridge Project, a partnership between the Koch Institute for Integrative Cancer Research at MIT and the Dana-Farber/Harvard Cancer Center (T.J. and S.S.); and fellowship awards. This work was supported in part by the NIH Pre-Doctoral Training Grant T32GM007287). T.J. is a David H. Koch Professor of Biology and Daniel K. Ludwig Scholar.

A.M.C., L.L., and T.J. designed the study. A.M.C, I.D, S.N., C.C., T.N., M.Y., and E.D. performed all of the mouse experiments in the lab of T.J.; A.M.C. performed the computational analyses associated with the scRNA-seq datasets.

CHAPTER 3

ANTIGEN DOMINANCE HIERARCHIES SHAPE TCF1+ PROGENITOR CD8 T CELL PHENOTYPES IN TUMORS

Megan L. Burger[1], **Amanda M. Cruz[1,2]**, Grace E. Crossland[1], Giorgio Gaglia[3,4,5], Cecily C. Ritch[3,4,5], Sarah E. Blatt[1], Arjun Bhutkar[1], David Canner[1,2], Tamina Kienka[1], Sara Z. Tavana[1], Alexia L. Barandiaran[1], Andrea Garmilla[1], Jason M. Schenkel[1,3], Michelle Hillman[1], Izumi de los Rios Kobara[1], Amy Li[1,2], Alex M. Jaeger[1], William L. Hwang[1,6,7], Peter M. K. Westcott[1], Michael J. Manos[8,9], Marta Holovatsk[8,9], F. Stephen Hodi[8,9,10], Scott J. Rodig[3,9], Aviv Regev[1,2,6,11], Sandro Santagata[3,4,5,12], Tyler Jacks[1,2]*†

[1]David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, 500 Main Street, Cambridge, MA 02139, USA

[2]Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

[3]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

[4]Ludwig Center at Harvard, Harvard Medical School, Boston, MA 02115, USA

[5]Laboratory of Systems Pharmacology, Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

[6]Broad Institute of MIT and Harvard, 415 Main Street, Cambridge, MA 02142, USA

[7]Department of Radiation Oncology, Massachusetts General Hospital, Boston, MA 02114, USA

[8]Melanoma Disease Center, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA.

[9]Center for Immuno-oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA.

[10]Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA.

[11]Genentech, 1 DNA Way, South San Francisco, CA 94080, USA

[12]Department of Oncologic Pathology, Dana Farber Cancer Institute, Boston, MA 02215, USA

*Corresponding authors: Tyler Jacks, tjacks@mit.edu
†Lead contact

AUTHOR CONTRIBUTIONS

M.L.B, A.M.C., G.E.C., A.B., G.G., R.H.H., D.C., J.M.S., A.L., A.M.J., W.L.H., F.S.H., S.J.R., S.S., A.R. and T.J. designed the study; M.L.B., G.E.C., T.K., S.T., A.L.B., A.G. and I.D.K. performed all of the mouse experiments in the lab of T.J.; A.M.C., S.E.B. and A.B. performed the computational analyses associated with the scRNA-seq datasets; M.L.B., G.E.C., W.L.H., M.H. and P.M.K.W. performed the mouse scRNA-seq experiment in the labs of T.J. and A.R.; G.G. and C.C.R. performed the t-CyCIF imaging in the lab of S.S.; M.J.M. and M.H. provided melanoma tissue samples and patient information.

Figure 1 was produced by M. L. B. and assembled with the help of A. M. C. Figure 2, Figure 3, Supplemental Figure 3, Figure 4, and Supplemental Figure 4 were produced with analyses performed by A. M. C. in collaboration with M. L. B, except Figure 2D, which was created using analysis produced by A. B. and S. E. B. All benchwork and mouse experiments were performed by M. L. B, G. E. C, S. T., and T. K. All Figures contain data originally published in Burger et al. 2021.

ABSTRACT

The responses elicited by CD8 T cells to different antigens occur simultaneously but are largely characterized in isolation. In mouse lung adenocarcinoma, we find that an antigen dominance hierarchy occurs when antigens with different pMHC binding properties are co-expressed in tumors. CD8 T Cells specific to the dominant antigen, which binds MHC with

the greatest stability, undergo larger clonal expansion compared to cells specific to the subdominant antigen, which binds MHC with less stability. Interestingly, T-cells specific to the subdominant antigen become enriched for a TCF1+ progenitor cell state, associated with positive response to immune checkpoint blockade (ICB) therapy, but do not preferentially benefit from ICB compared to T-cells specific to the dominant antigen. This is because the subdominant antigen response is differentially enriched for a dysfunctional population of TCF1+ cells marked by CCR6 and Tc17-like differentiation. However, this population can be eliminated through vaccination, which markedly improves the subdominant antigen response. These findings ultimately may inform a therapeutic strategy to elicit multi-faceted neoantigen responses in tumors.

INTRODUCTION

The immune response to solid tumors is largely driven by recognition and targeting of cells expressing mutant proteins that are presented to immune cells. Neoantigens are peptides produced by these proteins that are then presented on the surface of cells when loaded onto major histocompatibility (MHC) molecules (Schumacher, Scheper, and Kvistborg 2019). CD8 T cells are responsible for introducing cytotoxic stress in transformed cells and have otherwise been implicated as a population that expands in response to immune checkpoint blockade (ICB) therapy, specifically, anti-PD1 or anti-PDL1 (Pfannenstiel et al. 2019). Many forms of immunotherapy have been developed to target tumor-specific neoantigens in patients, including adoptive T cell therapies (E. Tran et al. 2016) and therapeutic vaccines (Hollingsworth and Jansen 2019).

Neoantigens that are targeted by immunotherapy are often selected computationally by predicting immunogenic epitopes from whole exome sequencing data (Peters, Nielsen, and Sette 2020; Wells et al. 2020). Importantly, however, the vast majority of these neoantigens fail to elicit a productive immune response, and reflect a deficit in the understanding of which neoantigens are relevant for immunotherapy as well as an understanding of the behavior of immune responses to them during immunotherapy (Schumacher, Scheper, and Kvistborg 2019). Weak immune responses to neoantigens could be due to T cell evasion mechanisms that create selective pressures against neoantigen expressing tumor cells, driving neoantigen loss or immunoediting. In any case, these complexities have made selection of therapeutically actionable neoantigens and characterization of the immune responses they elicit very limited.

Alternatively, weak neoantigen responses could be influenced by antigen immunodominance hierarchies that are established during the immune anti-tumor response (H. Schreiber et al. 2002). In the setting of acute viral infections, it has been previously shown that one or two immunogenic epitopes saturate the T cell response, thereby suppressing responses to other neoantigens (Yewdell 2006). Importantly, it was shown that T cells can respond to tumor neoantigens that do not drive immune responses by endogenous T cells (Strønen et al. 2016), which suggests that endogenous T cell responses to neoantigens are inhibited. Importantly, however, these suppressed responses can be rescued in melanoma via therapeutic vaccination (Carreno et al. 2015; Ott et al. 2017; Sahin et al. 2017). As such, it is possible that the immune responses to some neoantigens are suppressed as a consequence of

competition with responses to other neoantigens, which may provide a therapeutic opportunity to reinvigorate endogenous T cell responses through therapeutic vaccination. Still, however, the effect of immunodominance hierarchies has not been well characterized in the context of tumor immune responses.

The subset of CD8 T cells that expand in response to ICB in mice and humans have been previously characterized as a progenitor population within the CD8 compartment that is marked by expression of TCF1 (*Tcf7*) (Kurtulus et al. 2019; Q. Guo et al. 2019; Sade-Feldman et al. 2019; Siddiqui et al. 2019). These progenitor CD8 cells are of particular importance to immunotherapy because a subset of them have been described as capable of functionally differentiating into dysfunctional CD8 T cells (B. C. Miller et al. 2019a). In the context of dominance hierarchies of tumor neoantigens, this population of T cells has not been described. Further, the exact behavior of functionally differentiating progenitor CD8 T cells in response to ICB is poorly understood.

The $\underline{K}ras^{LSL\text{-}G12D}$;Tr$\underline{p}53^{fl/fl}$ ($\underline{P}$) model of lung adenocarcinoma, in which multi focal tumors of the lung are initiated through intratracheal delivery of lentiviral Cre (DuPage et al., 2009) can be harnessed to express experimentally defined neoantigens in a tumor-specific manner through incorporation of neoantigens to lentiviral Cre vectors (DuPage et al., 2011). In the absence of these defined neoantigens, the anti-tumor immune response is weak because KP tumor cells are transformed by experimentally defined mutations in *Kras* and *Trp53*, and otherwise do not frequently harbor mutations that can elicit a T cell response (DuPage et al., 2011). Expression of strong neoantigens in this manner can drive a productive immune response that eventually becomes diminished as tumors grow and adapt to this selective pressure (DuPage et al., 2011).

Importantly, this is not believed to occur as a result of dysregulating neoantigen presentation pathways or neoantigen loss (DuPage et al., 2011). Frequently, LucOS is expressed in tumor cells of KP mice to study neoantigen specific responses. LucOS contains two model CD8 T cell antigens, SIYRYYGL (SIY), which is a synthetic peptide, and SIINFEKL (SIIN), which is derived from chicken ovalbumin, that are both expressed through fusions to luciferase protein.

In the KP model, tumor progression occurs over the course of approximately 5 months, during which tumors progress through histological grades that faithfully recapitulate human clinical disease (Jackson et al. 2005; DuPage, Dooley, and Jacks 2009). The dynamic range afforded by this model permits longitudinal characterization of anti-tumor immune responses. Utilization of the model antigens SIIN and SIY also allow for longitudinal characterization of immune responses to specific neoantigens, because T cells specific for these antigens can be stained with tetramer (Dupage et. al, 2011). Further, tumor progression kinetics in response to expression of LucOS have been previously characterized (Dupage et. al, 2011). Importantly, expression of LucOS confers simultaneous expression of SIIN and SIY model antigens, thereby making the KP LucOS model system a viable experimental system to evaluate the contribution each neoantigen response has on the global anti-tumor immune response.

In this study, we find an antigen dominance hierarchy between SIIN and SIY that ultimately limits expansion of the subdominant SIY T cell response. Further, we find progenitor CD8 T cell populations are heterogeneous and differ between SIIN and SIY responses, and additionally describe a dysfunctional subset of progenitor cells that is abolished following therapeutic vaccination against SIIN and SIY.

RESULTS

Temporal dynamics of CD8 neoantigen response in Lung Adenocarcinoma

To characterize multi-modal neoantigen responses in the context of lung adenocarcinoma, we utilized a genetically engineered mouse model of lung adenocarcinoma in which lung tumors are generated *in situ* through lentiviral delivery of Cre recombinase. Upon expression of Cre, oncogenic *Kras*^{G12D} becomes expressed and *Trp53* becomes deleted, generating multifocal tumors. tumor-specific neoantigen expression is achieved by initiating tumors in KP mice with a lentivirus that contains LucOS and Cre using a lentivirus that encodes LucOS (Dupage et. al 2011)(**Figure 1A**). Importantly, we were able to track T-cells that recognize SIIN or SIY by staining them with H-2K^{b} peptide-MHC tetramer and assessing protein expression via flow cytometry.

Because cytotoxicity mediated by CD8 T-cells play a central role in the immune response to a tumor (Raskov et al. 2020), we sought to characterize the CD8 SIIN- and SIY- response in KP lung tumors. We found that the CD8 T cell expansion in response to SIIN was significantly larger compared to the SIY at 5 weeks after tumor initiation, but this difference gradually diminished at subsequent post-tumor initiation timepoints (**Figure 1B**). Notably, the expansion observed in response to SIIN contracted sharply between 5 and 8 weeks, whereas the SIY response remains largely constant (**Figure 1C**).

To further examine the phenotypic differences between SIIN and SIY specific CD8 T-cells, proliferation was assessed by Ki-67 staining. There was no significant difference observed in the proportion of cells proliferating when comparing the SIIN- and

SIY- specific CD8 T-cell response at any of the observed timepoints (**Figure 1D**). Because there are overall less SIY-specific cells that can be detected than SIIN-specific cells, although the proportion of Ki67 expressing cells is the same, the absolute number of proliferating cells is greater in the SIIN response compared to SIY.

We hypothesized that these differences in expansion of SIIN- and SIY- specific cells may be a consequence of CD8 T cell functionality, and sought to measure co-expression of co-inhibitory receptors associated with T-cell dysfunction and exhaustion: PD-1, LAG-3, and TIM-3. The number of cells co-expressing these markers was significantly higher for the SIIN response compared to SIY at 5 and 8 weeks post tumor initiation. Similar to the absolute abundance of SIIN- and SIY- specific CD8 T-cells, the difference in expression of these co-inhibitory receptors between antigen responses was not observed at later time points, 12 and 20 weeks (**Figure 1E**). The degree to which T cell responses contract over time clearly distinguish global response to SIIN and SIY. The absolute number of SIY-specific cells stays relatively constant over time, and contrasts with SIIN-specific cells, which become dramatically less abundant over time. Notably, however, the number and proportion of cells co-expressing inhibitory receptors are indistinguishable by 12 weeks, suggesting that although the kinetics of the SIIN- and SIY- response differ, both eventually become dysfunctional.

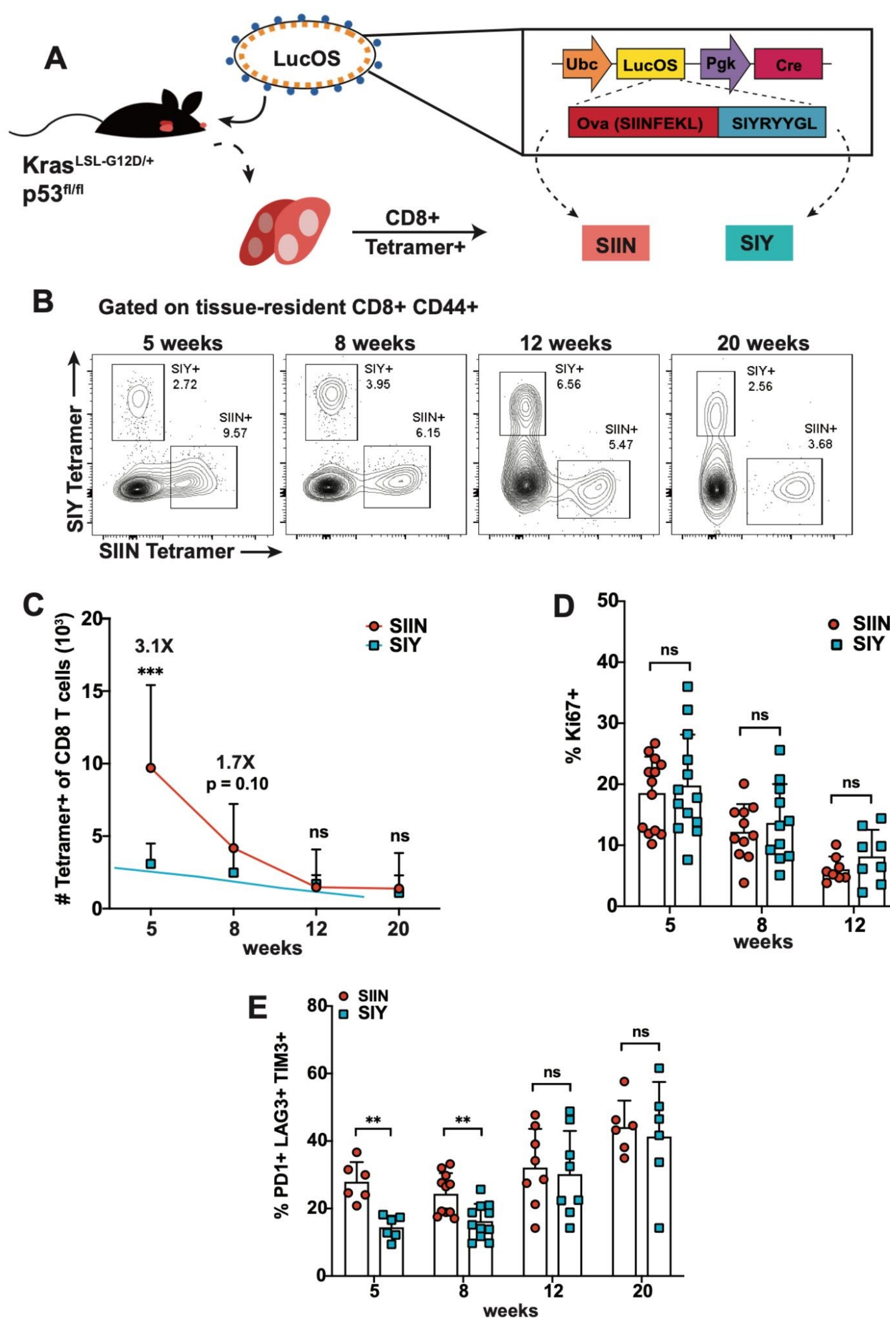


Figure 1. Temporal dynamics of CD8 neoantigen response in Lung Adenocarcinoma.

Data is adapted from Burger et. al 2021. Data are representative of ≥ 3 independent experiments per time point. Each data point represents an individual mouse, n ≥ 5 mice per group in each experiment. Results here and in the following figures are expressed as the mean + SD. Statistics were calculated by two-tailed Student's t test: ns = not significant, *p < 0.05, ** p < 0.01, *** p < 0.001.

(**A)** KP LucOS Model. Lentivirus containing *Cre* and SIIN and SIY fused to luciferase (LucOS) is intratracheally delivered to KP mice to initiate lung adenomas and adenocarcinomas.
(**B**) Representative flow cytometry plots depicting percentages of SIIN- and SIY-specific tissue-resident CD8 T cells (CD8+ CD44+) isolated from KP lung tissue. Specificity for SIIN and SIY is assessed by H-2K^b peptide-MHC tetrameter stain.
(**C**) Summary of absolute numbers of SIIN- and SIY- specific CD8 T cells over 5, 8, 12, and 20 weeks post tumor initiation.
(**D**) Proportion of CD8 T cells specific for SIIN- and SIY- Ki67+ cells as assessed by flow cytometry at 5, 8, and 12 weeks post tumor initiation.
(**E**) Proportion of dysfunctional CD8 T cells specific for SIIN- and SIY- which co-express co-inhibitory receptors PD-1, LAG3, and TIM3.

SIY-Specific CD8 T Cells are Enriched for a TCF1+ Progenitor Phenotype

To further understand the phenotypic differences between the SIIN and SIY response, 5' single-cell RNA-sequencing (scRNA-seq) was performed on SIIN- and SIY-specific CD8 T cells isolated at 5 weeks after tumor initiation, the time point where the responses to the two antigens differs the most. Dimensionality reduction was performed on RNA expression data from these antigen specific cells and visualized in a Uniform Manifold Approximation and Projection (UMAP) embedding (**Figure 2A**). Cells were then assigned to clusters (C0-C10, see Methods) in an unsupervised manner (**Figure 2B**).

To annotate cell clusters in a biologically informative manner, differential gene expression analysis was performed to identify positive and negative markers for each cluster (**Figure 2C**). CD8 T cell exhaustion has been extensively characterized in the context of chronic viral infection, most prominently lymphocytic choriomeningitis virus (LCMV) infection. In this setting, gene signatures associated with progenitors that can give rise to terminally exhausted T cells were derived. These signatures were then used to map the transcriptional landscape of SIIN- and SIY- specific cells by utilizing ProjectTIL atlases that contain these signatures (Raskov et al. 2020; Andreatta et al. 2021) (**Figure 2D**). After consideration of previously published gene expression signatures characteristic of functionally different T cell populations to those of these cell clusters (**Supplemental Figure 2**), each cell cluster was then manually annotated (**Figure 2C**). Taken together, we identified two progenitor clusters that were transcriptionally distinct (C4 and C8) as well as two dysfunctional/exhausted clusters that were highly continuous with one another, but reflect slightly different states of

dysfunction (C2 and C3). A number of cell clusters could be reliably annotated by associated function of the cells within it; for example, clusters with cell states associated with activation (C0 and C6), and exhaustion or memory (C1, C2, and C3). With respect to antigen specificity, SIIN-specific cells appeared to be enriched in several clusters, including exhausted clusters C2 and C3 (**Figure 2E**). In contrast, other clusters were enriched for SIY-specific cells, including progenitor clusters C4 and C8.

Still, gene expression in some clusters could not be clearly associated with previously described CD8 T cell populations. This is particularly evident for some clusters, such as C7, which had expression of genes characteristic of naive, short-lived effector, and effector memory T cells. Notably, C7 in particular was composed almost entirely of SIY-specific cells (**Figure 2E**). Initially, gene expression signatures seemed to suggest these cells may be in a naive-like state. Upon further investigation, it was determined that genes characteristic of naive cells, *Lef1* and *Sell*, were not expressed in these cells. Additionally, activation markers *Cd69* and *Cd44*, which are characteristic of effector cells, were lowly expressed. (**Supplemental Figure 2B-C**). Taken together, it became evident that C7 may have a gene expression signature that is similar, but distinct relative to those associated with previously characterized cell states. Of markers most significantly differentially expressed in C7, very few genes positively marked C7. It's most significant positive marker was *AY036118.1*, a transcript which is associated with the ETS-related transcription factor 1, ERF1. ERF is a potent transcriptional repressor that acts downstream of RAS/MAPK signaling to negatively regulate cell proliferation (Grånäs et al. 2006). Importantly, many genes expressed in C7 are consistent with a repressed proliferative state.

Unsurprisingly, the transcriptional profiles of SIIN- and SIY- specific CD8 T cells were very similar, and all clusters contained at least one cell specific to SIIN or SIY (**Figure 2A**). However, most cell clusters were enriched for cells specific for one antigen relative to the other (**Figure 2E**). Clusters that were not significantly enriched for SIIN or SIY-specific cells, C9 and C10, were small clusters with gene expression signatures reflective of transient responses or programs, interferon signaling and proliferation (**Figure 2C**). The markers associated with clusters that were enriched for cells specific to one antigen over the other were largely related to functionality or functional differentiation of T cells (**Supplemental Figure 2A, Figure 2C**), and suggests that the functionality of response to SIIN and SIY are different. Importantly, this is largely congruent with observations made about the behavior of tetramer stained SIIN- and SIY-specific T cells by flow cytometry. (**Figure 1**). The clusters that were enriched for SIIN-specific cells were generally reflective of activated cells with sustained exposure to antigen. The clusters that had the largest enrichment of SIIN-specific cells were C1 and C2 (**Figure 2E**), which contain cells that appear to progressively adopt gene expression profiles associated with T cell exhaustion. Because these cells of this experiment were harvested at 5 weeks, which precedes contraction of the SIIN-specific response, it is likely that these cells are indeed in the process of becoming dysfunctional. In contrast, the clusters that were enriched for cells specific for SIY, which is subdominant to SIIN, appeared to reflect less differentiated cell states. Amongst these clusters, C4 and C8 contained cells that express genes associated with TCF1+ progenitor cells. This is a particularly important distinction between the SIIN and SIY response, because TCF1+

progenitor cells are thought to give rise to many of the cell states reflected in SIIN-enriched exhausted-like clusters.

The remaining clusters that were enriched for SIY-specific cells, C0 and C7, represent cells with contrasting activation states. C0, which is modestly enriched for SIY-specific cells, appears to be reflective of cell states associated with early responses to productive activation signals and survival (**Figure 2C**). Conversely, the repressed activation cluster, C7, showed the greatest magnitude of enrichment for SIY-specific cells and was depleted for markers of activation. Interestingly, despite contrasting activation states in C0 and C7, both clusters have similar and overlapping positions in UMAP space (**Figure 2B**). It seems likely that SIY-specific cells that do not receive enough activation signals as a consequence of antigen dominance hierarchies enter the cell state reflected in C7 and arrest, rather than expand.

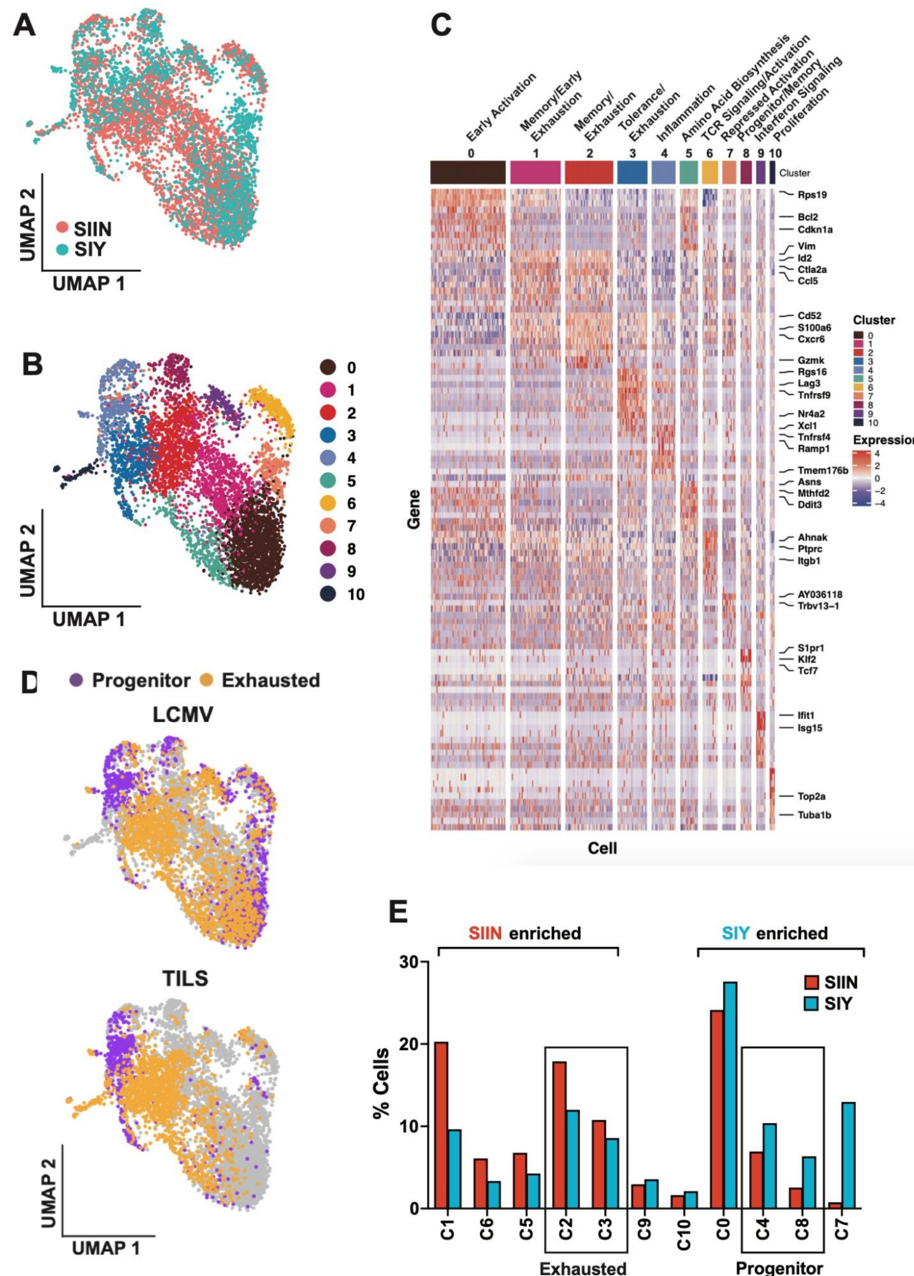


Figure 2. SIY-Specific CD8 T Cells are Enriched for a TCF1+ Progenitor Phenotype
(A) UMAP projection of 5' scRNA-seq gene expression data of SIIN- and SIY- specific CD8 T cells at 5 weeks after tumor initiation. N = 10 mice, 4,023 SIIN-specific and 1,861 SIY-specific cells.
(**B**) Clustering schematic of scRNA-seq data in (A) depicting C0-C10.
(**C**) Heatmap of the top 10 differentially expressed genes per cluster from (B). Genes highlighted on the right of the heatmap were used to annotate clusters that are denoted at the top of the heatmap.

(**D**) ProjectTIL classification using lymphocytic choriomeningitis virus (LCMV) infection and tumor infiltrating lymphocyte (TIL) signatures as reference atlases. SIY-specific cells are enriched for progenitor phenotypes (TIL $p = 0.002$, LCMV $p = 2.51E-15$) and SIIN-specific cells are enriched for an exhausted phenotype (TIL $p = 4.01E-17$; LCMV $p = 1.52E-10$).
(**E**). Quantification of SIIN- and SIY-specific cell assignments to clusters in (B). Brackets indicate clusters significantly ($p < .05$) enriched for either antigen specific cell. SIIN is enriched in C2 ($p = 2.9E-9$) and C3 ($p = 5E-3$) while SIY is enriched in C4 ($p = 4.9E-6$) and C8 ($p = 3.7E-12$).

Analysis of TCR Clonotypes from SIIN versus SIY scRNA-seq Data Delineates Relationships Between CD8 T Cell States

The connectivity of clusters C4 and C8 to C2 and C3 on the UMAP (**Figure 2B**) suggests transcriptional similarity between these cell populations, consistent with the previously proposed lineage relationship between progenitor and exhausted cells (**Figure 2B**)(Siddiqui et al., 2019). We further examined this relationship by analyzing the distribution of individual TCR clonotypes (with ≥ 5 cells) containing at least one cell assigned to progenitor clusters C4 or C8 (**Supplemental Figure 3A**). This was also performed across all clonotypes (**Figure 3A**).

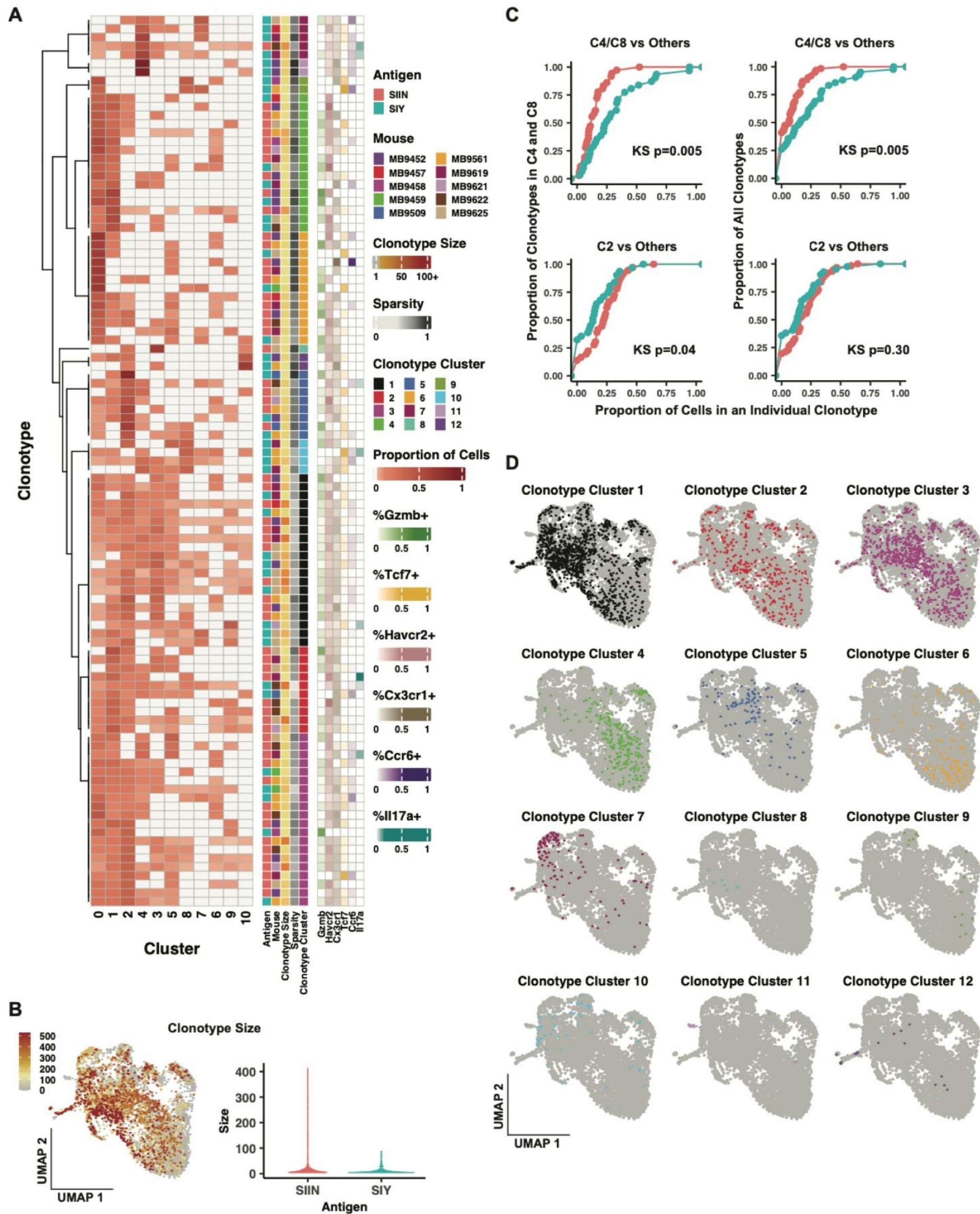
Because the observed clonotypes were relatively small, we reasoned that repeated patterns of functional differentiation may stratify groups of clonotypes in a manner that reflected their functional behavior. Unsupervised hierarchical clustering on the distribution of the clonotypes across clusters C2, C3, C4, and C8 largely segregated SIY clonotypes into the top half of the heatmap and SIIN clonotypes into the bottom half (**Supplemental Figure 3A**; see “antigen” side-bar). Across all clonotypes (**Figure 3A**) this trend is also observed, however, many more clonotypes which did not have cells assigned to progenitor clusters C4 or C8 appeared to have similar dynamics when comparing SIIN- and SIY-specific clonotypes. Together, this indicates that SIIN and SIY clonotypes that contain progenitor cells have unique, antigen-specific distribution patterns across clusters of cell states. Importantly, the SIIN- and SIY-specific clonotypes appear to be distributed across clusters equally, as assessed by Gini index, but do have differences in the clusters they are found (**Figure 3A**). Further, clonotypes were scored

for expression of key phenotypic genes, *Gzmb, Havcr2, Cx3cr1, Tcf7, Ccr6* and *Il17a*, as a proxy for their functional phenotypes (**Figure 3A**).

SIIN clonotypes that contained progenitor cells were distributed across most clusters, but were enriched for cells assigned to exhausted cluster C2 (**Supplemental Figures 3A and Figure 3C**); this further supports a lineage relationship between progenitor and exhausted cell states and suggests that SIIN clonotypes are well-progressed on the path to exhaustion. In contrast, both progenitor cell-containing clonotypes and total SIY clonotypes were biased in cell distribution to clusters C4 and C8 (**Supplemental Figures 3A, Figure 3A and Figure 3C**). Additionally, while there were a similar number of SIIN and SIY clonotypes (≥ 2 cells: SIIN 153, SIY 149), SIY clonotypes were smaller in size than SIIN clonotypes (**Figures 3B**). These observations suggest that SIY clonotypes are repressed in clonal expansion and differentiation to an exhausted cell state compared to SIIN clonotypes. Notably, however, sampling bias due to the smaller input of SIY cells may also contribute to reduced SIY clonotype size (see methods).

Flow cytometric analysis, published in **Burger et al, 2021,** confirmed that SIY-specific cells were enriched for a progenitor cell phenotype at 5 weeks, referred to hereafter as “TCF1+ progenitor” and defined as CD8α+ CD44+ TCF1+ TIM3- cells. Consistent with previous reports (Miller et al., 2019; Siddiqui et al., 2019), the majority of these cells expressed the inhibitory receptor PD1 (**Figure S3B**). Few expressed GZMB, indicating that they are generally not cytotoxic, but were more proliferative than TCF1- cells and were able to produce the effector cytokines IFNɣ and TNFα (**Burger et al, 2021**, **Supplemental Figure 3C**). Additionally, SIY-specific cells were enriched for

TCF1+ progenitor cells compared to SIIN-specific cells at 5 and 8 weeks (**Supplemental Figure 3D**). Notably, the absolute number of TCF1+ progenitor cells was similar between the responses across all time points. In summary, these data indicate that SIY-specific cells are enriched for a less differentiated, TCF1+ progenitor cell state early in the response to KP LucOS lung tumors, while SIIN-specific cells differentiate more rapidly to an exhausted cell state.



A
Clonotype
Cluster
Antigen
SIIN
SIY
Mouse
MB9452
MB9457
MB9458
MB9459
MB9509
MB9561
MB9619
MB9621
MB9622
MB9625
Clonotype Size
1 50 100+
Sparsity
0 1
Clonotype Cluster
Proportion of Cells
%Gzmb+
%Tcf7+
%Havcr2+
%Cx3cr1+
%Ccr6+
%Il17a+
Antigen
Mouse
Clonotype Size
Sparsity
Clonotype Cluster
Gzmb
Havcr2
Cx3cr1
Tcf7
Ccr6
Il17a
B
Clonotype Size
UMAP 1
UMAP 2
Size
SIIN
SIY
Antigen
C
C4/C8 vs Others
C2 vs Others
KS p=0.005
KS p=0.04
KS p=0.30
Proportion of Clonotypes in C4 and C8
Proportion of All Clonotypes
Proportion of Cells in an Individual Clonotype
D
Clonotype Cluster 1
Clonotype Cluster 2
Clonotype Cluster 3
Clonotype Cluster 4
Clonotype Cluster 5
Clonotype Cluster 6
Clonotype Cluster 7
Clonotype Cluster 8
Clonotype Cluster 9
Clonotype Cluster 10
Clonotype Cluster 11
Clonotype Cluster 12
UMAP 1
UMAP 2

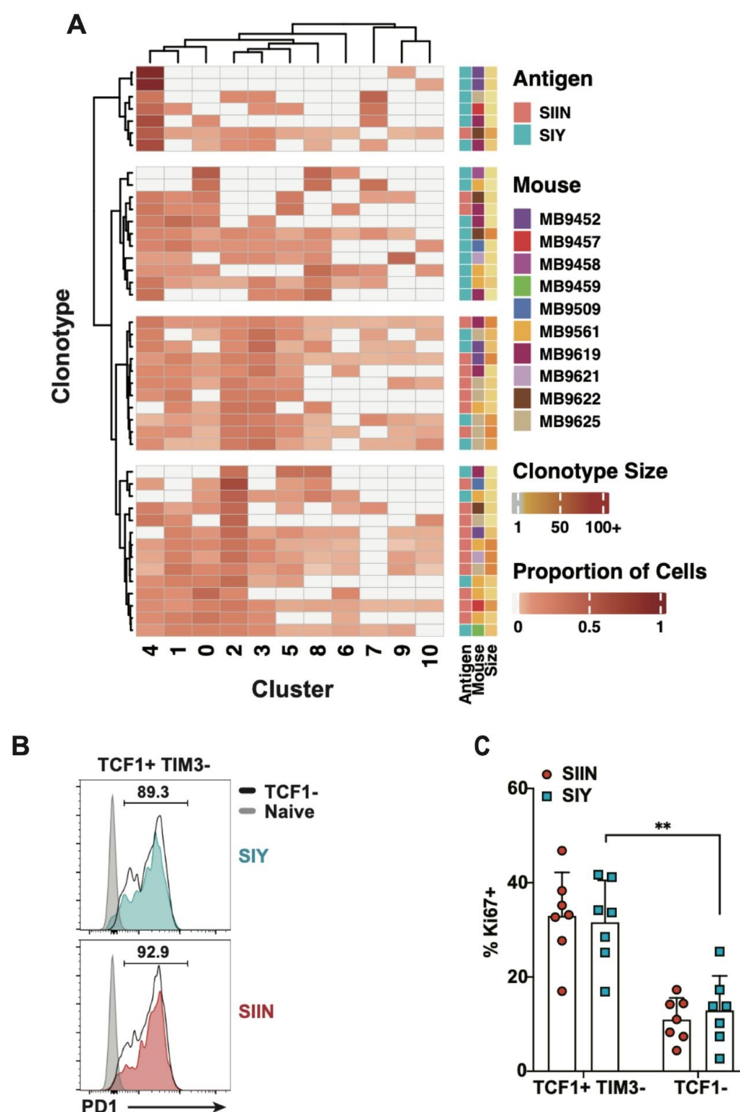
Figure 3. Analysis of TCR Clonotypes from SIIN versus SIY scRNA-seq Data Delineates Relationships Between CD8 T Cell States

(**A**) Heatmap representation of the distribution of TCR clonotypes (≥ 5 cells) across all clusters. Proportion of cells scale: 1 = contains 100% of cells assigned to the clonotype, 0 = contains 0% of cells assigned to the clonotype. Hierarchical clustering was performed on clonotypes (rows) and the clusters (columns). “Clonotype clusters”, antigen and mouse assignments, and the number of cells in each clonotype (size) are indicated on side-bars. A measure of the spread in distribution of each clonotype across clusters, or “sparsity” (Gini index)(Hurley, 2009), is also included. Additional side-bars show the proportion of cells in each cluster expressing key phenotypic genes: *Gzmb, Havcr2, Cx3cr1, Tcf7, Ccr6* and *Il17a*.

(**B**) UMAP depicting clonotype size for individual cells belonging to SIIN and SIY clonotypes. Scale bar indicates the number of cells in a clonotype. Cells that belong to very small clonotypes or do not have a clonotype assignment are shown in gray. The violin plot shows a comparison of the size of SIIN versus SIY clonotypes containing ≥ 2 cells.

(**C**) ECDF plots depicting the distribution of SIIN and SIY clonotypes to the indicated clusters. Enrichment is expressed as p-value by KS test.

(**D**) UMAP plots highlighting cells assigned to each of the clonotype clusters depicted on the clonotype cluster side-bar in (A). Clonotype cluster 7 is enriched for a Tc17 phenotype.



Supplemental Figure 3. SIY-specific CD8 T cells are enriched for TCF1+ cells and are less differentiated than SIIN-specific CD8 T Cells.
(A) Heatmap depicting proportions of TCR clonotypes (rows) assigned to each cluster (columns), for clonotypes ≥ 5 cells and ≥ 1 cell assigned to Tcf1+ progenitor clusters C4 or C8. Annotations for antigen specificity, mouse/source, and size (cell numbers) of each clonotype are indicated. Proportion of 1 = 100% of clonotype.
(**B-C**) Flow cytometric analysis of PD1 (B) and Ki67 (C) expression by TCF1+ TIM3- (colored) versus TCF1- (black line) SIIN- and SIY-specific CD8 T cells at 5 weeks. The naïve population (grey) represents CD44- tetramer- CD8+ T cells. Results are representative of 3 independent experiments.
(**D)** Percentage and absolute number of progenitor TCF1+ TIM3- cells of SIIN- and SIY-specific CD8 T cells over time. Representative of ≥ 3 independent experiments.

The Subdominant CD8 T Cell Response is Enriched for a CCR6+ TCF1+ Progenitor Cell Subset with a Tc17 Differentiation Trajectory

Competition with the SIIN response might continue to repress SIY-specific TCF1+ progenitor cell differentiation in the context of ICB and account for why SIY-specific cells do not respond better than SIIN-specific cells to treatment. Alternatively, it may be that SIY-specific TCF1+ progenitor cells are intrinsically less functional than their SIIN counterparts. In the scRNA-seq data, two clusters were marked by “progenitor” gene signatures (C4 and C8) and were proximal to two clusters marked by “exhausted” gene signatures (C2 and C3). To explore this apparent heterogeneity in the progenitor and exhausted cells, we separately analyzed the cells contained within these four clusters (**Figure 4A**). Interestingly, we found that *Tcf7*-expressing cells within C4 and C8 occupied distal regions of the UMAP, suggesting C4 and C8 contain distinct progenitor cell populations. C4 and C8 were both enriched for a number of genes previously associated with TCF1+ progenitor cells (e.g. *Tcf7, Xcl1, Slamf6, Ccr7)*, but were also characterized by unique gene signatures (**Figure 4B**). C8 was marked by genes associated with memory T cells, including the trafficking regulators *Klf2* and *S1pr1,* as well as the survival receptor *Il7r* (**Figure 4B and S4A**)(Best et al., 2013). Conversely, C4 was characterized by expression of markers of T cell dysfunction, tolerance and anergy, and most strikingly, showed strong enrichment for a signature of Tc17 cells (**Figure S4B, 4B, S4A and S4B**) (Linehan et al., 2018; Parish et al., 2009; Safford et al., 2005).

Tc17 cells, marked by expression of the chemokine receptor CCR6, are an IL17A-producing CD8 T cell subset commonly associated with autoimmune

inflammation (Srenathan et al., 2016). *Ccr6* was highly expressed in C4 and its expression overlapped with *Tcf7* expression; however, cells exhibiting other hallmarks of differentiated Tc17 cells, including expression of the transcription factor *Rorc* (i.e. RORɣT) and the cytokine *Il17a,* were predominantly found adjacent to *Tcf7*-expressing cells within C4 (**Figure 4B**). To determine if these cells shared an ancestral relationship to *Tcf7*-expressing cells, we used Monocle3 to infer lineage trajectories (Cao et al., 2019; Trapnell et al., 2014). This analysis predicted a trajectory connecting *Tcf7*-expressing cells in C4 with this putative Tc17 population (**Figure 4C**). *Il17a* expression was found outside clonotype cluster 7, but in a smaller number of clonotypes compared to expression of *Ccr6*, *Tcf7* or effector/exhaustion genes *Gzmb, Cx3cr1* and *Havcr2 (***Figure S4A**)*.* This suggests that a subset of TCRs might preferentially drive differentiation down the Tc17 pathway; however, most of these clonotypes expressed *Il17a* in a small proportion of cells and were distributed across multiple clusters (**Figure 3A**), indicating that clonotypes giving rise to Tc17 cells also give rise to other cell states. Flow cytometry analyses at 5 weeks later confirmed the presence of CCR6+ SIIN- and SIY-specific cells and revealed enrichment for this population in the SIY response (**Burger et al 2021)**. Altogether, these results uncover previously undescribed heterogeneity amongst TCF1+ progenitor cells. High expression of markers of dysfunction/tolerance, low GZMB expression and differentiation to a Tc17 phenotype suggest CCR6+ TCF1+ cells represent an unconventional TCF1+ population with reduced functionality.

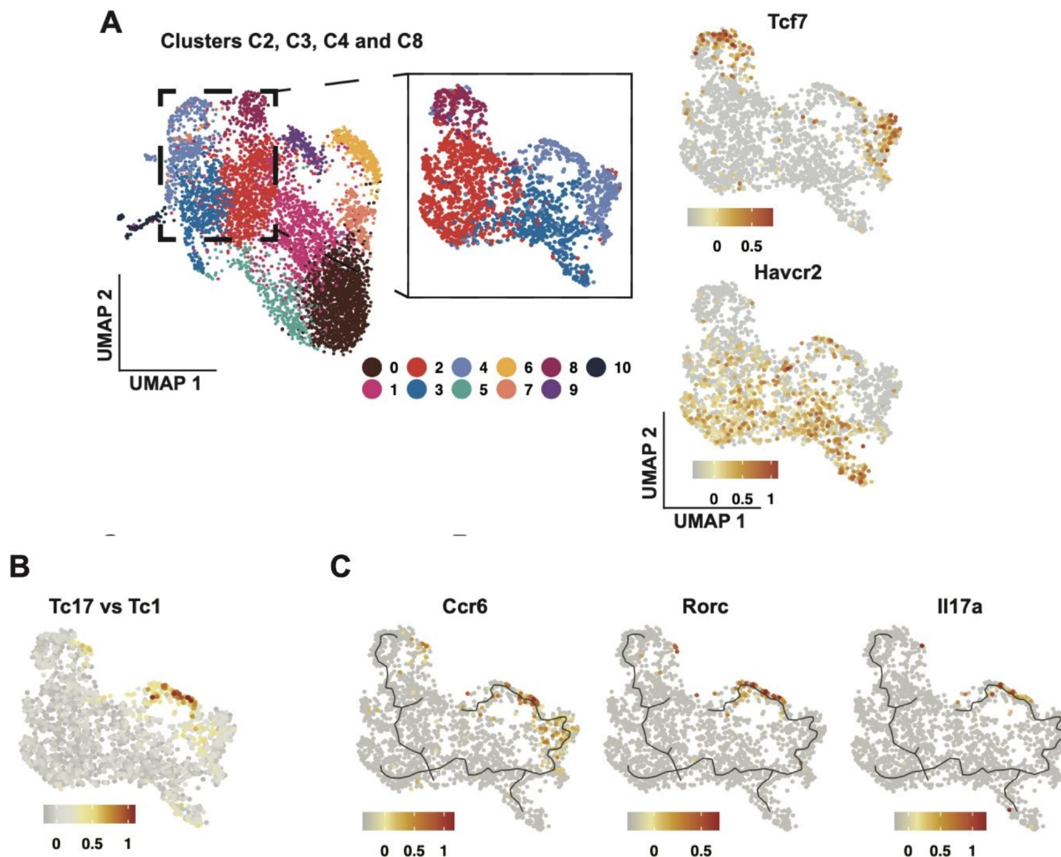


Figure 4. The Subdominant CD8 T Cell Response is Enriched for a CCR6+ TCF1+ Progenitor Cell Subset with a Tc17 Differentiation Trajectory

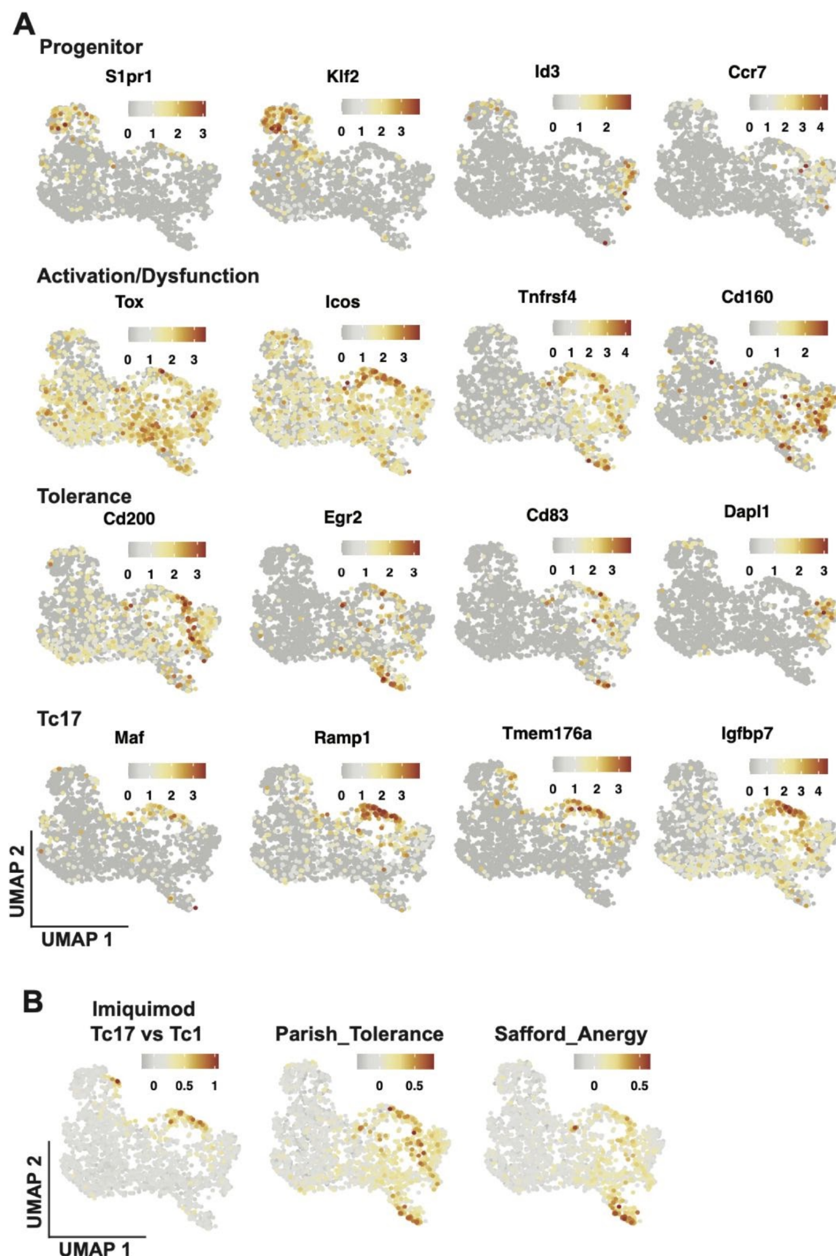
(**A**) Characterization of cells assigned to clusters C2, C3, C4 and C8 from the mouse scRNA-seq data (Figure 2B) with a separate UMAP embedding showing expression of *Tcf7* and *Havcr2*.

(**B**) Expression (mean log(expression +1)) of genes associated with the indicated classifications across clusters C2, C3, C4 and C8.

(**C**) Scoring of individual cells from (A) for enrichment of a gene signature differentially upregulated in Tc17 versus Tc1 CD8 T cells in the skin of mice infected with *S. epidermidis* (Linehan et al., 2018).

(**D**) Monocle3 lineage trajectories that connect cells expressing *Ccr6/Tcf7* and *Rorc/Il17a* overlaid with UMAP visualizations of Tc17-related gene expression.

(**E**) UMAP plots from (A) highlighting a group of five TCR clonotypes (clonotype cluster 7, see also Figure S4A and S4D) that fall predominantly within cluster C4 and span cells expressing *Ccr6/Tcf7* and *Rorc/IL17a* along the Monocle3-predicted trajectory in (D).



Supplemental Figure 4. Tc17-like TCF1+ Progenitor CD8 T Cells in KP LucOS Tumors

(A) Expression of select genes from the heatmap in Figure 5B projected onto the UMAP from Figure 4A.

(B) Scoring of cells from the mouse scRNA-seq analysis in Figure 4A for expression of gene signatures for Tc17 vs Tc1 from skin cells of Imiquimod treated mice (Linehan et al., 2018), T cell deletional tolerance (Parish et al., 2009) and T cell anergy (Safford et al., 2005).

DISCUSSION

The protective effects of T cells responding to the tumor-specific neoantigens, SIIN and SIY are most evident at early time points after tumor initiation (DuPage et al. 2011). In this study, we show the response to SIIN dominates the response to SIY; SIIN-specific T cells are more abundant and more cytotoxic than those specific for SIY (**Figure 1, Burger et. al 2021**). As such, the tumor protective effects are likely to be mediated primarily by the SIIN response. At timepoints where T cells appear to lose their ability to restrict tumor growth, the SIIN response also appears to contract (DuPage et al. 2011)(**Figure 1**). This further supports the notion that the SIIN response not only dominates the response to SIY in terms of antigen specific CD8 T Cell abundance, but also dominates the kinetics and productivity of the overall T cell response to LucOS.

At the 5 week timepoint, many SIIN-specific cells have become fully activated, cytotoxic, and differentiated (**Figure 1, Burger et. al 2021**). At the same time, many of these cells acquire gene expression profiles indicative of progressive dysfunction preceding terminal exhaustion. This is not surprising, because the T cell response to SIIN contracts shortly thereafter (**Figure 1**). In parallel, the cells responding to SIY appear to have cell states with less functional or suspended differentiation. However, the kinetics of the SIIN and SIY response do not just differ temporally (**Figure 1**); the SIY response is never observed to expand to the same degree as the SIIN response when co-expressed with SIIN. Later, **Burger et al 2021** went on to show that when antigens are expressed alone, the phenotypes of SIY-specific cells were lost.

Additionally, when SIIN was substituted with another neoantigen epitope that is thought to have a lower affinity for MHC than SIY, the observed antigen dominance hierarchy was reversed, instead favoring SIY as the dominant antigen (**Burger et al 2021**). Taken together, these observations suggest that the SIY antigen can elicit a meaningful T cell response, and that the difference between the SIIN and SIY responses when both antigens are co-expressed are specifically a consequence of co-expression and antigen dominance.

While the SIY response does not amount to the same magnitude of response to SIIN in the LucOS setting, this does not inherently mean that the subdominant SIY response is incapable of doing so. Intriguingly, although the SIY response never undergoes the phases of expansion and contraction that are characteristic of CD8 T cell responses, SIY-specific cells do appear to become dysfunctional. If the primary difference between the SIIN and SIY response are the kinetics of functional differentiation, it is entirely possible that suppression of the SIY response is further compounded by an increasingly immunosuppressive tumor microenvironment. By the time the SIIN response begins to contract, tumors are larger and have already adapted to selective pressures imposed by SIIN-specific cytotoxic T cells. As such, when the suppressive influence of the SIIN response becomes alleviated, the SIY response may be further blocked by the tumor itself.

While absolute numbers of Ki67 expressing cells differ between the SIIN and SIY response at 5 weeks post tumor initiation (**Figure 1D**), multiple factors may contribute to the difference in observed expansion of the response to SIIN compared to SIY. This may be a consequence of the fact that observed CD8 T cell expansion may

reflect both proliferation and expansion of tissue resident T cells as well as peripheral T cells which then infiltrate into the tumor-bearing lung.

One of the challenges associated with TCR clonotype analysis is that the abundance of a clonotype can be influenced by technical factors, such as sequencing dropout or underlying biological abundance of a clonotype. In the case of the SIIN- and SIY- specific CD8 T Cell response, we observed more clonotypes overall for the SIY response (**Figure 3B**), but these clonotypes were less expanded than SIIN-specific clonotypes. In less expanded clonotypes, because there are less cells, it is more difficult to determine the differentiation potential or distribution of cell states associated with the clonotype when there are inherently a small number of cells to begin with. To work around this issue, clonotypes were clustered according to their observed distribution patterns across cell clusters (**Figure 3A**) to aggregate clonotypes with similar distribution patterns. However, the observed distribution patterns are likely to be less robust for smaller clonotypes; as such, we inherently have less confidence in clonotype cluster assignments of small clonotypes. In spite of this technical constraint, when cells are pseudocolored by their associated clonotype size (**Figure 3D**), we observe that expanded clonotypes occupy different transcriptional spaces than clonotypes of small sizes. If we assume that sequencing dropout occurs uniformly across cells, then this at the minimum suggests that the degree of observed expansion for a given clonotype does influence transcriptional state.

It is important to note that interpretation of how descendants of a particular clone are distributed across cell states observed in this dataset is dependent on the assumption that cells within a clonotype have some sort of shared ancestry. Functional

differentiation of lymphocytes frequently follow paradigms of multilineage priming (Laslo et al. 2006) in which daughter cells of a given progenitor differentiate into multiple cell types simultaneously. Stated differently, the distribution of a clonotype across cell states can reflect one or many differentiation trajectories. This is likely to be most confounding in very large, expanded clonotypes which generally have at least one cell assigned to nearly all cell clusters described in our data.

In addition, the observed clonotype sizes of the SIIN- and SIY- response are (**Figure 3B**) consistent with the number of tetramer stained SIIN- and SIY- cells quantified in independent experiments by flow cytometry (**Figure 1C**). In both the SIIN- and SIY- response, the number of clonotypes that become very expanded relative to the number of clonotypes detected is very small. For both responses, the observed distribution of clonotype sizes are consistent with previously published work which suggests that antigen responses, irrespective of antigen dominance hierarchies, are dominated by certain clonotypes. In other words, productive antigen responses generally have low TCR diversity associated with them. This has been reported in a number of different biological contexts, including aging (Britanova et al. 2014), and ankylosing spondylitis (Hanson et al. 2020).

Intriguingly, the most significant marker of C7, *AY036118* is a poorly annotated gene in the mm10 reference genome primarily because it has nucleotide sequences that complicate read mapping. The gene contains a 705 base pair sequence with 95% homology to 18S rRNA in the 3’UTR (Kong et al. 2008) and has conflicting computational annotations. As a consequence, it is annotated as a pseudogene in Ensembl and ncRNA by NCBI. In spite of these annotations, clear evidence exists to

support protein expression of Erf1 (Lindemann et al. 2001; Grånäs et al. 2006). Erf1 has been functionally characterized as a potent repressor of cell proliferation and MAPK signaling (Sgouras et al. 1995; Papadaki et al. 2007). Many of the genes differentially downregulated in C7 were related to cell proliferation and activation; as such, we annotated cells in C7 as having cell states with Repressed Activation (**Figure 2C**). In agreement with this observation, C7 also had very few cells assigned to expanded clonotypes, suggesting that cells that adopt this repressed activation state may be unable to expand (**Figure 3B**).

In the context of an antigen dominance hierarchy, as reported here, this generalization persists. However, the largest SIY- specific clonotypes are markedly smaller than the most expanded SIIN- specific clonotypes (**Figure 3B**). In addition, the fact that there are more productive clonotypes identified in the SIY- specific response relative to the SIIN-specific response, in spite of greater SIIN- input cell numbers, further suggests TCR diversity may be inversely correlated with degree of observed clonotype expansion. However, it is less clear whether or not clonotype expansion simply reduces the likelihood that the true underlying TCR sequence diversity is captured, or whether expansion of one clonotype is able to influence expansion of another. In the context of antigen dominance, further questions can be raised about the influence of expanded clonotypes specific for dominant antigens on expansion of clonotypes specific for subdominant antigens.

In this study, the functionality of T cells that respond to dominant and subdominant antigens was characterized. These associations could be further extended to observations made about TCR diversity and expansion; if associations between TCR

diversity and antigen response are reliable and robust, measures of clonotype expansion and TCR diversity could be used to predict the productivity of response to associated neoantigens, which may be of particular utility in settings where neoantigen expression is not experimentally defined. These associations may be especially useful in clinical studies of neoantigen responses, where it is far more difficult to characterize and assess neoantigen responses.

However, the utility of these associations is starkly limited by the inability to associate a given TCR sequence with its associated neoantigen. This may be overcome by computational algorithms that can predict TCRs that respond to the same neoantigen (Glanville et al. 2017) and through empirical characterization of the relationship between TCR sequence identity and cognate neoantigens identity. Complementary efforts to identify features of neoantigens that can elicit productive T cell responses may lead to further insight when characterized with respect to TCR diversity and expansion.

Taken together, it is abundantly clear that the global response to a single neoantigen is dictated by heterogeneous responses of TCR clonotypes specific for that neoantigen. In the case of the most expanded SIIN-specific CD8 cells, observations made of the global response to a given antigen may be dominated by the behavior of one or very few clonotypes. This may explain variation that is often observed in flow cytometry based analyses. Notably, even SIIN-specific T cells occupy transcriptional states enriched for subdominant SIY-specific cells, suggesting that neoantigen identity does not entirely predict, but does influence, how T cells will respond to it. It is entirely possible that similarities between the heterogeneous SIIN- and SIY- responses are a

result of internal competition between clonotypes that recognize the same neoantigen, and are further compounded by competition between different neoantigens.

In work that is not included in this thesis, but is published in **Burger et al 2021**, further experiments were performed to functionally characterize the heterogeneous progenitor populations observed in this study. It was found that expression of *Ccr6* can distinguish between CD8 progenitor populations, and these populations have meaningful differences in their ability to drive productive immune responses. Additionally, it was found that this population of CCR6+ cells is lost following therapeutic vaccination against SIY, further providing evidence that SIY specific T cells are being primed suboptimally when in competition with SIIN. In future studies, characterization of both TCRs and neoantigens that are capable of eliciting productive immune responses may lead to more meaningful comparison of dysfunctional T cell responses. Stated simply, there may be multiple manifestations of T cell responses that are unproductive or dysfunctional that cannot be stratified by associated neoantigen identity. Overall, however, this work demonstrates the importance and influence that antigen dominance hierarchies have on functional differentiation of CD8 T cells.

MATERIALS AND METHODS

Lentiviral Tumor Induction

Tumors were induced in KP mice that were at least 8 weeks old through intratracheal delivery of lentivirus containing Cre recombinase and model neoantigens (2.5 x 10^4 PFU) as previously described (DuPage et al. 2011; DuPage, Dooley, and Jacks 2009). Mice were randomized for analysis and ICB and/or vaccination therapy.

Tissue Collection and Flow Cytometry

Analyses were performed on tissue-resident immune cells, which were distinguished from circulating immune cells through retroorbital injection of an anti-CD45 antibody 2-3 minutes prior to euthanasia (PE-CF594 or AlexaFluor780; 30-F11; BD Bioscience) (Anderson et al. 2014). Lung tissue was harvested and dissociated with a combination of manual cutting with spring scissors and a 30 minute treatment with collagenase IV treatment (125 U/mL) (Worthington Biochemical) combined with DNAse I treatment (40 U/mL)(Sigma-Aldrich) at 37°C. Lung tissue was further dissociated using gentleMACS dissociator m_lung_2.0.1 protocol in gentleMACS C tubes (Miltenyi Biotec) and passage through a 70 μm strainer. For spleen and lymph node tissues, dissociation was performed with a 70 μm cell strainer into RPMI 1640 media with 1% heat-inactivated fetal bovine serum. For all tissues, cells were pelleted by centrifugation and resuspended in 1X RBC Lysis Buffer (eBioscience) on ice for 10 minutes. Cells were then resuspended in Phosphate Buffered Saline (PBS) and transferred to a 96-well U-bottom plate.

Dead cells were excluded by staining viable cells 20 minutes on ice with Zombie or Tonobo Ghost Dye (Invitrogen and Tonbo Biosciences, respectively). Subsequently, cells were stained for surface antibodies purchased from ThermoFisher Scientific, BD BioSciences, or Biolegend in PBS with 1% Heat-inactivated FBS for 15-30 minutes on ice. Antibodies used: CD8ɑ (53-6.7), CCR6 (29-2L17), CX3CR1 (SA011F11), CXCR3 (CXCR3-173), CD44 (IM7), IL7R (A7R34), LAG3 (C9B7W), PD1 (RMP1-30), SLAMF6 (13G3), TIGIT (1G9), TIM3 (RMT3-23), OX40 (OX-86), ICOS (7E.17G9), CD200 (OX-90), CD83 (Michel-19). During this incubation, cells were concurrently stained with

H-2K^{b} peptide-MHC tetramers specific to SIINFEKL, SIYRYYGL, SIINYEKL, mALG8 or mLAMA4 (monomer, NIH Tetramer Core Facility; PE and APC streptavidin, Invitrogen). Cells were fixed to permit intracellular staining for 1 hour at room temp (eBioscience Fixation/Permeabilization Kit, ThermoFisher Scientific) and then stained overnight at 4°C with antibodies purchased from Cell Signaling Technology, ThermoFisher Scientific, BD Biosciences, Biolegend or Miltenyi Biotec. Antibodies used: TCF1/TCF7 (C63D9), RORɣT (B2D), TBET (eBio4B10), Granzyme B (GB11), Ki67 (B56), TOX (REA473), EGR2 (erongr2).

All samples were analyzed on a BD Biosciences LSR Fortessa or LSR II Flow Cytometry Analyzer.

Cytokine production was evaluated following depletion of tumor and myeloid cells in lung tissue, identified by incubation with 2 μg of purified antibodies specific for Ly-6G, EpCAM, and F4/80 (Biolegend) at 4°C for 20 minutes. Cells were subsequently incubated with 125 μl of sheep anti-rat Dynabeads (Invitrogen) at 4°C for 30 minutes while rotating and stained for CD45.2 (Brilliant Violet 510, Biolegend). Excluded cells were pelleted with a Dynabeads magnet (Invitrogen), leaving T cells in the supernatant that were then transferred to a clean tube, washed with PBS containing 1% heat inactivated PBS, pelleted by centrifugation, and resuspended in T Cell media (RPMI 1640 with 10% FBS, 10 mM HEPES, 1 mM sodium pyruvate, 1X MEM Non-essential amino acids, 2 mM L-glutamine, 0.275 mM beta-mercaptoethanol, and 50 U/mL penicillin-streptomycin). Cells were then transferred to a 96-well U-bottom plate, pelleted and resuspended in T cell media plus 1X Monensin (Biolegend), 1X Golgi Plug (BD Bioscience) and SIINFEKL or SIYRYYGL peptide (167 nM; New England Peptide)

to assay IFNɣ and TNFα production. Cells were stained for IL17A by resuspending in T cell media with PMA (2.5 ng/ml; EMD Millipore), Ionomycin (1 µM; Sigma-Aldrich), 1X Monensin, and 1X Golgi Plug and incubation at 37°C for 4-5 hours. Unstimulated controls were generated by reserving 10% of each aliquot. The remaining 90% of cells in each aliquot were then stained for viability and surface markers as described above through incubation with antibodies overnight at 4°C. Antibodies used: IFN-ɣ (XMG1.2), TNFα (MP6-XT22) and IL17A (17B7) from ThermoFisher Scientific, BD Biosciences or Biolegend.

scRNA-seq Data Processing

Base calls, mapping/alignment, and counts of scRNA seq 5' RNA expression data were performed using Cell Ranger, version 3.1.0 (Zheng et al. 2017). Cell Ranger 3.1.0 was also used to map VDJ and Cell Hashing libraries. RNA expression data was aligned to the GRCm38/mm10 reference mouse transcriptome (version 3.0.0) and VDJ sequencing data was aligned to the prebuilt mouse (GRCm38/mm10) VDJ reference supplied by 10X Genomics (version 3.1.0) (Zheng et al. 2017).

Approximately 6,472 cells and 15,939 genes were detected for SIIN-specific CD8 T-Cell libraries at a sequencing depth of 80.4%. 3,646 cells and 14,834 genes were detected for RNA SIY- specific CD8 T-cells at a sequencing depth of 76.9%. VDJ libraries (containing TCR sequences) captured an estimated 4,713 (SIIN) and 2,705 cells (SIY) with 3,857 (SIIN) and 2,196 (SIY) of those cells containing productive V-J Spanning Pair. Cell Hashing libraries captured approximately 6,073 (SIIN) and 3,319 (SIY) cells at a sequencing saturation of 10.3% (SIIN) and 13.2% (SIY) for SIIN- and SIY- specific CD8 T Cell libraries, respectively.

Seurat (version 4.0.0) was used to transform normalized counts by centered-log ratio (CLR) and demultiplex cell hashing data (Butler et al. 2018; Ruf-Zamojski et al. 2018). A positive quantile threshold of 0.98 to infer which mouse each cell was harvested from.

RNA expression counts were normalized and natural log-scaled in Seurat (version 4.0.0). Seurat was also used to select variable features, and perform differential gene expression analysis. Dying cells, probable doublets, and low quality data were filtered out by imposing requirements for individual cells to express at least 100 but less than 4000 genes and have at least 20,000 reads, with a maximum of 5% of reads aligning to the mitochondrial genome. Additionally, cells called as doublets by cell hashing were also removed.

Expression for all genes in SIIN- and SIY- libraries were first centered by subtracting average expression of each gene and subsequently scaled by dividing gene expression levels by their standard deviations. SIIN- and SIY- libraries were then merged in Seurat (“merge.data = TRUE”). For each cell passing quality control thresholds, metadata assignments for V(D)J clonotypes and cell hashing (mouse of origin) were made using Python with Pandas and Numpy (McKinney 2017; Harris et al. 2020)

Cell Clustering, and Differential Expression Analysis

Dimensionality reductions were performed using the 2,000 most variable genes, selected in Seurat using the vst method (Butler et al. 2018; Ruf-Zamojski et al. 2018). An estimation of principal components (PCs) used for further dimensionality reduction was estimated by performing Principal component analysis (PCA) for the first 50 PCs

and JackStraw analysis/elbow method. This estimation was further refined to a final 30 PCs through manual evaluation of features. These 30 PCs were used to construct a shared nearest neighbor graph (SNN, k = 20) and perform Louvain clustering using default parameters in Seurat (default parameters, resolution; (Meo et al. 2011). Cells were embedded into 2-dimensional space by Uniform Manifold Approximation and Projection (UMAP) algorithm (McInnes et al. 2018; Butler et al. 2018) with default parameters in Seurat. Genes differentially expressed between cell clusters were identified by Wilcoxon Ranked Sum test (FindAllMarkers, min.pct = 0.25).

Heatmaps with differentially expressed genes for each cluster were produced using ComplexHeatmap (cluster_columns = FALSE, cluster_rows = FALSE). Statistical enrichment was tested by a hypergeometric test in R (phyper, alpha = 0.05). Gene expression visualizations in UMAP space were generated in Seurat using the FeaturePlot function (order = TRUE) or in Monocle3 using the plot_cells function. Data from cells assigned to C2, C3, C4 and C8 were separately analyzed in Monocle3, version 0.2.3.0 (Junyue Cao et al. 2019; Qiu et al. 2017; Trapnell et al. 2014). Raw data was normalized using default parameters. PCA was performed and the first 20 PCs were utilized to generate a UMAP embedding (default parameters, except: umap.n_neighbors = 30L, umap.fast_sgd = FALSE, preprocess_method = “PCA”; McInnes et al., 2018). Cells were grouped using leiden clustering (Traag, Waltman, and van Eck 2019) for trajectory analyses (resolution = 0.001) with otherwise default parameters. To predict differentiation trajectories, a principal graph (visualized in UMAP space) was learned in Monocle3 (learn_graph, use_partition = FALSE, rann.k = 20). Further visualizations were created by exporting UMAP coordinates for these cells from

Monocle3 to Seurat for gene expression and signature analysis. Differentially expressed genes were calculated in Seurat, as described above.

Genes associated with CD8 T cell functionality were identified through manual consideration of genes differentially expressed between cell clusters and previously published research. The methodology utilized by gene expression calculations made by Monocle3's plot_cells_by_group function was used to score expression of these functionality associated genes. Subsequent gene expression scores were visualized in heatmaps produced by ComplexHeatmap in R (cluster_columns = FALSE, cluster_rows = FALSE).

scRNA seq T-Cell Subtype Classification

Using a reference tumor-infiltrating lymphocyte (TIL) atlas and lymphocytic choriomeningitis virus (LCMV)-specific CD8 T cell atlas, individual cells were aligned and annotated using ProjecTILs R package, version 0.5.1 (Andreatta et al. 2021). Normalized expression data from Seurat was provided as input and annotations for cell states were created using a nearest-neighbor algorithm ("cellstate.predict") and visualized "as-is", without a confidence threshold, in UMAP space. These assignments were further validated by confirming cell state annotations were the same after imposing a confidence score threshold of 0.5.

TIL and LCMV progenitor state enrichment for SIIN- and SIY- specific T-cells were analyzed by hypergeometric test (phyper, Stats R package; alpha = 0.05). Similarly, progenitor or exhausted cell state enrichment in cell clusters was evaluated by a hypergeometric test, described above.

Gene Signature Scores

Individual cells were scored for expression of previously published mouse gene signatures by calculating mean expression for each signature subtracted by aggregated expression of control signatures in Seurat using the AddModuleScore function. Prior to scoring, genes in signatures that were not detected in mouse scRNA seq data were removed from each published signature.

Progenitor exhausted and terminally exhausted signatures were derived from (B. C. Miller et al. 2019b), the deletional tolerance signature was derived from (Parish et al. 2009), and the anergy signature was derived from (Safford et al. 2005). Tc17 and Tc1 gene signatures were derived from (Linehan et al. 2018) and produced by alignment to the mouse genome (NCBI37/mm9) using Bowtie (version 1.2.3)(Langmead et al. 2009), quantification of feature counts with rsem (version 1.3.1)(B. Li and Dewey 2011), mm9 annotation with UCSC (genome.ucsc.edu), identification of pairwise differentially expressed genes with DESeq2 (Love, Huber, and Anders 2014) in R (version 3.6.0). CCR6+TCF7+ and CCR6-TCF7+ signatures were derived de novo in mouse scRNA-seq data Signatures were then filtered genes as described above. Individual cell scores were visualized in UMAP space.

TCR Clonotype Analysis

Clonotype identification and assignments to individual cells was performed using Cell Ranger as described above. Clonotypes are expected to originate from a single mouse; as such, rare clonotypes (22 of 652) that appear to originate from multiple mice, which is determined by cell hashing, were identified using Pandas in Python and excluded from downstream clonotype analyses. Most of the identified clonotypes were

composed of less than 5 cells (548 of 652) that is likely a result of their physiological abundance or extent of clonal expansion. Because it is impossible to distinguish whether clonotype sizes are a result of biological abundance or from technical dropout, clonotypes with less than 5 cells were removed from indicated analyses. Ultimately, 103 clonotypes remained.

Distribution of cells assigned to each clonotype across cell clusters (C0-C10) was evaluated by calculating the proportion of cells in each clonotype assigned to each cluster. This permits comparison of clonotypes with different sizes. For visualization in a heatmap, rows (clonotypes) of the heatmap were ordered by hierarchical clustering (method = "ward.D") using default pairwise euclidean distance and columns (cell clusters) were ordered by hierarchical complete-linkage clustering of pairwise Jaccard distance (hclust, method = "ward.D"). Subsequently, clonotypes were assigned to 12 clonotype clusters by hierarchical clustering, using euclidean distance as a distance metric (h = 0.6, cutree, R).

Identity-based annotations for each clonotype of this heatmap were generated using clonotype metadata. Gene-expression based annotations for each clonotype were generated by quantifying the proportion of cells with gene expression of Gzmb, Havcr2, Cx3cr1, Tcf7, Ccr6, and Il17a > 0.5. Clonotypes statistically enriched for expression of these genes were identified by hypergeometric test (phyper, Stats R package; alpha = 0.05). Additionally, to provide a metric for how equally distributed each clonotype is across clusters, we calculated the Gini index for each clonotype as a measure of sparsity.

To test statistical enrichment of SIIN- and SIY-specific clonotypes, a 2-dimensional 2-sample KS test was performed; differences in proportion of cells in each cluster or combination of clusters were visualized with an empirical cumulative distribution plot.

Clonotype sizes were calculated in Python using Pandas. To further evaluate how the size of a clonotype influences transcriptional profile, we assigned each cell a value equivalent to the size of its respective clonotype and visualized these values in UMAP space using Seurat.

ACKNOWLEDGEMENTS

We would like to acknowledge R. Herbst for interpretation of scRNA-seq phenotypes; T. Delorey and D. Phillips in the lab of A.R. for technical work on the mouse scRNA-seq experiment; C. Backlund for input on vaccination techniques; Koch Institute's Robert A. Swanson Biotechnology Center for assistance with flow cytometry and histology; Broad Genomics Platform and Whitehead Institute Genomics Technology Core for assistance with scRNA-seq; K. Yee, J. Texeira, K. Anderson, K. Mercer and M. Magendantz for administrative and logistical support.

This work was supported by the Howard Hughes Medical Institute (T.J.); the Ludwig Center at Harvard (S.S.); NIH U54 CA225088 (S.S.); NIH T32HL007627(G.G.); NCI Cancer Center Support Grant P30-CA1405; the Bridge Project, a partnership between the Koch Institute for Integrative Cancer Research at MIT and the Dana-Farber/Harvard Cancer Center (T.J. and S.S.); and fellowship awards from the Jane Coffin Childs Memorial Fund for Medical Research and the Ludwig Center for

Molecular Oncology at MIT (M.L.B.). T.J. is a David H. Koch Professor of Biology and

Daniel K. Ludwig Scholar.

CHAPTER 4

Amanda M. Cruz, Jose B. Cruz, Tyler Jacks

SPECULATIONS AND FUTURE DISCUSSIONS

Models for Evolution of Tumor Heterogeneity

One perspective of tumor evolution, discussed in Chapter 1, can be guided by Cancer Stem Cell (CSC) models of tumor heterogeneity, which operate under the premise that a stem-like population can seed other subclones of a tumor. If a singular CSC population existed that was responsible for the development of phenotypic tumor heterogeneity through random and stochastic differentiation, a brief period in which very little heterogeneity is observed would be expected, followed by a drastic increase in heterogeneity once a CSC population has matured. In KP lung tumors, discussed in Chapter 2, our findings are relatively consistent with these models, except that the number of cell states observed in the primary tumor eventually plateau. Presumably, the limitation of accessible transcriptional tumor cell states are a consequence of selective pressures. Although the number of transcriptional subpopulations of a tumor may increase with time, their associated transcriptional trajectories predicted by Monocle3 largely converge towards a *Hmga2*$^{+}$ metastatic-like state. This convergence may reflect increasing selective pressure, or fine-tuned adaptation to the tumor microenvironment. An important caveat to these speculations about selective pressures of the tumor

microenvironment is that the tumor microenvironment changes dynamically as tumors progress. For example, in Chapter 3, we demonstrate that the CD8 T cells of the tumor microenvironment undergo dynamic changes in functional differentiation that ultimately produce selective pressures that change over time. This highlights the possibility that tumor cell subpopulations that are abundant and are positively selected for at one stage of tumor progression may be negatively selected against or even eliminated at another stage of tumor progression. As such, while computationally predicted transcriptional trajectories of tumor cells may appear to converge towards an *Hmga2*$^{+}$ state, it is possible that this convergence reflects “pruning” of tumor cell subpopulations.

In consideration of the dynamically changing tumor microenvironment, it is entirely possible that multiple CSC states may be simultaneously present within a tumor in order to be better suited to adapt to certain kinds of selective pressures from the tumor microenvironment, but not others. Consequently, multiple populations may exist within a tumor that are functionally distinct, but all have stem-like abilities that flexibly allow tumors to generate diverse cell states in order to adapt to various forms of selective pressure. Conversely, one could imagine that a single CSC population exists, but itself undergoes changes over time as tumors progress that have meaningful functional implications. These hypotheses have been supported by the discovery of a functionally distinct Wnt producing niche in KP tumor cells (Tammela et al. 2017).

Many hypotheses that build upon the CSC model typically operate under contexts where the cell of origin of a tumor is believed to have homeostatic stem-like abilities. However, in many of these cases and in the case of AT2 cells in particular, which are believed to be the cell of origin in the KP model (C. F. B. Kim et al. 2005), the

potential of these stem cells are still lineage restricted. In order for a stem cell to give rise to the diverse cell states and dysregulated identities observed in tumors, they must undergo extensive functional changes with respect to differentiation potential to generate a heterogeneous tumor. Often, terminally differentiated cells are thought to have terminal evolutionary trajectories under physiological conditions. The associated stability of transcriptional states for terminally differentiated cells are often attributed to heritable changes in epigenetic landscape of terminally differentiated cells.

Fascinatingly, the changes in differentiation state of tumor cells challenge the permanence of transcriptional states that are usually associated with terminally differentiated cells. Tumor cells exploit the lack of permanence of epigenetic states in order to generate a greater diversity of transcriptional states. The biological implication of the chromatin landscape of a healthy, differentiated cell is to restrict the permutations of cellular machinery, circuitry, and programs that are accessible to the cell in order to restrict cellular responses to change and homeostatic processes to ultimately maintain the tissue and its overall function in the context of the entire organism. As such, in order for these cells to become capable of functionally differentiating into a diverse set of cell states, many epigenetic changes must occur. While these epigenomic states are reversible, they are very stable, and in order to become plastic, a substantial destabilizing force is required. One of the most apparent and widely accepted sources of oncogenic stress are those that result from destabilizing “hits” of oncogenic mutations.

Tumor evolution as a chaotic process

A system is considered to be chaotic if it is bounded, deterministic, and has characteristic hypersensitivity to perturbations (Toker, Sommer, and D'Esposito 2020).

Tumors may evolve as a result of sensitivity to tumor-initiating signals

For an organism to develop and sustain life, gene expression must be tightly regulated. This has inherently led to the evolution of genes, circuitry, machinery, and multicellular systems that suppress and prevent transformation or tumor formation. A single mutation can confer subtle effects on these tumor suppressive networks and systems. Over time, the effects of a single oncogenic point mutation can accumulate to cross a threshold that disrupts canonical processes enough to confer strong susceptibility to cancer development (H. Lee et al. 1999; J. M. Dunn et al. 1988; Lynch et al. 2015). To counteract this sensitivity, regulatory systems of development and homeostasis exert control over the biological processes that underlie cancer development through surveillance, repair, and feedback (Filipski et al. 2002; Bruchovsky et al. 1996; L. Huang and Mellor 2014). In many cases, sufficient control is maintained long enough for an organism to develop, mature, and age; when it is not, cancer is likely to occur. Temporal feedback to perturbation may also occur, supporting hypotheses that cells may be transformed and progress to a clinically detectable tumor over relatively long time scales (P. C. Nowell 1976). This is further supported by the clinical cases where measurable disease kinetics, such as primary tumor growth or rate of metastasis, progressively accelerate over time.

Evidence for sensitivity to transformation has been described in the field of cancer biology in numerous ways, particularly through characterization of individual

oncogenic mutations and clonal outgrowth of cells (i.e. tumor progression begins with change in a single cell) (S. Y. Luo and Lam 2013; Shlush and Hershkovitz 2015; Greaves and Maley 2012; H. Lee et al. 1999), which collectively provide evidence that a single, or few, mutations can sufficiently drive oncogenic transformation. Mathematical models of tumor evolution support clonal expansion of tumor subclones that contain mutations that confer specific mutational and fitness properties (Heide et al. 2018). However, not all mutations destabilize regulatory systems enough to cause cancer (Martincorena and Campbell 2015), which may be a consequence of the effects they confer, tissue specificity (García-Nieto, Morrison, and Fraser 2019), or protection from tumor suppressive pathways (Heuer et al. 2020; Xuyi Wang, Simpson, and Brown 2015). In turn, chaotic behavior of evolutionary trajectories are possible when regulatory systems become sufficiently destabilized.

Tumor evolution is deterministic and bounded

The extent of variation in clinical and experimental observations made throughout the course of tumor progression is superficially consistent with a model for stochastic, rather than deterministic, evolution throughout tumor progression. In Chapter 2, a time-series analysis of cell states in KP lung tumor evolution was attempted through single-cell RNA sequencing. In this study, the observed tumor cell states were highly reproducible across mice, but varied in abundance across samples harvested from the same time point. This manifested in variation both at the state of the overall tumor and of its respective tumor cells.

At a given time point, the overall grade of a tumor that is harvested from the lungs of the animal follows a probability distribution that is dependent on time (Jackson

et al. 2005). In this way, the probability of harvesting a tumor that has a high histological grade is higher at longer time points than it is at shorter time points after tumor initiation. Additional complexity is created by intra-tumor heterogeneity of cell states; abundance of cell states observed in heterogeneous tumors are a function of tumor progression (Marjanovic et al. 2020). If a higher grade tumor is harvested, the state of any given cell within that tumor follows probabilistic distribution patterns that are dictated by the grade or progression of the tumor it belongs to.

Notably, the distribution of tumor cell states at a given time point *is still dependent* on time. As such, tumor evolution can be perceived as a deterministic process at the macroscopic level, governed by biological noise generated at least partially from the stochastic influence of latent variables. Although these variables are likely to affect tumor cells somewhat stochastically, across many tumors and mice, their influence exhibits predictable behavior and is consistent with mathematical models of stochastic processes (Ditlevsen and Samson 2013). These latent variables produce variation in timescales associated with tumor evolution, discussed in Chapter 1 Section 5.1, such that transcriptional changes occur in tumor cells in a stochastic manner. This temporally manifests in evolutionary behavior consistent with a nonstationary process, a function that is influenced by a stochastic process that itself is macroscopically deterministic with properties that vary in a time-dependent manner (Gagniuc 2017). In other words, the time scales over which tumor evolution occurs is dictated by stochastic biological influences that ultimately make time series based analyses unreliable. However, the stochastic behavior of these biological influences have properties that are themselves deterministic, and evolve as a function of time.

Analysis of nonstationary processes is a complex issue frequently encountered in data forecasting. Generalizations made about nonstationary processes have suggested that behavior of nonstationary processes can be approximated by assuming local stationarity, proposed by Dahlhaus (Dahlhaus and Giraitis 1998), wherein the properties that define the stochastic effect of biological influences on time scales of tumor evolution are assumed constant across small intervals of time. This approach inherently creates and defines multiple kinds of time scales over which tumor evolution can be examined. The first is the absolute or chronological time scale of evolution of a given tumor, and is regarded as ‘observed time’ or ‘absolute time’. The second is ‘rescaled time’, which is defined by deterministic changes between two states of the system (Van Bellegem and von Sachs 2004). Characteristically, in between states associated with the beginning and end of the process being studied, the spectrum of states observed will increase in an asymptotic manner. This is largely consistent with the observed convergent evolution of KP lung tumor cell states and with bounded behavior of the system.

Biological pseudotime is a rescaled unit of time

If we reconstruct tumor evolution on an axis of pseudotime, rather than absolute time, the systems which dictate tumor cell state appear to change in a dynamical and deterministic manner. If construction of pseudotime across longitudinally-sampled tumor cells indeed corrects for the stochastic effects of latent variables that produce confounding biological noise, as discussed in Chapter 1 and in the section above, the reproducible nature of KP lung tumor evolution loosely fits the criteria for deterministic systems. Pseudotime can be considered an “arrow” of time that, although governed by

physical processes that proceed as a function of absolute (symmetrical) time, described in Chapter 1 Section 5.1, manifest in an 'asymmetric' manner (Roberts and Quispel 1992). Simply, the amount of time that is required to generate cell states with equal distance along an axis of biological pseudotime is not constant across tumor evolution (**Figure 1**).

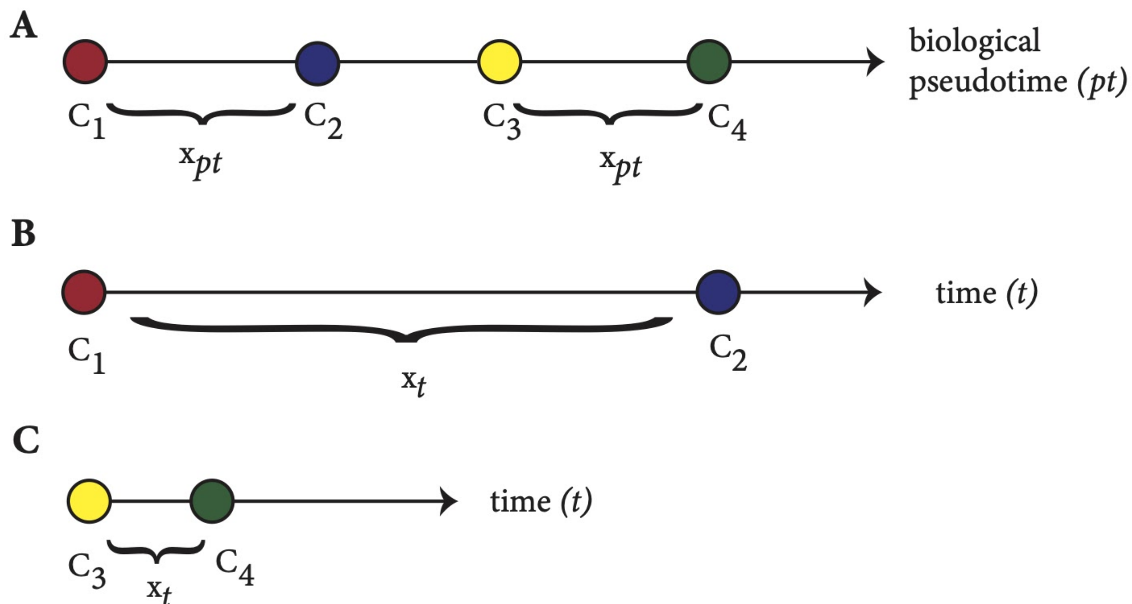


Figure 1. A visualization of asymmetry in biological pseudotime. A theoretical model for tumor evolutionary trajectories, *f*, is differentiable with respect to absolute time, *t*, or biological pseudotime, *p*, such that a distance, *x*, can be calculated between two cells with different evolutionary states.

$$x_t = \int f(t)\, dt \text{ and } x_{pt} = \int f(p)\, dp$$

(**A)** Assume two pairs of cell states in tumor evolution, (C_1, C_2) and (C_3, C_4) exist with equidistant relative positions in biological pseudotime (x_{pt}). The absolute or chronological time required for a cell to evolve from state C_1 to C_2 , (**B**) given by

$x_t = \int_{C_1}^{C_2} t \, d(t)$ can be nonequivalent to that of evolution from state C_3 to C_4, (**C**) given by

$$x_t = \int_{C_3}^{C_4} t \, d(t)$$

Evolution of tumor cell state is a time-dependent process. Evidently, this implies tumor evolutionary processes roughly follow behavior of a dynamical system, which is defined by its ability to be differentiated with respect to time (Katok and Hasselblatt 1995). As it is proposed in this thesis, tumor evolution is also differentiable with respect to biological pseudotime, a variable scale of absolute time. Importantly, pseudotime is differentiable (or dependent on) absolute time; as such, biological pseudotime likely can be described by a multivariate function that is dependent on changes in gene expression that occur as a consequence of the passage of absolute time (**Figure 1**).

The argument that tumor evolution occurs in a deterministic manner seemingly contradicts long-held models of stochastic tumor evolution (Foo, Leder, and Michor 2011; Bose and Trimper 2009; M. J. Williams et al. 2016). Notably, these models functionally define cell state in tumor evolution by mutational status; in contrast, in this work, tumor evolutionary status is assessed by cell states defined by transcriptional or chromatin accessibility profile. Tumor evolution appears to be macroscopically reproducible, but occurs with a high degree of variation from tumor to tumor, mouse to mouse, and possibly even cell to cell with respect to time. For this reason, it is likely that the relationship between biological pseudotime (tumor evolution) and absolute time cannot be defined in a reproducible manner without statistical modeling, and is therefore

stochastic. Tumor evolution, however, is deterministic along the axis of and dynamical with respect to biological pseudotime.

Differentiability with respect to pseudotime, a hallmark of dynamical systems, may seem counter-intuitive in the context of experimental approaches often utilized in biological research. Biologists usually seek to explain biological phenomena with respect to absolute, chronological time, or through time-series. Systems-level approaches to modeling biological processes in these time series led to the emergence of mathematical models that describe some of these processes as having behavior characteristic of chaotic systems, frequently in the context of population dynamics and enzyme kinetics (Olsen and Degn 1985; Parry 1979). In the era of single-cell technologies, biological phenomena are profiled in high-dimensional transcriptional or chromatin accessibility space that has created new opportunities to model temporal dynamics of heterogeneous and noisy biological systems.

Phenomena of asymmetry in time or variant time scales have been repeatedly described in other contexts, particularly in physics (Roberts and Quispel 1992). Many experimental descriptions of chaotic systems have a similar behavior to that of KP tumor evolution, in which chaos can be deterministically modeled in the absence of stochastic terms and may manifest in bifurcation of the behavior of the system towards stability or chaos (Glass 2009). In this case, periods of time or pseudotime between branch events created through bifurcation and chaotic behavior can be thought of as the time scales over which assumptions of local stationarity are true. In other words, I predict periods of stasis exist in tumor evolution in which selective pressures are relatively constant and biological systems exist in a state of equilibrium. Further,

between these time periods, and amongst time periods with similar properties across tumor evolution, approximations can be made about the behavior of evolutionary trajectories that can be described in a piecewise manner (Van Bellegem and von Sachs 2004).

The remaining requirement of chaotic systems, which require bounded behavior, is largely satisfied by dynamics of KP lung tumor evolution. If transcription in tumor cells became entirely dysregulated, trajectories of tumor cells would follow the dynamics of a random walk across transcriptional space or a high-dimensional manifold. Stated plainly, gene expression in cells would simply become random. Of course, not every system in a tumor cell becomes dysregulated. An important caveat to this perspective is that bounded behavior of accessible states may occur in *any* biological process as a consequence of constraints imposed by gene regulatory networks (S. Huang 2012) and may not be unique to tumor evolution itself. Complete transcriptional dysregulation is arguably never observed in a viable cell because it would not be able to sustain life. That said, life can still be sustained in a wide variety of cell states that are apparent in comparison of cell states across different tissues. As such, it is still remarkable that reproducibility can be observed in KP lung tumor evolution, particularly despite latent influences that cannot be controlled for that generate biological noise.

Comparisons to models of mutational tumor cell evolution

Models of tumor evolution in the context of mutations have been largely focused on distinguishing highly recurrent destabilizing alterations, whose fitness benefits are dependent on stochastic processes such as mutation, described in Chapter 1 Section

1.3 and by (Hanahan and Weinberg 2011; Dagogo-Jack and Shaw 2018) from mutations that do not confer selective advantages.

Acquisition of additional mutations are not required for and does not drive tumor evolution in the KP model (DuPage et al. 2011), as such, stochastic random mutation is unlikely to have a strong influence on tumor evolutionary trajectories in this specific context. However, in other studies of the K and KP model, introduction of additional mutations confer an effect on cell state in a manner that affects the evolutionary fitness, and therefore trajectories, of tumor cells (Rogers et al. 2018). As such, conclusions generated from analysis of the KP model are inherently limited in their ability to model response to acquisition of mutations in other cancer-driving genes. That said, further work describing evolutionary trajectories in KP tumors with additional genetic mutations may lead to insight on how mutations in such genes may affect overall disease progression.

Despite the lack of acquired mutations in the KP model, evolution of KP lung tumors may share similar evolutionary dynamics to observed mutational trends in other tumor contexts. In the context of mutational status, tumor evolution and heterogeneity has been previously proposed to be a deterministic process (M. J. Williams et al. 2016), in which neutral evolution governs mutational-trajectories of tumor cells, and dictates equal growth rates of clones with distinguishing neutral mutations. This model is derived from the Luria-Delbruck model of bacterial evolution. Despite support for this model in studies with mice (Driessens et al. 2012), the interpretation of the data presented by Williams et. al. drew criticism because it assumes a constant rate of cell death, mutation, and proliferation as well as synchronous division (Tarabichi et al. 2018).

However, the solution proposed in the Williams model agrees with convergent solutions generated from simulations in which these parameters are stochastically defined (Heide et al. 2018). When incorporated with models where mutation and subsequent subclone fitness is defined stochastically, the Williams model performs well, except in extreme situations. It performs poorly in situations where a subclone either acquires a mutation that confers a strong selective advantage but does not decrease genomic stability, or when a subclone acquires a mutation with moderate selective advantage that decreases genomic stability (Heide et al. 2018). In these situations where the model fails, subclonal expansion is likely driven by a selective advantage.

Transcriptional Bursting May Underlie Heterogeneity of Cell State

Although the variation in kinetics of tumor progression creates analytical challenges in cancer biology, it is believed to be caused by variables with meaningful biological influence. This kinetic variation is often treated as biological noise, but nonlinear kinetic processes are prevalent in biology (“Nonlinear Oscillations in Chemical and Biological Systems” 1991), and latent variables that cause variation in tumor progression kinetics could provide meaningful changes or selective pressures that ultimately drive tumors to become heterogeneous.

In this thesis, the majority of discussions of cell state are based upon transcriptional profiles of individual cells. While it is difficult to model transcriptional dynamics in experimental systems that are as complex as mouse models, several pieces of evidence suggest that the kinetics of transcription in individual cells is nonlinear and subject to bursting behavior that is coordinated by precisely regulated systems (“Enhancer Control of Transcriptional Bursting” 2016).

While many regulatory systems in a tumor become disrupted, many studies have suggested that topologies of intracellular regulatory circuitry are not entirely lost, but are instead rewired (“Gene Networks with Transcriptional Bursting Recapitulate Rare Transient Coordinated High Expression States in Cancer” 2020). Some studies have further suggested that bursting behavior of transcriptional networks in a tumor have the capacity to generate functionally distinct rare tumor subpopulations (Zhao, Cheng, and Zhao 2017). In this study, it is proposed that highly connected gene regulatory network topologies result in higher throughput of gene expression outputs, both in terms of expression and duration (Zhao, Cheng, and Zhao 2017). Importantly, the amplitude and frequency of burst kinetics in this study are generally constant and maintained throughout transcriptional responses in highly connected gene networks. In gene networks with less connectivity, this regularity is partially lost (Zhao, Cheng, and Zhao 2017). Indeed, outside of biology, burst kinetics have been extensively used to model network constraints of information trafficking (Cruz 1991b). In this regime, it was demonstrated that engineering multiple regulators within a network with burst kinetics can increase the throughput of the system (Cruz 1991a), and parallels observations made by (Zhao, Cheng, and Zhao 2017). Taken together, if transcriptional burst kinetics are assumed, rewiring and disruption of regulatory networks in tumor cells may lead to aberrant transcription that can drive evolution of heterogeneous subpopulations of cells.

In the KP lung tumor model, the AT2 cell of origin is a well-differentiated epithelial cell whose identity is maintained except in response to injury (Paris et al. 2020). Thus, in untransformed cells where gene regulatory networks are intact, transcriptional outputs are stable and result in stable, regulated expression of AT2 identity genes. In

contrast, untransformed T cells have characteristic functional heterogeneity that is an explicit feature of a productive overall T cell response (Gong, Linderman, and Kirschner 2014). Consequently, T cells have contrasting gene regulatory dynamics when compared to AT2 cells. In the setting of the tumor microenvironment, the factors that control heterogeneous functional differentiation are 'tuned' in a manner that can lead to dysfunctional T cells (Hashimoto et al. 2018).

In sum, these parallel observations suggest that transcriptional heterogeneity observed in single cells, including both CD8 T cells and tumor cells, may be symptomatic of underlying transcriptional regulatory dynamics. Importantly, loss of regulators in gene networks of tumor cells may result in differentiation (or change in differentiation status) that ultimately generates heterogeneous subclones of a tumor. Analytically, the ability to fit the observed transcriptional noise in tumor cells to models that describe multi-level burst kinetics is currently extremely limited. This is due to the fact that there are dynamic changes in transcriptional network topology that occur in response to changes of regulatory dynamics. However, in T cells, regulatory dynamics are exploited and tuned, but transcriptional network topology is not necessarily altered, which has led to emerging studies that describe the role of transcriptional bursting in T cell differentiation (DeMarino et al. 2020).

Relationships between Cell Signaling and Cell Identity

Signaling Architecture

If we think about applications of information theory to understanding the changes that result from the onset of oncogenic signaling - in our model, driven by Kras, we can think of transcriptional outputs from signaling as being constrained by channel capacity

of upstream signaling components. In the KP model, the “channel capacity” of KRAS can be interpreted to change through an oncogenic mutational event. Thereafter, KRAS activation is no longer rate limiting, and provides a constant, saturated signal to downstream signaling components, such as PI3K and MAPK signaling. This single oncogenic mutation, however, does not change the “channel capacity” of these downstream signaling events. It is important to note that these “channels” (i.e. signaling components) are not defined simply by their biochemical capacity to signal to downstream effectors as intrinsically defined through their protein structure. Instead, the parameters of each signaling “node”, such as components of PI3K and MAPK signaling pathways, are a function of the activity and availability of other signaling components that serve to change the “channel capacity” of signaling nodes. Stated otherwise, the ability for downstream signaling pathways of KRAS to respond to constant oncogenic signals from $KRAS^{G12D}$ gain-of-function activity is influenced by the architecture of further downstream effector networks. As such, in a cellular signaling network, the capacity of these channels are dynamic. An important limitation to this reasoning is that signaling pathways do not act in isolation. Signaling pathways downstream of KRAS activation are not solely activated by RAS signaling. They integrate both positive and negative signals from pathways that can act in parallel to RAS or upstream of it. These upstream or parallel signaling pathways can provide feedback on pathways downstream of RAS to tune channel capacity. In isolation, the activity of downstream signaling components behave according to their intrinsic constraints. In practice, however, signaling pathways act within a network, and the impacts of a saturating oncogenic signal will affect the entire signaling network. Importantly, it is very possible that the

global response to this form of oncogenic stress underlies differences in how cells respond to various forms of oncogenic stress.

In AT2 cells, the signaling network of a healthy and normal cell can be thought to be specialized to support AT2 function. In this manner, it is optimized for at least 3 purposes: 1) to maintain AT2 cell identity, 2) to respond to cellular stresses encountered by AT2 cells, and 3) to create a “kill-switch” to prevent proliferation of cells with some kind of defect. Cellular signaling events relay information about the extracellular environment of the cell. The purpose of AT2 cells is to produce surfactants to prevent the lung from collapsing and to participate in structural formation of the lung epithelium (see Chapter 1). Thus, the steady-state signaling network of AT2 cells has been optimized to perform these exact functions. These functions could include detection of biomechanical stress, production of surfactants, and proliferation in response to injury. This functional optimization is not mutually exclusive with maintenance of cell identity, as cell features of cell identity, particularly cell shape, are dictated by functionality. Stated plainly, form fits function. Lastly, the kill-switch is important to ensure homeostasis in the lung epithelium, and kill cells that have acquired DNA damage. This could prevent transformation or outgrowth of cells transformed with $Kras^{G12D}$ and may underlie the discrepancies observed between the number of cells infected with virus to initiate tumors and the number of over lesions and tumors that are later observed in the lungs of these animals.

An interesting nuance of AT2 cells is that they have been shown to be able to trans-differentiate into AT1 cells in response to certain kinds of stress (see Chapter 1), which represents an outcome characteristic of all of the above described purposes of a

cellular signaling system. The primary difference in this instance is that the programmed response to this particular stress signal is a change in cell identity. If a cell survives acute saturation of activating signals in one branch of its intracellular signaling network, it may adapt to the signal through feedback that inherently changes the topology or dynamics of the network itself. Because these adaptive changes of cellular signaling alter cell functionality and thus cell identity, it is possible trans-differentiation is a programmed mechanism to change the dynamics of cellular signaling in canonical, non-transformed AT2 cells.

This is conflicting in some aspects. One of the major purposes of the optimized cellular signaling dynamics of AT2 cells is both to maintain cellular identity, but also to provide a programmed response to change cellular identity. The difference between the two outcomes is a function of stress signals. Under a highly simplified boolean logic, the absence of a trans-differentiation stress signal could result in maintenance of AT2 cell identity, whereas the presence of this signal results in trans-differentiation. Inherent to this theory is the ability to sense this stress signal, and that this sensing mechanism has the capacity to change cellular signaling dynamics as well as transcriptional and epigenomic state.

It is well established that AT2 cells and AT1 cells have distinct transcriptional profiles (Travaglini et al. 2020), which almost certainly confer different topologies or architectures of cellular signaling networks. In line with this logic, epigenetic changes are also likely to mediate changes in these topologies. In contrast to the AT2 and AT1 states observed in KP lung tumor evolution, an important distinction about stress responses that induce differentiation of AT2 cells is that cell state changes occur

independently of mutational changes in signaling components. The intrinsic biochemical nature of signaling components remains unchanged. As such, introducing a mutation in KRAS and thereby increasing its signaling output may exploit endogenous and programmed responses to stress signals that would normally cause cells to differentiate into AT1 cells. In the context of transformation, this may initiate destabilization of cell state.

When oncogenic KRAS is introduced into AT2 cells, the initial events that follow induction of oncogenic KRAS must operate within the constraints of the signaling network that underlies AT2 cell identity. Introduction of oncogenic Kras inherently changes the cellular signaling network architecture, but only downstream of Kras. Thus, the “channel capacity” for downstream signaling effector networks are unchanged and operate under their normal endogenous constraints. This implies 2 possible modes of responses to oncogenic Kras signaling, which are not mutually exclusive: it can be interpreted as a stress signal to differentiate into AT1 cells, or, it can cause endogenous feedback systems to respond to changes in KRAS signaling by activating a “kill switch”. If such a kill-switch response occurred invariably, introduction of oncogenic KRAS, even with simultaneous loss of P53, would invariably lead to cell death. Thus, the response to oncogenic KRAS is not unilateral, despite the fact that the oncogenic signals it provides are constant and inherent to the gain-of-function G12D mutation. This may, in fact, create selective pressures that drive dysregulation of cell identity in KP lung tumor evolution: trans-differentiate, or die.

Etv4 expression is induced at early stages of tumor initiation, although at a low level (Mainardi 2013). Although *Etv4* appears to be expressed relatively late in tumor

progression and is not expressed in the normal lung, discussed in Chapter 2 and in (Mainardi 2013), cells which have lost *Etv4* appear to be selected against. Similarly, *Etv5* is also required for Kras tumor initiation in the lung (Z. Zhang et al. 2017), but is expressed in untransformed AT2 cells and has patterns of gene expression changes during tumor evolution that are distinct from those of *Etv4*. The Pea3 transcription factor subfamily is particularly important for processes important for lung development and cell identity, and are also effectors of MAPK signaling (O'Hagan and Hassell 1998; "ATXN1L, CIC, and ETS Transcription Factors Modulate Sensitivity to MAPK Pathway Inhibition" 2017). Importantly, the ETS transcription factors coordinate with multiple other transcription factors and are known to influence cell identity (Marra and Wingert 2016; "A Systematic Approach to Identify Candidate Transcription Factors That Control Cell Identity" 2015). Given the demonstrated importance of the Pea3 transcription factors in KP lung tumor evolution, it is possible these transcription factors mediate cell identity changes. However, further characterization is required to better understand the relationships between the ETS transcription factors in KP lung tumor progression.

Refining models of the functional differentiation of tumor-specific CD8 T Cells

The primary conclusion from the study discussed in Chapter 3 is that the response a CD8 T cell has to an antigen is more heterogeneous than previously thought. In fact, the influence of antigen dominance hierarchies represent conclusive evidence that the functional behavior and differentiation of CD8 T-cells is partially determined by the dynamics of the overall anti-tumor response and are not unilaterally determined by the peptide sequence of an epitope or the neoantigen itself. This agrees with the discovery that exogenous antigen specific CD8 T-cells have productive immune

responses to antigens that endogenous CD8 T-cells do not respond to (Strønen et al. 2016).

One of the ways this multi-faceted behavior of the immune response manifests biologically is through subtle, but functionally important, transcriptional changes that in turn affect the functionality of cytotoxic T-cells. For example, T-cells specific to the subdominant SIY antigen have very similar transcriptional profiles to that of cells specific to the dominant SIIN antigen. Functional differentiation trajectories appear to be somewhat stochastic, but overall, SIIN specific cells appear to be poised to differentiate into states that appear more dysfunctional, whereas SIY specific cells are poised to remain in a slightly less differentiated state that ultimately affects their cytotoxic functions. Burger et. al. later went on to show that these phenotypes are not observed when antigens are expressed alone.

Many immunology studies are being conducted with intent to better characterize properties of neoantigens that are capable of stimulating a productive anti-tumor immune response. However, the study presented in Chapter 3 provides clear evidence that response to neoantigens is meaningfully impacted by other neoantigen responses in anti-tumor responses. In light of the complexity of the CD8 T cell response alone, when the influence of other tumor-related lymphocytes are considered, the tumor-immune interaction dynamics becomes extraordinarily complex. Research in immunology has been conducted at single-cell resolution for many years via flow cytometry, and it has long been appreciated that responses are both complex and heterogeneous. Through many years of characterizing and refining these heterogeneous responses, enough of a biological framework was provided to

meaningfully interpret the transcriptional states of CD8 tumor cells in Chapter 3. As sources of heterogeneity that drive diversity of immune responses become better understood, this will create opportunities for applications of systems biology that can be used to create highly informed and robust models of immune responses in cancer. Ultimately, this can be incorporated with models of tumor evolution to better understand the interplay of the immune system and tumor progression.

Conclusions

The tumor microenvironment consists of multiple systems that interact in a dynamic manner and ultimately determine how tumors evolve over time. Approaches in biological research are largely reductionist in nature, and have been useful in studies that have allowed generalizations about how a gene acts in multiple biological contexts. However, this kind of approach has inherently created limitations in the kinds of discoveries that can be made in these studies. A notable historical example of these limitations comes from efforts to characterize and understand cancer as a singular disease. Intuitively, this made sense, as many hallmarks of cancer are observed across cancers that arise from different tissues (Hanahan and Weinberg 2011). Today, and over the last few decades, it is widely appreciated that cancer is a collection of diseases which manifest and progress in a tissue specific manner. Still, for many years, repeated attempts were made to discover “pan-cancer” driver genes, despite the fact that there is compelling evidence for tissue specific behavior even amongst the most widespread oncogenically implicated genes. Many oncogenic signaling pathways are ubiquitous across tissues of an organism, and yet little efforts are made to understand how

context-specificity is achieved in these systems or how they become dysregulated in cancer. These specificities are largely ignored in cancer research.

With the resolution afforded by single-cell technologies, and as biological research is conducted at increasingly higher resolution, many studies have described variation in the resulting high-dimensional data that have made it difficult to make meaningful biological conclusions from it. In some cases, the variation captured at these resolutions have explained conceptual inconsistencies between conclusions made from bulk profiling studies and the behavior of the underlying biological system (Li, Amy, and Ph. D. Massachusetts Institute of Technology 2018). Single-cell profiling has made it abundantly clear that biological processes are extraordinarily heterogeneous at the single-cell level, even in homeostatic processes. For this reason, it has been particularly difficult to use these data to predict the functionality of cell states captured in single-cell studies. Additionally, in many single-cell studies of tumor cells (including those presented in this thesis), the conclusions made from empirical data largely challenge existing models of tumor evolution and heterogeneity.

Models of computational biology are frequently disregarded by experimental biologists because they often require assumptions that are incompatible with experimental systems. As biological research generates increasingly more data on biological processes, the field will, by necessity, become increasingly more dependent on bioinformatic and computational interpretation of these data. Integration of descriptive approaches utilized by experimental biologists and explicit definition of “ground truths” is one way that experimentalists can reduce assumptions required for computational modeling of biological processes. In turn, this will also require scientists

responsible for interpretation of the data to define and continuously challenge assumptions they make about the biology of the system they are studying. This is extremely important when handling high-dimensional data, as the mathematical transformations required to approximate trends and structures of high-dimensional data can impart artifacts that can lead to erroneous interpretations.

Importantly, these technologies have allowed biologists the opportunity to make meaningful connections between stochastic behavior of biological systems and biological processes. Stochastic processes have been long described in biology in settings where biological processes can be measured with enough resolution to observe stochastic variation (Toker, Sommer, and D'Esposito 2020). Importantly, the behavior of these processes are macroscopically deterministic. Despite the fact that many single-cell omic datasets have been published to date that profile biological processes with similar properties to those that are chaotic, very few interpretations of these data attempt to define the properties of these stochastic processes.

In the tumor evolution study presented in this thesis, the temporal inter-tumoral and intra-tumor heterogeneity of the KP mouse model, which manifests through variation in both cell state and time, have permitted relatively comprehensive longitudinal profiling of evolutionary processes in primary tumors. These analyses are largely contingent on the assumption that the composition of a heterogeneous tumor simultaneously reflects its evolutionary past and present. Critically, this assumption is supported by empirical evidence of tumor cell states generally associated with early stages of tumor progression being represented in the most aggressive and advanced tumors. The majority of the cells profiled in this study were derived from a limited

number of timepoints that are weeks apart; the comprehensive sampling of this evolutionary process was largely an unexpected benefit of variation in the model. Notably, these studies are conducted with syngeneic mice in a highly controlled environment and utilize experimentally defined mutations. In spite of this controlled environment, the KP lung tumor model still has extensive, but reproducible variation. With this same reasoning, it may be possible to extend the assumption that heterogeneous cell states reflect the evolutionary past and present of a tumor to biological settings in which longitudinal sampling is impossible, or in settings with uncontrolled environments (e.g. patient derived tumor samples).

To better understand how heterogeneous evolutionary behaviors of tumor cells affect other cells of the tumor microenvironment, the dynamics of response to tumor-specific antigens in CD8 T cells was profiled in the KP Lung tumor LucOS model. Interestingly, many of the biological conclusions made from this study, which is conducted at a single time point after tumor initiation, agree with studies that longitudinally profiled many of these same cells across a series of timepoints (Schenkel et al. 2020). Arguably, these similarities are a result of the same temporal variation in tumor progression and evolution that permitted comprehensive longitudinal profiling of the KP tumor cells themselves. If this is true, this similarity across these datasets provides compelling evidence for the strength of relationships between tumor cells and the other biological systems they interact with, in this case, CD8 T-cells.

The tumor-specific CD8 T-cell compartment represents only one system of many in the tumor microenvironment, which interacts not only with tumor cells, but also other cells in the immune system, endothelial cells, and fibroblasts. Although biological

systems, especially in autochthonous models of cancer, are often studied in isolation, the interactions between these systems are always present and introduce additional layers of complexity to overt biological processes such as tumor evolution. Interactions between biological systems have been extensively modeled in reductionist and highly experimental biological settings, but have rarely been modeled in settings such as tumor evolution beyond probabilistic models of variant allele frequencies. In tumor cells, the topology of gene regulatory networks evolve over time and limit the ability to model intracellular biological systems of tumor cells.

Stated otherwise, assuming the underlying variation observed in tumor progression is entirely random will inherently limit discoveries about latent influences or dynamics that give rise to heterogeneity in cell state. Identification of these latent variables is impossible if studying biological processes with entirely reductionist approaches. The conclusions made from studies in this thesis largely ignore the influence of other systems and simply treat the observed variation as biological noise. However, through modeling networked interactions both inter- and intra-cellularly, the underlying dynamics that produce noise can be identified. Critically, these identified dynamics may have meaningful influences over tumor progression.

The other biological systems that influence tumor cells and CD8 T cell phenotypic trajectories occur within individual cells and are defined by intracellular signaling, genetic, metabolic, and mechanical networks. The existing interpretations are dependent on the treatment of observed cell states and transcriptional changes as a network, but do not explicitly consider constraints of gene regulatory (or any other) networks, because they have not been characterized with the required precision. In

future work, introduction of genetic perturbations may allow estimation of these network constraints and may provide more robust interpretations or conclusions.

While interactions between biological systems create major analytical complications in interpretation of biological data, there are many existing approaches which make sense of high-dimensional data in non-biological settings that are also influenced by other systems that cannot be controlled for. Moreover, the descriptive and speculative discussions made in the final chapter of this thesis are derived from insights made from mathematical modeling of complex data across many different fields, including particle physics, astrophysics, data science, machine learning, natural language processing, computer vision, quantum and statistical mechanics, macroeconomics, electrical engineering, and meteorological forecasting. This is not to say that the phenomena described in these non-biological settings are the same as those in biological settings. Instead, I assert that greater importance should be placed upon quantitative approaches used to interpret data from higher-order systems with stochastic behavior, which in turn, may provide systems-level insights in complex biological settings such as tumor evolution.

REFERENCES

Beauchemin, Kyle Jaymes. 2016. “Characterization of the Murine Lung Transcriptome during Pre- and Post-Natal Development in Three Inbred Strains of Laboratory Mice.” Edited by Carol J. Bult and Viravuth P. Yin. Ann Arbor, United States: The University of Maine. https://www.proquest.com/dissertations-theses/characterization-murine-lung-transcriptome-during/docview/1844967052/se-2.

AbdulJabbar, Khalid, Shan E. Ahmed Raza, Rachel Rosenthal, Mariam Jamal-Hanjani, Selvaraju Veeriah, Ayse Akarca, Tom Lund, et al. 2020. “Geospatial Immune

Variability Illuminates Differential Evolution of Lung Adenocarcinoma." *Nature Medicine* 26 (7): 1054–62.
Abler, Lisa L., Suzanne L. Mansour, and Xin Sun. 2009. "Conditional Gene Inactivation Reveals Roles for Fgf10 and Fgfr2 in Establishing a Normal Pattern of Epithelial Branching in the Mouse Lung." *Developmental Dynamics: An Official Publication of the American Association of Anatomists* 238 (8): 1999–2013.
Coley, William B. "A Brief History of Immunotherapy." 2014. Targeted Oncology. August 21, 2014. https://www.targetedonc.com/publications/special-reports/2014/immunotherapy-issue3/a-brief-history-of-immunotherapy.
Agarwal, M. L., W. R. Taylor, M. V. Chernov, O. B. Chernova, and G. R. Stark. 1998. "The p53 Network." *The Journal of Biological Chemistry* 273 (1): 1–4.
Alberg, Anthony J., and Jonathan M. Samet. 2003. "Epidemiology of Lung Cancer." *Chest* 123 (1 Suppl): 21S – 49S.
Alexandrov, Ludmil B., Young Seok Ju, Kerstin Haase, Peter Van Loo, Iñigo Martincorena, Serena Nik-Zainal, Yasushi Totoki, et al. 2016. "Mutational Signatures Associated with Tobacco Smoking in Human Cancer." *Science* 354 (6312): 618–22.
Amini, Sasan, Dmitry Pushkarev, Lena Christiansen, Emrah Kostem, Tom Royce, Casey Turk, Natasha Pignatelli, et al. 2014. "Haplotype-Resolved Whole-Genome Sequencing by Contiguity-Preserving Transposition and Combinatorial Indexing." *Nature Genetics* 46 (12): 1343–49.
Anderson, Kristin G., Katrin Mayer-Barber, Heungsup Sung, Lalit Beura, Britnie R. James, Justin J. Taylor, Lindor Qunaj, et al. 2014. "Intravascular Staining for Discrimination of Vascular and Tissue Leukocytes." *Nature Protocols*. https://doi.org/10.1038/nprot.2014.005.
Andoni, Alexandr, and Piotr Indyk. 2006. "Near-Optimal Hashing Algorithms for Approximate Nearest Neighbor in High Dimensions." In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, 459–68.
Andreatta, Massimo, Jesus Corria-Osorio, Sören Müller, Rafael Cubas, George Coukos, and Santiago J. Carmona. 2021. "Interpretation of T Cell States from Single-Cell Transcriptomics Data Using Reference Atlases." *Nature Communications* 12 (1): 2965.
Antonia, Scott J., Augusto Villegas, Davey Daniel, David Vicente, Shuji Murakami, Rina Hui, Takayasu Kurata, et al. 2018. "Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC." *The New England Journal of Medicine* 379 (24): 2342–50.
Appleman, Leonard J., and Vassiliki A. Boussiotis. 2003. "T Cell Anergy and Costimulation." *Immunological Reviews* 192 (April): 161–80.
"A Stochastic Model in Tumor Growth." 2006. *Journal of Theoretical Biology* 242 (2): 329–36.
"A Systematic Approach to Identify Candidate Transcription Factors That Control Cell Identity." 2015. *Stem Cell Reports* 5 (5): 763–75.
"ATXN1L, CIC, and ETS Transcription Factors Modulate Sensitivity to MAPK Pathway Inhibition." 2017. *Cell Reports* 18 (6): 1543–57.
Baghban, Roghayyeh, Leila Roshangar, Rana Jahanban-Esfahlan, Khaled Seidi, Abbas

Ebrahimi-Kalan, Mehdi Jaymand, Saeed Kolahian, Tahereh Javaheri, and Peyman Zare. 2020. “Tumor Microenvironment Complexity and Therapeutic Implications at a Glance.” *Cell Communication and Signaling: CCS* 18 (1): 59.

Bahar, Rumana, Claudia H. Hartmann, Karl A. Rodriguez, Ashley D. Denny, Rita A. Busuttil, Martijn E. T. Dollé, R. Brent Calder, et al. 2006. “Increased Cell-to-Cell Variation in Gene Expression in Ageing Mouse Heart.” *Nature* 441 (7096): 1011–14.

Baitsch, Lukas, Petra Baumgaertner, Estelle Devêvre, Sunil K. Raghav, Amandine Legat, Leticia Barba, Sébastien Wieckowski, et al. 2011. “Exhaustion of Tumor-Specific CD8+ T Cells in Metastases from Melanoma Patients.” *The Journal of Clinical Investigation* 121 (6): 2350–60.

Balkwill, Frances R., Melania Capasso, and Thorsten Hagemann. 2012. “The Tumor Microenvironment at a Glance.” *Journal of Cell Science* 125 (Pt 23): 5591–96.

Balkwill, Fran, and Alberto Mantovani. 2001. “Inflammation and Cancer: Back to Virchow?” *The Lancet* 357 (9255): 539–45.

Bandara, Samuel, Johannes P. Schlöder, Roland Eils, Hans Georg Bock, and Tobias Meyer. 2009. “Optimal Experimental Design for Parameter Estimation of a Cell Signaling Model.” *PLoS Computational Biology* 5 (11): e1000558.

Barkauskas, Christina E., Michael J. Cronce, Craig R. Rackley, Emily J. Bowie, Douglas R. Keene, Barry R. Stripp, Scott H. Randell, Paul W. Noble, and Brigid L. M. Hogan. 2013. “Type 2 Alveolar Cells Are Stem Cells in Adult Lung.” *Journal of Clinical Investigation*. https://doi.org/10.1172/jci68782.

Beauchemin, Kyle Jaymes. 2016. “Characterization of the Murine Lung Transcriptome during Pre- and Post-Natal Development in Three Inbred Strains of Laboratory Mice.” Edited by Carol J. Bult and Viravuth P. Yin. Ann Arbor, United States: The University of Maine.

Becht, Etienne, Leland McInnes, John Healy, Charles-Antoine Dutertre, Immanuel W. H. Kwok, Lai Guan Ng, Florent Ginhoux, and Evan W. Newell. 2018. “Dimensionality Reduction for Visualizing Single-Cell Data Using UMAP.” *Nature Biotechnology*, December. https://doi.org/10.1038/nbt.4314.

Bellusci, S., J. Grindley, H. Emoto, N. Itoh, and B. L. Hogan. 1997. “Fibroblast Growth Factor 10 (FGF10) and Branching Morphogenesis in the Embryonic Mouse Lung.” *Development* 124 (23): 4867–78.

Beltra, Jean-Christophe, Sasikanth Manne, Mohamed S. Abdel-Hakeem, Makoto Kurachi, Josephine R. Giles, Zeyu Chen, Valentina Casella, et al. 2020. “Developmental Relationships of Four Exhausted CD8+ T Cell Subsets Reveals Underlying Transcriptional and Epigenetic Landscape Control Mechanisms.” *Immunity* 52 (5): 825–41.e8.

Bjerkvig, Rolf, Berit B. Tysnes, Karen S. Aboody, Joseph Najbauer, and A. J. A. Terzis. 2005. “Opinion: The Origin of the Cancer Stem Cell: Current Controversies and New Insights.” *Nature Reviews. Cancer* 5 (11): 899–904.

Bonnet, D., and J. E. Dick. 1997. “Human Acute Myeloid Leukemia Is Organized as a Hierarchy That Originates from a Primitive Hematopoietic Cell.” *Nature Medicine* 3 (7): 730–37.

Bose, Thomas, and Steffen Trimper. 2009. “Stochastic Model for Tumor Growth with Immunization.” *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics* 79 (5 Pt 1): 051903.

Boveri, Theodor. 1914. *Zur Frage Der Entstehung Maligner Tumoren*. Fischer.
Bowen, William S., Abhishek K. Svrivastava, Lalit Batra, Hampartsoum Barsoumian, and Haval Shirwan. 2018. “Current Challenges for Cancer Vaccine Adjuvant Development.” *Expert Review of Vaccines* 17 (3): 207–15.
Brabletz, Thomas, Raghu Kalluri, M. Angela Nieto, and Robert A. Weinberg. 2018. “EMT in Cancer.” *Nature Reviews. Cancer* 18 (2): 128–34.
Brinkman, Eva K., Tao Chen, Mario Amendola, and Bas van Steensel. 2014. “Easy Quantitative Assessment of Genome Editing by Sequence Trace Decomposition.” *Nucleic Acids Research* 42 (22): e168.
Britanova, Olga V., Ekaterina V. Putintseva, Mikhail Shugay, Ekaterina M. Merzlyak, Maria A. Turchaninova, Dmitriy B. Staroverov, Dmitriy A. Bolotin, et al. 2014. “Age-Related Decrease in TCR Repertoire Diversity Measured with Deep and Normalized Sequence Profiling.” *Journal of Immunology* 192 (6): 2689–98.
Broca, Paul. 1866. *Traité des tumeurs*. P. Asselin.
Brody, Robert, Yiduo Zhang, Marc Ballas, Mohd Kashif Siddiqui, Palvi Gupta, Craig Barker, Anita Midha, and Jill Walker. 2017. “PD-L1 Expression in Advanced NSCLC: Insights into Risk Stratification and Treatment Selection from a Systematic Literature Review.” *Lung Cancer* 112 (October): 200–215.
Bruchovsky, N., R. Snoek, P. S. Rennie, K. Akakura, L. S. Goldenberg, and M. Gleave. 1996. “Control of Tumor Progression by Maintenance of Apoptosis.” *The Prostate. Supplement* 6: 13–21.
Bruggen, P. van der, C. Traversari, P. Chomez, C. Lurquin, E. De Plaen, B. Van den Eynde, A. Knuth, and T. Boon. 1991. “A Gene Encoding an Antigen Recognized by Cytolytic T Lymphocytes on a Human Melanoma.” *Science* 254 (5038): 1643–47.
Buckley, S. M. K., S. J. Howe, V. Sheard, N. J. Ward, C. Coutelle, A. J. Thrasher, S. N. Waddington, and T. R. McKay. 2008. “Lentiviral Transduction of the Murine Lung Provides Efficient Pseudotype and Developmental Stage-Dependent Cell-Specific Transgene Expression.” *Gene Therapy* 15 (16): 1167–75.
Buenrostro, Jason D., Beijing Wu, Howard Y. Chang, and William J. Greenleaf. 2015. “ATAC-Seq: A Method for Assaying Chromatin Accessibility Genome-Wide.” *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]* 109 (January): 21.29.1–21.29.9.
Butler, Andrew, Paul Hoffman, Peter Smibert, Efthymia Papalexi, and Rahul Satija. 2018. “Integrating Single-Cell Transcriptomic Data across Different Conditions, Technologies, and Species.” *Nature Biotechnology* 36 (5): 411–20.
Cagle, Philip T., Lynette M. Sholl, Neal I. Lindeman, Randa Alsabeh, Dimitrios X. G. Divaris, Philip Foulis, Gemma Lee, et al. 2014. “Template for Reporting Results of Biomarker Testing of Specimens from Patients with Non-Small Cell Carcinoma of the Lung.” *Archives of Pathology & Laboratory Medicine* 138 (2): 171–74.
Cancer Genome Atlas Research Network. 2014. “Comprehensive Molecular Profiling of Lung Adenocarcinoma.” *Nature* 511 (7511): 543–50.
Cannon, Walter B. 1929. “ORGANIZATION FOR PHYSIOLOGICAL HOMEOSTASIS.” *Physiological Reviews* 9 (3): 399–431.
Cao, Jian, and Qin Yan. 2020. “Cancer Epigenetics, Tumor Immunity, and Immunotherapy.” *Trends in Cancer Research* 6 (7): 580–92.
Cao, Junyue, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M. Ibrahim, Andrew

J. Hill, Fan Zhang, et al. 2019. “The Single-Cell Transcriptional Landscape of Mammalian Organogenesis.” *Nature* 566 (7745): 496–502.

Cao, Zhi-Jie, Lin Wei, Shen Lu, De-Chang Yang, and Ge Gao. 2020. “Searching Large-Scale scRNA-Seq Databases via Unbiased Cell Embedding with Cell BLAST.” *Nature Communications* 11 (1): 3458.

Carballo, Gabriela Basile, Jéssica Ribeiro Honorato, Giselle Pinto Farias de Lopes, and Tania Cristina Leite de Sampaio E. Spohr. 2018. “A Highlight on Sonic Hedgehog Pathway.” *Cell Communication and Signaling: CCS* 16 (1): 11.

Carbone, David P., Martin Reck, Luis Paz-Ares, Benjamin Creelan, Leora Horn, Martin Steins, Enriqueta Felip, et al. 2017. “First-Line Nivolumab in Stage IV or Recurrent Non–Small-Cell Lung Cancer.” *The New England Journal of Medicine* 376 (25): 2415–26.

Cardoso, Wellington V., and Jining Lü. 2006. “Regulation of Early Lung Morphogenesis: Questions, Facts and Controversies.” *Development*. https://doi.org/10.1242/dev.02310.

Cardoso, Wellington V., and Jeffrey A. Whitsett. 2008. “Resident Cellular Components of the Lung: Developmental Aspects.” *Proceedings of the American Thoracic Society* 5 (7): 767–71.

Carreno, Beatriz M., Vincent Magrini, Michelle Becker-Hapak, Saghar Kaabinejadian, Jasreet Hundal, Allegra A. Petti, Amy Ly, et al. 2015. “Cancer Immunotherapy. A Dendritic Cell Vaccine Increases the Breadth and Diversity of Melanoma Neoantigen-Specific T Cells.” *Science* 348 (6236): 803–8.

Castellano, Esther, and Eugenio Santos. 2011. “Functional Specificity of Ras Isoforms: So Similar but so Different.” *Genes & Cancer* 2 (3): 216–31.

Cavalli, Giacomo, and Edith Heard. 2019. “Advances in Epigenetics Link Genetics to the Environment and Disease.” *Nature* 571 (7766): 489–99.

Celià-Terrassa, Toni, and Yibin Kang. 2016. “Distinctive Properties of Metastasis-Initiating Cells.” *Genes & Development* 30 (8): 892–908.

Chabner, Bruce A., and Thomas G. Roberts Jr. 2005. “Timeline: Chemotherapy and the War on Cancer.” *Nature Reviews. Cancer* 5 (1): 65–72.

Chapelle, Albert de la. 2003. “Microsatellite Instability.” *The New England Journal of Medicine*.

Cheng, Feixiong, Han Liang, Atul J. Butte, Charis Eng, and Ruth Nussinov. 2019. “Personal Mutanomes Meet Modern Oncology Drug Discovery and Precision Health.” *Pharmacological Reviews* 71 (1): 1–19.

Cheng, Y., and E. W. Newell. 2016. “Deep Profiling Human T Cell Heterogeneity by Mass Cytometry.” *Advances in Immunology*. https://doi.org/10.1016/bs.ai.2016.02.002.

Chen, Lieping, and Dallas B. Flies. 2013. “Molecular Mechanisms of T Cell Co-Stimulation and Co-Inhibition.” *Nature Reviews. Immunology* 13 (4): 227–42.

Chicard, Mathieu, Leo Colmet-Daage, Nathalie Clement, Adrien Danzon, Mylène Bohec, Virginie Bernard, Sylvain Baulande, et al. 2018. “Whole-Exome Sequencing of Cell-Free DNA Reveals Temporo-Spatial Heterogeneity and Identifies Treatment-Resistant Clones in Neuroblastoma.” *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 24 (4): 939–49.

Chotteau-Lelievre, A., P. Dolle, V. Peronne, L. Coutte, Y. de Launoit, and X. Desbiens.

2001. “Expression Patterns of the Ets Transcription Factors from the PEA3 Group during Early Stages of Mouse Development.” *Mechanisms of Development* 108 (1-2): 191–95.

Chotteau-Lelièvre, Anne, Xavier Desbiens, Hélène Pelczar, Pierre-Antoine Defossez, and Yvan de Launoit. 1997. “Differential Expression Patterns of the PEA3 Group Transcription Factors through Murine Embryonic Development.” *Oncogene*. https://doi.org/10.1038/sj.onc.1201261.

Christofori, Gerhard. 2006. “New Signals from the Invasive Front.” *Nature* 441 (7092): 444–50.

Chuang, Jody C., Henning Stehr, Ying Liang, Millie Das, Jane Huang, Maximilian Diehn, Heather A. Wakelee, and Joel W. Neal. 2017. “ERBB2-Mutated Metastatic Non-Small Cell Lung Cancer: Response and Resistance to Targeted Therapies.” *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer* 12 (5): 833–42.

Clemente, C. G., M. C. Mihm Jr, R. Bufalino, S. Zurrida, P. Collini, and N. Cascinelli. 1996. “Prognostic Value of Tumor Infiltrating Lymphocytes in the Vertical Growth Phase of Primary Cutaneous Melanoma.” *Cancer* 77 (7): 1303–10.

Clevers, Hans. 2011. “The Cancer Stem Cell: Premises, Promises and Challenges.” *Nature Medicine* 17 (3): 313–19.

Consortium, The Tabula Muris, The Tabula Muris Consortium, Overall Coordination, Logistical Coordination, Organ collection and processing, Library preparation and sequencing, Computational Data Analysis, et al. 2018. “Single-Cell Transcriptomics of 20 Mouse Organs Creates a Tabula Muris.” *Nature*. https://doi.org/10.1038/s41586-018-0590-4.

Cosi, Irene, Annamaria Pellecchia, Emanuele De Lorenzo, Eugenio Torre, Michela Sica, Gabriella Nesi, Rosario Notaro, and Maria De Angioletti. 2020. “ETV4 Promotes Late Development of Prostatic Intraepithelial Neoplasia and Cell Proliferation through Direct and p53-Mediated Downregulation of p21.” *Journal of Hematology & Oncology* 13 (1): 112.

Coussens, Lisa M., and Zena Werb. 2002. “Inflammation and Cancer.” *Nature* 420 (6917): 860–67.

Cox, Adrienne D., and Channing J. Der. 2010. “Ras History: The Saga Continues.” *Small GTPases* 1 (1): 2–27.

Crispen, Paul L., Rosalia Viterbo, Stephen A. Boorjian, Richard E. Greenberg, David Y. T. Chen, and Robert G. Uzzo. 2009. “Natural History, Growth Kinetics, and Outcomes of Untreated Clinically Localized Renal Tumors under Active Surveillance.” *Cancer* 115 (13): 2844–52.

Croker, A. K., and A. L. Allan. 2008. “Cancer Stem Cells: Implications for the Progression and Treatment of Metastatic Disease.” *Journal of Cellular and Molecular Medicine* 12 (2): 374–90.

Crudele, Julie M., and Jeffrey S. Chamberlain. 2018. “Cas9 Immunity Creates Challenges for CRISPR Gene Editing Therapies.” *Nature Communications* 9 (1): 1–3.

Cruz, R. L. 1991a. “A Calculus for Network Delay. II. Network Analysis.” *IEEE Transactions on Information Theory*. https://doi.org/10.1109/18.61110.

———. 1991b. “A Calculus for Network Delay. I. Network Elements in Isolation.” *IEEE*

Transactions on Information Theory. https://doi.org/10.1109/18.61109.
Curtsinger, J. M., C. S. Schmidt, A. Mondino, D. C. Lins, R. M. Kedl, M. K. Jenkins, and M. F. Mescher. 1999. “Inflammatory Cytokines Provide a Third Signal for Activation of Naive CD4+ and CD8+ T Cells.” *Journal of Immunology* 162 (6): 3256–62.
Curtsinger, J. W., H. H. Fukui, A. A. Khazaeli, A. Kirscher, S. D. Pletcher, D. E. Promislow, and M. Tatar. 1995. “Genetic Variation and Aging.” *Annual Review of Genetics* 29: 553–75.
Cusanovich, Darren A., Riza Daza, Andrew Adey, Hannah A. Pliner, Lena Christiansen, Kevin L. Gunderson, Frank J. Steemers, Cole Trapnell, and Jay Shendure. 2015. “Multiplex Single Cell Profiling of Chromatin Accessibility by Combinatorial Cellular Indexing.” *Science* 348 (6237): 910–14.
Dagogo-Jack, Ibiayi, and Alice T. Shaw. 2018. “Tumour Heterogeneity and Resistance to Cancer Therapies.” *Nature Reviews. Clinical Oncology* 15 (2): 81–94.
Dahlhaus, Rainer, and Liudas Giraitis. 1998. “On the Optimal Segment Length for Parameter Estimates for Locally Stationary Time Series.” *Journal of Time Series Analysis / a Journal Sponsored by the Bernoulli Society for Mathematical Statistics and Probability* 19 (6): 629–55.
Dean L, Kane M. Trastuzumab Therapy and ERBB2 Genotype. 2015 Aug 5 In: Pratt VM, Scott SA, Pirmohamed M, et al., editors. Medical Genetics Summaries [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2012-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK310376/
Decker, William K., Rodrigo F. da Silva, Mayra H. Sanabria, Laura S. Angelo, Fernando Guimarães, Bryan M. Burt, Farrah Kheradmand, and Silke Paust. 2017. “Cancer Immunotherapy: Historical Perspective of a Clinical Revolution and Emerging Preclinical Animal Models.” *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2017.00829.
DeMarino, Catherine, Maria Cowen, Michelle L. Pleet, Daniel O. Pinto, Pooja Khatkar, James Erickson, Steffen S. Docken, et al. 2020. “Differences in Transcriptional Dynamics Between T-Cells and Macrophages as Determined by a Three-State Mathematical Model.” *Scientific Reports* 10 (1): 1–22.
De Petris, L., L. M. Orre, L. Kanter, M. Pernemalm, H. Koyi, R. Lewensohn, and J. Lehtiö. 2009. “Tumor Expression of S100A6 Correlates with Survival of Patients with Stage I Non-Small-Cell Lung Cancer.” *Lung Cancer* 63 (3). https://doi.org/10.1016/j.lungcan.2008.06.003.
Deplancke, Bart, Daniel Alpern, and Vincent Gardeux. 2016. “The Genetics of Transcription Factor DNA Binding Variation.” *Cell* 166 (3): 538–54.
Desai, Tushar J., Douglas G. Brownfield, and Mark A. Krasnow. 2014. “Alveolar Progenitor and Stem Cells in Lung Development, Renewal and Cancer.” *Nature* 507 (7491): 190–94.
Ding, Jiarui, and Aviv Regev. 2021. “Deep Generative Model Embedding of Single-Cell RNA-Seq Profiles on Hyperspheres and Hyperbolic Spaces.” *Nature Communications* 12 (1): 1–17.
Ditlevsen, Susanne, and Adeline Samson. 2013. “Introduction to Stochastic Models in Biology.” *Lecture Notes in Mathematics*. https://doi.org/10.1007/978-3-642-32157-3_1.

"DNA Isolation Buffer." 2019. *Cold Spring Harbor Protocols* 2019 (5): db.rec106443.
Dobosz, Paula, and Tomasz Dzieciątkowski. 2019. "The Intriguing History of Cancer Immunotherapy." *Frontiers in Immunology* 10 (December): 2965.
Doench, John G., Nicolo Fusi, Meagan Sullender, Mudra Hegde, Emma W. Vaimberg, Katherine F. Donovan, Ian Smith, et al. 2016. "Optimized sgRNA Design to Maximize Activity and Minimize off-Target Effects of CRISPR-Cas9." *Nature Biotechnology* 34 (2): 184–91.
Dong, Liang, Richard C. Zieren, Wei Xue, Theo M. de Reijke, and Kenneth J. Pienta. 2019. "Metastatic Prostate Cancer Remains Incurable, Why?" *Asian Journal of Urology* 6 (1): 26–41.
Donoho, David L., and Others. 2000. "High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality." *AMS Math Challenges Lecture* 1 (2000): 32.
Driessens, Gregory, Benjamin Beck, Amélie Caauwe, Benjamin D. Simons, and Cédric Blanpain. 2012. "Defining the Mode of Tumour Growth by Clonal Analysis." *Nature* 488 (7412): 527–30.
Duma, Narjust, Rafael Santana-Davila, and Julian R. Molina. 2019. "Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment." *Mayo Clinic Proceedings. Mayo Clinic* 94 (8): 1623–40.
Dunn, Gavin P., Lloyd J. Old, and Robert D. Schreiber. 2004. "The Immunobiology of Cancer Immunosurveillance and Immunoediting." *Immunity*. https://doi.org/10.1016/j.immuni.2004.07.017.
Dunn, J. M., R. A. Phillips, A. J. Becker, and B. L. Gallie. 1988. "Identification of Germline and Somatic Mutations Affecting the Retinoblastoma Gene." *Science* 241 (4874): 1797–1800.
DuPage, Michel, Ann F. Cheung, Claire Mazumdar, Monte M. Winslow, Roderick Bronson, Leah M. Schmidt, Denise Crowley, Jianzhu Chen, and Tyler Jacks. 2011. "Endogenous T Cell Responses to Antigens Expressed in Lung Adenocarcinomas Delay Malignant Tumor Progression." *Cancer Cell* 19 (1): 72–85.
DuPage, Michel, Alison L. Dooley, and Tyler Jacks. 2009. "Conditional Mouse Lung Cancer Models Using Adenoviral or Lentiviral Delivery of Cre Recombinase." *Nature Protocols* 4 (7): 1064–72.
Eccles, Suzanne A., and Danny R. Welch. 2007. "Metastasis: Recent Discoveries and Novel Treatment Strategies." *The Lancet* 369 (9574): 1742–57.
Eling, Nils, Michael D. Morgan, and John C. Marioni. 2019. "Challenges in Measuring and Understanding Biological Noise." *Nature Reviews. Genetics* 20 (9): 536–48.
Elowitz, Michael B., Arnold J. Levine, Eric D. Siggia, and Peter S. Swain. 2002. "Stochastic Gene Expression in a Single Cell." *Science* 297 (5584): 1183–86.
"Enhancer Control of Transcriptional Bursting." 2016. *Cell* 166 (2): 358–68.
Esteller, Manel. 2008. "Epigenetics in Cancer." *The New England Journal of Medicine* 358 (11): 1148–59.
"ETS-Transcription Factor ETV1 Regulates Stromal Expansion and Metastasis in Pancreatic Cancer." 2016. *Gastroenterology* 151 (3): 540–53.e14.
Evans, M. J., L. J. Cabral, R. J. Stephens, and G. Freeman. 1975. "Transformation of Alveolar Type 2 Cells to Type 1 Cells Following Exposure to NO2." *Experimental and Molecular Pathology* 22 (1): 142–50.
Eva Y. H. P. Lee, William J. Muller. 2010. "Oncogenes and Tumor Suppressor Genes."

Cold Spring Harbor Perspectives in Biology 2 (10). https://doi.org/10.1101/cshperspect.a003236.

"Evolutionary Bioscience as Regulatory Systems Biology." 2011. *Developmental Biology* 357 (1): 35–40.

Fawkner-Corbett, David, Agne Antanaviciute, Kaushal Parikh, Marta Jagielowicz, Ana Sousa Gerós, Tarun Gupta, Neil Ashley, et al. 2021. "Spatiotemporal Analysis of Human Intestinal Development at Single-Cell Resolution." *Cell* 184 (3): 810–26.e23.

Feinberg, A. P., and B. Vogelstein. 1983. "Hypomethylation Distinguishes Genes of Some Human Cancers from Their Normal Counterparts." *Nature* 301 (5895): 89–92.

Fell, Jesse Weldon. 1857. *A Treatise on Cancer: And Its Treatment*. J. Churchill.

Feng, Zhaohui, Wenwei Hu, Elisa de Stanchina, Angelika K. Teresky, Shengkan Jin, Scott Lowe, and Arnold J. Levine. 2007. "The Regulation of AMPK beta1, TSC2, and PTEN Expression by p53: Stress, Cell and Tissue Specificity, and the Role of These Gene Products in Modulating the IGF-1-AKT-mTOR Pathways." *Cancer Research* 67 (7): 3043–53.

Ferris, B. G., Jr, and D. S. Pollard. 1960. "Effect of Deep and Quiet Breathing on Pulmonary Compliance in Man." *The Journal of Clinical Investigation* 39 (January): 143–49.

Filipski, Elisabeth, Verdun M. King, Xiaomei Li, Teresa G. Granda, Marie-Christine Mormont, Xuhui Liu, Bruno Claustrat, Michael H. Hastings, and Francis Lévi. 2002. "Host Circadian Clock as a Control Point in Tumor Progression." *Journal of the National Cancer Institute* 94 (9): 690–97.

Fincher, Christopher T., Omri Wurtzel, Thom de Hoog, Kellie M. Kravarik, and Peter W. Reddien. 2018. "Cell Type Transcriptome Atlas for the Planarian." *Science* 360 (6391). https://doi.org/10.1126/science.aaq1736.

Findlay, Victoria J., Amanda C. LaRue, David P. Turner, Patricia M. Watson, and Dennis K. Watson. 2013. "Understanding the Role of ETS-Mediated Gene Regulation in Complex Biological Processes." *Advances in Cancer Research*. https://doi.org/10.1016/b978-0-12-407190-2.00001-0.

Fitzmaurice, Christina, Daniel Dicker, Amanda Pain, Hannah Hamavid, Maziar Moradi-Lakeh, Michael F. MacIntyre, Christine Allen, et al. 2015. "The Global Burden of Cancer 2013." *JAMA Oncology* 1 (4): 505–27.

Flier, Jeffrey S., Lisa H. Underhill, and Harold F. Dvorak. 1986. "Tumors: Wounds That Do Not Heal." *New England Journal of Medicine*. https://doi.org/10.1056/nejm198612253152606.

Foo, Jasmine, Kevin Leder, and Franziska Michor. 2011. "Stochastic Dynamics of Cancer Initiation." *Physical Biology* 8 (1): 015002.

Foo, Jasmine, Lin L. Liu, Kevin Leder, Markus Riester, Yoh Iwasa, Christoph Lengauer, and Franziska Michor. 2015. "An Evolutionary Approach for Identifying Driver Mutations in Colorectal Cancer." *PLoS Computational Biology* 11 (9): e1004350.

Frank, David B., Ian J. Penkala, Jarod A. Zepp, Aravind Sivakumar, Ricardo Linares-Saldana, William J. Zacharias, Katharine G. Stolz, et al. 2019. "Early Lineage Specification Defines Alveolar Epithelial Ontogeny in the Murine Lung." *Proceedings of the National Academy of Sciences of the United States of America*

116 (10): 4362–71.
Frank, Natasha Y., Tobias Schatton, and Markus H. Frank. 2010. “The Therapeutic Promise of the Cancer Stem Cell Concept.” *The Journal of Clinical Investigation* 120 (1): 41–50.
Fridman, Wolf Herman, Franck Pagès, Catherine Sautès-Fridman, and Jérôme Galon. 2012. “The Immune Contexture in Human Tumours: Impact on Clinical Outcome.” *Nature Reviews. Cancer* 12 (4): 298–306.
Gagniuc, Paul A. 2017. *Markov Chains: From Theory to Implementation and Experimentation*. John Wiley & Sons.
Galon, Jérôme, Anne Costes, Fatima Sanchez-Cabo, Amos Kirilovsky, Bernhard Mlecnik, Christine Lagorce-Pagès, Marie Tosolini, et al. 2006. “Type, Density, and Location of Immune Cells within Human Colorectal Tumors Predict Clinical Outcome.” *Science* 313 (5795): 1960–64.
Garber, Judy E., and Kenneth Offit. 2005. “Hereditary Cancer Predisposition Syndromes.” *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 23 (2): 276–92.
García-Nieto, Pablo E., Ashby J. Morrison, and Hunter B. Fraser. 2019. “The Somatic Mutation Landscape of the Human Body.” *Genome Biology* 20 (1): 298.
“Gene Networks with Transcriptional Bursting Recapitulate Rare Transient Coordinated High Expression States in Cancer.” 2020. *Cell Systems* 10 (4): 363–78.e12.
Ghebranious, N., and L. A. Donehower. 1998. “Mouse Models in Tumor Suppression.” *Oncogene* 17 (25): 3385–3400.
Gibney, E. R., and C. M. Nolan. 2010. “Epigenetics and Gene Expression.” *Heredity* 105 (1): 4–13.
Glanville, Jacob, Huang Huang, Allison Nau, Olivia Hatton, Lisa E. Wagar, Florian Rubelt, Xuhuai Ji, et al. 2017. “Identifying Specificity Groups in the T Cell Receptor Repertoire.” *Nature* 547 (7661): 94–98.
Glass, Leon. 2009. “Introduction to Controversial Topics in Nonlinear Science: Is the Normal Heart Rate Chaotic?” *Chaos* 19 (2): 028501.
Gong, Chang, Jennifer J. Linderman, and Denise Kirschner. 2014. “Harnessing the Heterogeneity of T Cell Differentiation Fate to Fine-Tune Generation of Effector and Memory T Cells.” *Frontiers in Immunology* 5. https://doi.org/10.3389/fimmu.2014.00057.
Goodwin, Sara, John D. McPherson, and W. Richard McCombie. 2016. “Coming of Age: Ten Years of next-Generation Sequencing Technologies.” *Nature Reviews. Genetics* 17 (6): 333–51.
Grånäs, Charlotta, Betina Kerstin Lundholt, Frosty Loechel, Hans-Christian Pedersen, Sara Petersen Bjørn, Viggo Linde, Christian Krogh-Jensen, Eva-Maria Damsgaard Nielsen, Morten Praestegaard, and Søren Jensby Nielsen. 2006. “Identification of RAS-Mitogen-Activated Protein Kinase Signaling Pathway Modulators in an ERF1 Redistribution Screen.” *Journal of Biomolecular Screening* 11 (4): 423–34.
Greaves, Mel. 2010. “Cancer Stem Cells: Back to Darwin?” *Seminars in Cancer Biology* 20 (2): 65–70.
Greaves, Mel, and Carlo C. Maley. 2012. “Clonal Evolution in Cancer.” *Nature* 481 (7381): 306–13.
Gross, L. 1943. “Intradermal Immunization of C3H Mice against a Sarcoma That

Originated in an Animal of the Same Line." https://www.semanticscholar.org/paper/c1ea94ddbaf61772a08a2f494cb014802e66ecbd.

Guo, Mingzhou, Yaojun Peng, Aiai Gao, Chen Du, and James G. Herman. 2019. "Epigenetic Heterogeneity in Cancer." *Biomarker Research* 7 (October): 23.

Guo, Qianyu, Fan Huang, Christophe Goncalves, Sonia V. del Rincón, and Wilson H. Miller. 2019. "Translation of Cancer Immunotherapy from the Bench to the Bedside." *Advances in Cancer Research*. https://doi.org/10.1016/bs.acr.2019.03.001.

Gupta, Gaorav P., and Joan Massagué. 2006. "Cancer Metastasis: Building a Framework." *Cell* 127 (4): 679–95.

Gu, Zuguang, Roland Eils, and Matthias Schlesner. 2016. "Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data." *Bioinformatics* 32 (18): 2847–49.

Haghverdi, Laleh, Aaron T. L. Lun, Michael D. Morgan, and John C. Marioni. 2018. "Batch Effects in Single-Cell RNA-Sequencing Data Are Corrected by Matching Mutual Nearest Neighbors." *Nature Biotechnology* 36 (5): 421–27.

Hajdu, Steven I. 2011. "A Note from History: Landmarks in History of Cancer, Part 2." *Cancer* 117 (12): 2811–20.

———. 2012. "A Note from History: Landmarks in History of Cancer, Part 3." *Cancer* 118 (4): 1155–68.

Hakuma, Nobuyuki, Ichiro Kinoshita, Yasushi Shimizu, Koichi Yamazaki, Koichi Yoshida, Masaharu Nishimura, and Hirotoshi Dosaka-Akita. 2005. "E1AF/PEA3 Activates the Rho/Rho-Associated Kinase Pathway to Increase the Malignancy Potential of Non–Small-Cell Lung Cancer Cells." *Cancer Research*. https://doi.org/10.1158/0008-5472.can-05-0060.

Hamada, Hiromasa, Maria de la Luz Garcia-Hernandez, Joyce B. Reome, Sara K. Misra, Tara M. Strutt, Kai K. McKinstry, Andrea M. Cooper, Susan L. Swain, and Richard W. Dutton. 2009. "Tc17, a Unique Subset of CD8 T Cells That Can Protect against Lethal Influenza Challenge." *Journal of Immunology* 182 (6): 3469–81.

Hamdy, Freddie C., Jenny L. Donovan, J. Athene Lane, Malcolm Mason, Chris Metcalfe, Peter Holding, Michael Davis, et al. 2016. "10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer." *The New England Journal of Medicine* 375 (15): 1415–24.

Hanahan, Douglas. 2014. "Rethinking the War on Cancer." *The Lancet* 383 (9916): 558–63.

Hanahan, Douglas, and Robert A. Weinberg. 2011. "Hallmarks of Cancer: The next Generation." *Cell* 144 (5): 646–74.

Han, Guojun, Matthew H. Spitzer, Sean C. Bendall, Wendy J. Fantl, and Garry P. Nolan. 2018. "Metal-Isotope-Tagged Monoclonal Antibodies for High-Dimensional Mass Cytometry." *Nature Protocols* 13 (10): 2121–48.

Hanson, Aimee L., Hendrik J. Nel, Linda Bradbury, Julie Phipps, Ranjeny Thomas, Kim-Anh Lê Cao, Tony J. Kenna, and Matthew A. Brown. 2020. "T-Cell Receptor Immunosequencing Reveals Altered Repertoire Diversity and Disease-Associated Clonal Expansions in Ankylosing Spondylitis Patients." *Arthritis and Rheumatology* 72 (8): 1289–1302.

Hao, Dapeng, Li Wang, and Li-Jun Di. 2016. “Distinct Mutation Accumulation Rates among Tissues Determine the Variation in Cancer Risk.” *Scientific Reports* 6 (January): 19458.

Hapach, Lauren A., Jenna A. Mosier, Wenjun Wang, and Cynthia A. Reinhart-King. 2019. “Engineered Models to Parse Apart the Metastatic Cascade.” *NPJ Precision Oncology* 3 (August): 20.

Harding, F. A., J. G. McArthur, J. A. Gross, D. H. Raulet, and J. P. Allison. 1992. “CD28-Mediated Signalling Co-Stimulates Murine T Cells and Prevents Induction of Anergy in T-Cell Clones.” *Nature* 356 (6370): 607–9.

Hardoon, David R., Sandor Szedmak, and John Shawe-Taylor. 2004. “Canonical Correlation Analysis: An Overview with Application to Learning Methods.” *Neural Computation* 16 (12): 2639–64.

Harris, Charles R., K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, et al. 2020. “Array Programming with NumPy.” *Nature* 585 (7825): 357–62.

Hashimoto, Masao, Alice O. Kamphorst, Se Jin Im, Haydn T. Kissick, Rathi N. Pillai, Suresh S. Ramalingam, Koichi Araki, and Rafi Ahmed. 2018. “CD8 T Cell Exhaustion in Chronic Infection and Cancer: Opportunities for Interventions.” *Annual Review of Medicine* 69 (January): 301–18.

Hata, Aaron N., Matthew J. Niederst, Hannah L. Archibald, Maria Gomez-Caraballo, Faria M. Siddiqui, Hillary E. Mulvey, Yosef E. Maruvka, et al. 2016. “Tumor Cells Can Follow Distinct Evolutionary Paths to Become Resistant to Epidermal Growth Factor Receptor Inhibition.” *Nature Medicine* 22 (3): 262–69.

Hay, E. D. 1995. “An Overview of Epithelio-Mesenchymal Transformation.” *Acta Anatomica* 154 (1): 8–20.

He, Bing, and Kai Tan. 2016. “Understanding Transcriptional Regulatory Networks Using Computational Models.” *Current Opinion in Genetics & Development* 37 (April): 101–8.

Heide, Timon, Luis Zapata, Marc J. Williams, Benjamin Werner, Giulio Caravagna, Chris P. Barnes, Trevor A. Graham, and Andrea Sottoriva. 2018. “Reply to ‘Neutral Tumor Evolution?’” *Nature Genetics*.

Heppner, G. H., and B. E. Miller. 1983. “Tumor Heterogeneity: Biological Implications and Therapeutic Consequences.” *Cancer Metastasis Reviews* 2 (1): 5–23.

Herriges, John C., Jamie M. Verheyden, Zhen Zhang, Pengfei Sui, Ying Zhang, Matthew J. Anderson, Deborah A. Swing, Yan Zhang, Mark Lewandoski, and Xin Sun. 2015. “FGF-Regulated ETV Transcription Factors Control FGF-SHH Feedback Loop in Lung Branching.” *Developmental Cell*. https://doi.org/10.1016/j.devcel.2015.10.006.

Heuer, Jörn, Franziska Heuer, René Stürmer, Sönke Harder, Hartmut Schlüter, Nayara Braga Emidio, Markus Muttenthaler, Dörthe Jechorek, Frank Meyer, and Werner Hoffmann. 2020. “The Tumor Suppressor TFF1 Occurs in Different Forms and Interacts with Multiple Partners in the Human Gastric Mucus Barrier: Indications for Diverse Protective Functions.” *International Journal of Molecular Sciences* 21 (7). https://doi.org/10.3390/ijms21072508.

Hildanus, Wilhelm Fabricius. n.d. *Opera Observationum et Curationum Medico-Chirurgicarum*.

Hirota, S. 1998. “Gain-of-Function Mutations of c-Kit in Human Gastrointestinal Stromal Tumors.” *Science*. https://doi.org/10.1126/science.279.5350.577.
Hiroumi, Hiromitsu, Hirotoshi Dosaka-Akita, Koichi Yoshida, Masanobu Shindoh, Toshirou Ohbuchi, Kei Fujinaga, and Masaharu Nishimura. 2001. “Expression of E1AF/PEA3, an Ets-Related Transcription Factor in Human Non-Small-Cell Lung Cancers: Its Relevance in Cell Motility and Invasion.” *International Journal of Cancer*. https://doi.org/10.1002/ijc.1410.
Hlubek, Falk, Thomas Brabletz, Jan Budczies, Sabine Pfeiffer, Andreas Jung, and Thomas Kirchner. 2007. “Heterogeneous Expression of Wnt/beta-Catenin Target Genes within Colorectal Cancer.” *International Journal of Cancer. Journal International Du Cancer* 121 (9): 1941–48.
Hoefer, Carl, and Christopher Ray. 1992. “JOHN S. EARMAN [1989]: World Enough and Space-Time: Absolute versus Relational Theories of Space and Time. Cambridge, MA: MIT Press. Xiv 233 Pp. (including Bibliography and index).573580.” *The British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/43.4.573.
Hollenhorst, Peter C., David A. Jones, and Barbara J. Graves. 2004. “Expression Profiles Frame the Promoter Specificity Dilemma of the ETS Family of Transcription Factors.” *Nucleic Acids Research* 32 (18): 5693–5702.
Holliday, R. 1987. “The Inheritance of Epigenetic Defects.” *Science* 238 (4824): 163–70.
Hollingsworth, Robert E., and Kathrin Jansen. 2019. “Turning the Corner on Therapeutic Cancer Vaccines.” *NPJ Vaccines* 4 (February): 7.
Hong, David S., Marwan G. Fakih, John H. Strickler, Jayesh Desai, Gregory A. Durm, Geoffrey I. Shapiro, Gerald S. Falchook, et al. 2020. “KRASG12C Inhibition with Sotorasib in Advanced Solid Tumors.” *The New England Journal of Medicine* 383 (13): 1207–17.
Hoption Cann, S. A., J. P. van Netten, C. van Netten, and D. W. Glover. 2002. “Spontaneous Regression: A Hidden Treasure Buried in Time.” *Medical Hypotheses* 58 (2): 115–19.
Huang, Lei, and Andrew L. Mellor. 2014. “Metabolic Control of Tumour Progression and Antitumour Immunity.” *Current Opinion in Oncology* 26 (1): 92–99.
Huang, Sui. 2012. “Tumor Progression: Chance and Necessity in Darwinian and Lamarckian Somatic (mutationless) Evolution.” *Progress in Biophysics and Molecular Biology* 110 (1): 69–86.
Huang, Xuan, Lei Wu, and Yinsong Ye. 2019. “A Review on Dimensionality Reduction Techniques.” *International Journal of Pattern Recognition and Artificial Intelligence* 33 (10): 1950017.
Humphrey, Linda L., Mark Helfand, Benjamin K. S. Chan, and Steven H. Woolf. 2002. “Breast Cancer Screening: A Summary of the Evidence for the U.S. Preventive Services Task Force.” *Annals of Internal Medicine* 137 (5 Part 1): 347–60.
Hyatt, Brian A., Xiaofei Shangguan, and John M. Shannon. 2004. “FGF-10 Induces SP-C and Bmp4 and Regulates Proximal-Distal Patterning in Embryonic Tracheal Epithelium.” *American Journal of Physiology. Lung Cellular and Molecular Physiology* 287 (6): L1116–26.
Im, Se Jin, Masao Hashimoto, Michael Y. Gerner, Junghwa Lee, Haydn T. Kissick, Matheus C. Burger, Qiang Shan, et al. 2016. “Defining CD8+ T Cells That Provide

the Proliferative Burst after PD-1 Therapy.” *Nature* 537 (7620): 417–21.
Ingenito, E. P., L. Mark, J. Morris, F. F. Espinosa, R. D. Kamm, and M. Johnson. 1999. “Biophysical Characterization and Modeling of Lung Surfactant Components.” *Journal of Applied Physiology* 86 (5): 1702–14.
Institute, National Cancer, and National Cancer Institute. 2020. “Immune System and Related Disorders.” *Definitions*. https://doi.org/10.32388/ivf2w6.
Intlekofer, Andrew M., Arnob Banerjee, Naofumi Takemoto, Scott M. Gordon, Caitlin S. Dejong, Haina Shin, Christopher A. Hunter, E. John Wherry, Tullia Lindsten, and Steven L. Reiner. 2008. “Anomalous Type 17 Response to Viral Infection by CD8+ T Cells Lacking T-Bet and Eomesodermin.” *Science* 321 (5887): 408–11.
Irish, Jonathan M., Randi Hovland, Peter O. Krutzik, Omar D. Perez, Øystein Bruserud, Bjørn T. Gjertsen, and Garry P. Nolan. 2004. “Single Cell Profiling of Potentiated Phospho-Protein Networks in Cancer Cells.” *Cell* 118 (2): 217–28.
Ishikawa, Fumihiko, Shuro Yoshida, Yoriko Saito, Atsushi Hijikata, Hiroshi Kitamura, Satoshi Tanaka, Ryu Nakamura, et al. 2007. “Chemotherapy-Resistant Human AML Stem Cells Home to and Engraft within the Bone-Marrow Endosteal Region.” *Nature Biotechnology* 25 (11): 1315–21.
Jackson, E. L., N. Willis, K. Mercer, R. T. Bronson, D. Crowley, R. Montoya, T. Jacks, and D. A. Tuveson. 2001. “Analysis of Lung Tumor Initiation and Progression Using Conditional Expression of Oncogenic K-Ras.” *Genes & Development* 15 (24): 3243–48.
Jackson, Erica L., Kenneth P. Olive, David A. Tuveson, Roderick Bronson, Denise Crowley, Michael Brown, and Tyler Jacks. 2005. “The Differential Effects of Mutant p53 Alleles on Advanced Murine Lung Cancer.” *Cancer Research*. https://doi.org/10.1158/0008-5472.can-05-2193.
Jain, Amit, Cindy Lim, Eugene Mingjin Gan, David Zhihao Ng, Quan Sing Ng, Mei Kim Ang, Angela Takano, et al. 2015. “Impact of Smoking and Brain Metastasis on Outcomes of Advanced EGFR Mutation Lung Adenocarcinoma Patients Treated with First Line Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors.” *PloS One* 10 (5): e0123587.
Janiszewska, Michalina. 2020. “The Microcosmos of Intratumor Heterogeneity: The Space-Time of Cancer Evolution.” *Oncogene* 39 (10): 2031–39.
Janiszewska, Michalina, Lin Liu, Vanessa Almendro, Yanan Kuang, Cloud Paweletz, Rita A. Sakr, Britta Weigelt, et al. 2015. “In Situ Single-Cell Analysis Identifies Heterogeneity for PIK3CA Mutation and HER2 Amplification in HER2-Positive Breast Cancer.” *Nature Genetics* 47 (10): 1212–19.
Jhunjhunwala, Suchit, Christian Hammer, and Lélia Delamarre. 2021. “Antigen Presentation in Cancer: Insights into Tumour Immunogenicity and Immune Evasion.” *Nature Reviews. Cancer* 21 (5): 298–312.
Jin, Ming-Zhu, and Wei-Lin Jin. 2020. “The Updated Landscape of Tumor Microenvironment and Drug Repurposing.” *Signal Transduction and Targeted Therapy* 5 (1): 166.
Jin, Suoqin, Adam L. MacLean, Tao Peng, and Qing Nie. 2018. “scEpath: Energy Landscape-Based Inference of Transition Probabilities and Cellular Trajectories from Single-Cell Transcriptomic Data.” *Bioinformatics* 34 (12): 2077–86.
Jin, Xiang, Yinghui Guan, Hui Sheng, and Yang Liu. 2017. “Crosstalk in Competing

Endogenous RNA Network Reveals the Complex Molecular Mechanism Underlying Lung Cancer." *Oncotarget* 8 (53): 91270.

Jögi, Annika, Marica Vaapil, Martin Johansson, and Sven Påhlman. 2012. "Cancer Cell Differentiation Heterogeneity and Aggressive Behavior in Solid Tumors." *Upsala Journal of Medical Sciences* 117 (2): 217–24.

Johnson, L., K. Mercer, D. Greenbaum, R. T. Bronson, D. Crowley, D. A. Tuveson, and T. Jacks. 2001. "Somatic Activation of the K-Ras Oncogene Causes Early Onset Lung Cancer in Mice." *Nature* 410 (6832): 1111–16.

Jones, Peter A., and Stephen B. Baylin. 2007. "The Epigenomics of Cancer." *Cell* 128 (4): 683–92.

June, Carl H. 2007. "Adoptive T Cell Therapy for Cancer in the Clinic." *The Journal of Clinical Investigation* 117 (6): 1466–76.

Kaech, Susan M., and Weiguo Cui. 2012. "Transcriptional Control of Effector and Memory CD8 T Cell Differentiation." *Nature Reviews Immunology*. https://doi.org/10.1038/nri3307.

Kaern, Mads, Timothy C. Elston, William J. Blake, and James J. Collins. 2005. "Stochasticity in Gene Expression: From Theories to Phenotypes." *Nature Reviews. Genetics* 6 (6): 451–64.

Kaplan, Rosandra N., Rebecca D. Riba, Stergios Zacharoulis, Anna H. Bramley, Loïc Vincent, Carla Costa, Daniel D. MacDonald, et al. 2005. "VEGFR1-Positive Haematopoietic Bone Marrow Progenitors Initiate the Pre-Metastatic Niche." *Nature* 438 (7069): 820–27.

Katok, Anatole, and Boris Hasselblatt. 1995. "Introduction to the Modern Theory of Dynamical Systems." https://doi.org/10.1017/cbo9780511809187.

Kauffmann-Guerrero, Diego, Kathrin Kahnert, Rosemarie Kiefl, Laura Sellmer, Julia Walter, Jürgen Behr, and Amanda Tufman. 2021. "Systemic Inflammation and pro-Inflammatory Cytokine Profile Predict Response to Checkpoint Inhibitor Treatment in NSCLC: A Prospective Study." *Scientific Reports* 11 (1): 10919.

Kharchenko, Peter V., Lev Silberstein, and David T. Scadden. 2014. "Bayesian Approach to Single-Cell Differential Expression Analysis." *Nature Methods* 11 (7): 740–42.

Kim, Carla F. Bender, Erica L. Jackson, Amber E. Woolfenden, Sharon Lawrence, Imran Babar, Sinae Vogel, Denise Crowley, Roderick T. Bronson, and Tyler Jacks. 2005. "Identification of Bronchioalveolar Stem Cells in Normal Lung and Lung Cancer." *Cell* 121 (6): 823–35.

Kim, Jun-Kyum, Hee-Young Jeon, and Hyunggee Kim. 2015. "The Molecular Mechanisms Underlying the Therapeutic Resistance of Cancer Stem Cells." *Archives of Pharmacal Research* 38 (3): 389–401.

Kiselev, Vladimir Yu, Tallulah S. Andrews, and Martin Hemberg. 2019. "Challenges in Unsupervised Clustering of Single-Cell RNA-Seq Data." *Nature Reviews. Genetics* 20 (5): 273–82.

Klein, G., and E. Klein. 1985. "Evolution of Tumours and the Impact of Molecular Oncology." *Nature* 315 (6016): 190–95.

"Know Thy Mouse: Variability in Aged Mice." n.d. Accessed July 13, 2021. https://www.jax.org/news-and-insights/jax-blog/2020/march/variability-in-aged-mice.

Knudsen, Amy B., Ann G. Zauber, Carolyn M. Rutter, Steffie K. Naber, V. Paul

Doria-Rose, Chester Pabiniak, Colden Johanson, Sara E. Fischer, Iris Lansdorp-Vogelaar, and Karen M. Kuntz. 2016. “Estimation of Benefits, Burden, and Harms of Colorectal Cancer Screening Strategies: Modeling Study for the US Preventive Services Task Force.” *JAMA: The Journal of the American Medical Association* 315 (23): 2595–2609.

Knudson, A. G., Jr. 1971. “Mutation and Cancer: Statistical Study of Retinoblastoma.” *Proceedings of the National Academy of Sciences of the United States of America* 68 (4): 820–23.

Kong, Qiongman, Michael P. Stockinger, Yueming Chang, Hirofumi Tashiro, and Chien-Liang Glenn Lin. 2008. “The Presence of rRNA Sequences in Polyadenylated RNA and Its Potential Functions.” *Biotechnology Journal* 3 (8): 1041–46.

Koning, Harry J. de, Rafael Meza, Sylvia K. Plevritis, Kevin ten Haaf, Vidit N. Munshi, Jihyoun Jeon, Saadet Ayca Erdogan, et al. 2014. “Benefits and Harms of Computed Tomography Lung Cancer Screening Strategies: A Comparative Modeling Study for the U.S. Preventive Services Task Force.” *Annals of Internal Medicine* 160 (5): 311–20.

Kouzarides, Tony. 2007. “Chromatin Modifications and Their Function.” *Cell* 128 (4): 693–705.

Kreso, Antonija, and John E. Dick. 2014. “Evolution of the Cancer Stem Cell Model.” *Cell Stem Cell* 14 (3): 275–91.

Krog, Benjamin L., and Michael D. Henry. 2018. “Biomechanics of the Circulating Tumor Cell Microenvironment.” In *Biomechanics in Oncology*, edited by Cheng Dong, Nastaran Zahir, and Konstantinos Konstantopoulos, 209–33. Cham: Springer International Publishing.

Krzak, Monika, Yordan Raykov, Alexis Boukouvalas, Luisa Cutillo, and Claudia Angelini. 2019. “Benchmark and Parameter Sensitivity Analysis of Single-Cell RNA Sequencing Clustering Methods.” *Frontiers in Genetics* 10. https://doi.org/10.3389/fgene.2019.01253.

Kugler, Matthias C., Alexandra L. Joyner, Cynthia A. Loomis, and John S. Munger. 2015. “Sonic Hedgehog Signaling in the Lung. From Development to Disease.” *American Journal of Respiratory Cell and Molecular Biology* 52 (1): 1–13.

Kumagai, Shogo, Yosuke Togashi, Takahiro Kamada, Eri Sugiyama, Hitomi Nishinakamura, Yoshiko Takeuchi, Kochin Vitaly, et al. 2020. “The PD-1 Expression Balance between Effector and Regulatory T Cells Predicts the Clinical Efficacy of PD-1 Blockade Therapies.” *Nature Immunology* 21 (11): 1346–58.

Kurtulus, Sema, Asaf Madi, Giulia Escobar, Max Klapholz, Jackson Nyman, Elena Christian, Mathias Pawlak, et al. 2019. “Checkpoint Blockade Immunotherapy Induces Dynamic Changes in PD-1CD8 Tumor-Infiltrating T Cells.” *Immunity* 50 (1): 181–94.e6.

Kwak, Eunice L., Yung-Jue Bang, D. Ross Camidge, Alice T. Shaw, Benjamin Solomon, Robert G. Maki, Sai-Hong I. Ou, et al. 2010. “Anaplastic Lymphoma Kinase Inhibition in Non-Small-Cell Lung Cancer.” *The New England Journal of Medicine* 363 (18): 1693–1703.

Kwei, K. A., Y. H. Kim, L. Girard, J. Kao, M. Pacyna-Gengelbach, K. Salari, J. Lee, et al. 2008. “Genomic Profiling Identifies TITF1 as a Lineage-Specific Oncogene

Amplified in Lung Cancer." *Oncogene* 27 (25): 3635–40.
Kyttälä, Aija, Roksana Moraghebi, Cristina Valensisi, Johannes Kettunen, Colin Andrus, Kalyan Kumar Pasumarthy, Mahito Nakanishi, et al. 2016. "Genetic Variability Overrides the Impact of Parental Cell Type and Determines iPSC Differentiation Potential." *Stem Cell Reports* 6 (2): 200–212.
LaFave, Lindsay M., Vinay K. Kartha, Sai Ma, Kevin Meli, Isabella Del Priore, Caleb Lareau, Santiago Naranjo, et al. 2020. "Epigenomic State Transitions Characterize Tumor Progression in Mouse Lung Adenocarcinoma." *Cancer Cell* 38 (2): 212–28.e13.
Laken, S. J., G. M. Petersen, S. B. Gruber, C. Oddoux, H. Ostrer, F. M. Giardiello, S. R. Hamilton, et al. 1997. "Familial Colorectal Cancer in Ashkenazim due to a Hypermutable Tract in APC." *Nature Genetics* 17 (1): 79–83.
Lancia, Andrea, Elisa Merizzoli, and Andrea Riccardo Filippi. 2019. "The 8th UICC/AJCC TNM Edition for Non-Small Cell Lung Cancer Staging: Getting off to a Flying Start?" *Annals of Translational Medicine*.
Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. 2009. "Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome." *Genome Biology* 10 (3): R25.
Laslo, Peter, Chauncey J. Spooner, Aryeh Warmflash, David W. Lancki, Hyun-Jun Lee, Roger Sciammas, Benjamin N. Gantner, Aaron R. Dinner, and Harinder Singh. 2006. "Multilineage Transcriptional Priming and Determination of Alternate Hematopoietic Cell Fates." *Cell* 126 (4): 755–66.
Laviano, Alessandro, and Filippo Rossi Fanelli. 2012. "Toxicity in Chemotherapy — When Less Is More." *New England Journal of Medicine*. https://doi.org/10.1056/nejmcibr1202395.
Lawson, Devon A., Kai Kessenbrock, Ryan T. Davis, Nicholas Pervolarakis, and Zena Werb. 2018. "Tumour Heterogeneity and Metastasis at Single-Cell Resolution." *Nature Cell Biology*. https://doi.org/10.1038/s41556-018-0236-7.
Ledford, Heidi, Holly Else, and Matthew Warren. 2018. "Cancer Immunologists Scoop Medicine Nobel Prize." *Nature*. https://doi.org/10.1038/d41586-018-06751-0.
Lee, Catherine S., Newman J. Sund, Rüdiger Behr, Pedro L. Herrera, and Klaus H. Kaestner. 2005. "Foxa2 Is Required for the Differentiation of Pancreatic Alpha-Cells." *Developmental Biology* 278 (2): 484–95.
Lee, H., A. H. Trainer, L. S. Friedman, F. C. Thistlethwaite, M. J. Evans, B. A. Ponder, and A. R. Venkitaraman. 1999. "Mitotic Checkpoint Inactivation Fosters Transformation in Cells Lacking the Breast Cancer Susceptibility Gene, Brca2." *Molecular Cell* 4 (1): 1–10.
León, D. Aguilar, D. Aguilar León, M. J. Zumárraga, R. Jiménez Oropeza, A. K. Gioffré, A. Bernardelli, H. Orozco Estévez, A. A. Cataldi, and R. Hernández Pando. 2009. "Mycobacterium Boviswith Different Genotypes and from Different Hosts Induce Dissimilar Immunopathological Lesions in a Mouse Model of Tuberculosis." *Clinical & Experimental Immunology*. https://doi.org/10.1111/j.1365-2249.2009.03923.x.
Leun, Anne M. van der, Daniela S. Thommen, and Ton N. Schumacher. 2020. "CD8+ T Cell States in Human Cancer: Insights from Single-Cell Analysis." *Nature Reviews. Cancer* 20 (4): 218–32.
Levine, Arnold J. 1997. "p53, the Cellular Gatekeeper for Growth and Division." *Cell*.

https://doi.org/10.1016/s0092-8674(00)81871-1.
Levine, Jacob H., Erin F. Simonds, Sean C. Bendall, Kara L. Davis, El-Ad D. Amir, Michelle Tadmor, Oren Litvin, et al. 2015. "Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells That Correlate with Prognosis." *Cell* 162 (1): 184.
Li, Amy, and Ph. D. Massachusetts Institute of Technology. 2018. "Investigation of CD4+ T Cell Heterogeneity and Function in a Genetic Mouse Model of Lung Adenocarcinoma." Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/117885?show=full.
Li, Bo, and Colin N. Dewey. 2011. "RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome." *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-12-323.
Li, Carman Man-Chung, Vasilena Gocheva, Madeleine J. Oudin, Arjun Bhutkar, Shi Yun Wang, Saya R. Date, Sheng Rong Ng, et al. 2015. "Foxa2 and Cdx2 Cooperate with Nkx2-1 to Inhibit Lung Adenocarcinoma Metastasis." *Genes & Development* 29 (17): 1850–62.
Li, Mengyuan, Zhilan Zhang, Lin Li, and Xiaosheng Wang. 2020. "An Algorithm to Quantify Intratumor Heterogeneity Based on Alterations of Gene Expression Profiles." *Communications Biology* 3 (1): 505.
Lindemann, Ralph K., Pia Ballschmieter, Alfred Nordheim, and Jürgen Dittmer. 2001. "Transforming Growth Factor β Regulates Parathyroid Hormone-Related Protein Expression in MDA-MB-231 Breast Cancer Cells through a Novel Smad/Ets Synergism*." *The Journal of Biological Chemistry* 276 (49): 46661–70.
Linehan, Jonathan L., Oliver J. Harrison, Seong-Ji Han, Allyson L. Byrd, Ivan Vujkovic-Cvijin, Alejandro V. Villarino, Shurjo K. Sen, et al. 2018. "Non-Classical Immunity Controls Microbiota Impact on Skin Immunity and Tissue Repair." *Cell*. https://doi.org/10.1016/j.cell.2017.12.033.
Lin, S. C., S. X. Skapek, and E. Y. Lee. 1996. "Genes in the RB Pathway and Their Knockout in Mice." *Seminars in Cancer Biology* 7 (5): 279–89.
Li, Peng, Xiaodong Lv, Zhiqiang Zhang, and Shanshan Xie. 2019. "S100A6/miR193a Regulates the Proliferation, Invasion, Migration and Angiogenesis of Lung Cancer Cells through the P53 Acetylation." *American Journal of Translational Research* 11 (8): 4634.
Lipinski, M. M., and T. Jacks. 1999. "The Retinoblastoma Gene Family in Differentiation and Development." *Oncogene* 18 (55): 7873–82.
Liston, Adrian, and Daniel H. D. Gray. 2014. "Homeostatic Control of Regulatory T Cell Diversity." *Nature Reviews. Immunology* 14 (3): 154–65.
Little, Danielle R., Kamryn N. Gerner-Mauro, Per Flodby, Edward D. Crandall, Zea Borok, Haruhiko Akiyama, Shioko Kimura, Edwin J. Ostrin, and Jichao Chen. 2019. "Transcriptional Control of Lung Alveolar Type 1 Cell Development and Maintenance by NK Homeobox 2-1." *Proceedings of the National Academy of Sciences of the United States of America* 116 (41): 20545–55.
Little, Danielle R., Anne M. Lynch, Yun Yan, Haruhiko Akiyama, Shioko Kimura, and Jichao Chen. 2021. "Differential Chromatin Binding of the Lung Lineage Transcription Factor NKX2-1 Resolves Opposing Murine Alveolar Cell Fates in Vivo." *Nature Communications* 12 (1): 1–18.

Liu, Edison T. 2008. “Functional Genomics of Cancer.” *Current Opinion in Genetics & Development*. https://doi.org/10.1016/j.gde.2008.07.014.
Liu, Jialin, Chao Gao, Joshua Sodicoff, Velina Kozareva, Evan Z. Macosko, and Joshua D. Welch. 2020. “Jointly Defining Cell Types from Multiple Single-Cell Datasets Using LIGER.” *Nature Protocols* 15 (11): 3632–62.
Liu, Jiewei, Xiaorong Zhong, Juan Li, Baoxing Liu, Shanxian Guo, Jun Chen, Qingwei Tan, et al. 2012. “Screening and Identification of Lung Cancer Metastasis-Related Genes by Suppression Subtractive Hybridization.” *Thoracic Cancer* 3 (3): 207–16.
Liu, Lihui, Shang-Wei Chong, Nataraja V. Balasubramaniyan, Vladimir Korzh, and Ruowen Ge. 2002. “Platelet-Derived Growth Factor Receptor Alpha (pdgfr-Alpha) Gene in Zebrafish Embryonic Development.” *Mechanisms of Development* 116 (1-2): 227–30.
Liu, Qiaofei, Quan Liao, and Yupei Zhao. 2016. “Myeloid-Derived Suppressor Cells (MDSC) Facilitate Distant Metastasis of Malignancies by Shielding Circulating Tumor Cells (CTC) from Immune Surveillance.” *Medical Hypotheses* 87 (February): 34–39.
Liu, Ting, Fuli Shi, Ying Ying, Qiongfeng Chen, Zhimin Tang, and Hui Lin. 2020. “Mouse Model of Menstruation: An Indispensable Tool to Investigate the Mechanisms of Menstruation and Gynaecological Diseases.” *Molecular Medicine Reports* 22 (6): 4463.
Liu, Yuru, Haiyan Jiang, Howard C. Crawford, and Brigid L. M. Hogan. 2003. “Role for ETS Domain Transcription Factors Pea3/Erm in Mouse Lung Development.” *Developmental Biology* 261 (1): 10–24.
Li, Xingchen, and Jianliu Wang. 2020. “Mechanical Tumor Microenvironment and Transduction: Cytoskeleton Mediates Cancer Cell Invasion and Metastasis.” *International Journal of Biological Sciences* 16 (12): 2014–28.
Lo, Hin Ching, Zhan Xu, Ik Sun Kim, Aaron Muscarella, Jun Liu, Sarah Hein, Hai Wang, et al. 2020. “Circulating Tumor Cell Clusters Exhibit Enhanced Immune Evasion from Natural Killer Cells.” *The Journal of Immunology* 204 (1 Supplement): 88.18–88.18.
Lönnberg, Tapio, Valentine Svensson, Kylie R. James, Daniel Fernandez-Ruiz, Ismail Sebina, Ruddy Montandon, Megan S. F. Soon, et al. 2017. “Single-Cell RNA-Seq and Computational Analysis Using Temporal Mixture Modelling Resolves Th1/Tfh Fate Bifurcation in Malaria.” *Science Immunology* 2 (9). https://doi.org/10.1126/sciimmunol.aal2192.
Lorusso, Girieca, and Curzio Rüegg. 2008. “The Tumor Microenvironment and Its Contribution to Tumor Evolution toward Metastasis.” *Histochemistry and Cell Biology* 130 (6): 1091–1103.
Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550.
Luecken, Malte D., and Fabian J. Theis. 2019. “Current Best Practices in Single-Cell RNA-Seq Analysis: A Tutorial.” *Molecular Systems Biology* 15 (6): e8746.
Luetteke, N. C., T. H. Qiu, S. E. Fenton, K. L. Troyer, R. F. Riedel, A. Chang, and D. C. Lee. 1999. “Targeted Inactivation of the EGF and Amphiregulin Genes Reveals Distinct Roles for EGF Receptor Ligands in Mouse Mammary Gland Development.”

Development 126 (12): 2739–50.

Lu, M. M., H. Yang, L. Zhang, W. Shu, D. G. Blair, and E. E. Morrisey. 2001. “The Bone Morphogenic Protein Antagonist Gremlin Regulates Proximal-Distal Patterning of the Lung.” *Developmental Dynamics: An Official Publication of the American Association of Anatomists* 222 (4): 667–80.

Lun, A. 2019. “Further MNN Algorithm Development.”

Luo, Kunxin. 2017. “Signaling Cross Talk between TGF-β/Smad and Other Signaling Pathways.” *Cold Spring Harbor Perspectives in Biology* 9 (1): a022137.

Luo, Susan Y., and David Cl Lam. 2013. “Oncogenic Driver Mutations in Lung Cancer.” *Translational Respiratory Medicine* 1 (1): 6.

Luo, Wenxin, Panwen Tian, Yue Wang, Heng Xu, Lu Chen, Chao Tang, Yang Shu, et al. 2018. “Characteristics of Genomic Alterations of Lung Adenocarcinoma in Young Never-Smokers.” *International Journal of Cancer. Journal International Du Cancer* 143 (7): 1696–1705.

Lusitani, Z. n.d. “Praxis Medical Admiranda.” *Lugduni: J. Hugvetan* 1649.

Lynch, Henry T., Carrie L. Snyder, Trudy G. Shaw, Christopher D. Heinen, and Megan P. Hitchins. 2015. “Milestones of Lynch Syndrome: 1895-2015.” *Nature Reviews. Cancer* 15 (3): 181–94.

Maaten, Laurens van der. 2008. “Visualizing Data Using T-SNE.” 2008. https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf?fbclid=IwA.

Macosko, Evan Z., Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, et al. 2015. “Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets.” *Cell* 161 (5): 1202–14.

Maeda, Yutaka, Vrushank Davé, and Jeffrey A. Whitsett. 2007. “Transcriptional Control of Lung Morphogenesis.” *Physiological Reviews* 87 (1): 219–44.

Mailleux, A. A., D. Tefft, D. Ndiaye, N. Itoh, J. P. Thiery, D. Warburton, and S. Bellusci. 2001. “Evidence That SPROUTY2 Functions as an Inhibitor of Mouse Embryonic Lung Growth and Morphogenesis.” *Mechanisms of Development* 102 (1-2): 81–94.

Mainardi, Sara. 2013. “Identification and Characterization of Cancer Initiating Cells in K-Ras Driven Non Small Cell Lung Cancer.” repositorio.uam.es. https://repositorio.uam.es/handle/10486/660249.

Mandl, Judith N., Rachel Liou, Frederick Klauschen, Nienke Vrisekoop, João P. Monteiro, Andrew J. Yates, Alex Y. Huang, and Ronald N. Germain. 2012. “Quantification of Lymph Node Transit Times Reveals Differences in Antigen Surveillance Strategies of Naive CD4+ and CD8+ T Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (44): 18036–41.

Mani, Sendurai A., Wenjun Guo, Mai-Jing Liao, Elinor Ng Eaton, Ayyakkannu Ayyanan, Alicia Y. Zhou, Mary Brooks, et al. 2008. “The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells.” *Cell* 133 (4): 704–15.

Marjanovic, Nemanja Despot, Matan Hofree, Jason E. Chan, David Canner, Katherine Wu, Marianna Trakala, Griffin G. Hartmann, et al. 2020. “Emergence of a High-Plasticity Cell State during Lung Cancer Evolution.” *Cancer Cell* 38 (2): 229–46.e13.

Marjanovic, Nemanja D., Robert A. Weinberg, and Christine L. Chaffer. 2013. “Cell

Plasticity and Heterogeneity in Cancer." *Clinical Chemistry*. https://doi.org/10.1373/clinchem.2012.184655.
Marra, Amanda N., and Rebecca A. Wingert. 2016. "Epithelial Cell Fate in the Nephron Tubule Is Mediated by the ETS Transcription Factors etv5a and etv4 during Zebrafish Kidney Development." *Developmental Biology* 411 (2): 231–45.
Martincorena, Iñigo, and Peter J. Campbell. 2015. "Somatic Mutation in Cancer and Normal Cells." *Science* 349 (6255): 1483–89.
Martin, D. 2002. *Manifold Theory: An Introduction for Mathematical Physicists*. Elsevier.
Martínez-Jiménez, Francisco, Ferran Muiños, Inés Sentís, Jordi Deu-Pons, Iker Reyes-Salazar, Claudia Arnedo-Pac, Loris Mularoni, et al. 2020. "A Compendium of Mutational Cancer Driver Genes." *Nature Reviews. Cancer* 20 (10): 555–72.
Martins, Renato G., Thomas A. D'Amico, Billy W. Loo, Mary Pinder-Schenck, Hossein Borghaei, Jamie E. Chaft, Apar Kishor P. Ganti, et al. 2012. "The Management of Patients With Stage IIIA Non–Small Cell Lung Cancer With N2 Mediastinal Node Involvement." *Journal of the National Comprehensive Cancer Network: JNCCN* 10 (5): 599–613.
Ma, Shining, Tao Jiang, and Rui Jiang. 2018. "Constructing tissue specific Transcriptional Regulatory Networks via a Markov Random Field." *BMC Genomics* 19 (Suppl 10): 884.
Ma, Shuangge, and Ying Dai. 2011. "Principal Component Analysis Based Methods in Bioinformatics Studies." *Briefings in Bioinformatics* 12 (6): 714–22.
Mason, Robert J. 2006. "Biology of Alveolar Type II Cells." *Respirology* 11 Suppl (January): S12–15.
Massagué, Joan, and Anna C. Obenauf. 2016. "Metastatic Colonization by Circulating Tumour Cells." *Nature* 529 (7586): 298–306.
McCarthy, Edward F. 2006. "The Toxins of William B. Coley and the Treatment of Bone and Soft-Tissue Sarcomas." *The Iowa Orthopaedic Journal* 26: 154–58.
McFarland, Christopher D., Julia A. Yaglom, Jonathan W. Wojtkowiak, Jacob G. Scott, David L. Morse, Michael Y. Sherman, and Leonid A. Mirny. 2017. "The Damaging Effect of Passenger Mutations on Cancer Progression." *Cancer Research* 77 (18): 4763–72.
McGrail, D. J., P. G. Pilié, N. U. Rashid, L. Voorwerk, M. Slagter, M. Kok, E. Jonasch, et al. 2021. "High Tumor Mutation Burden Fails to Predict Immune Checkpoint Blockade Response across All Cancer Types." *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO* 32 (5): 661–72.
McGranahan, Nicholas, and Charles Swanton. 2019. "Neoantigen Quality, Not Quantity." *Science Translational Medicine* 11 (506). https://doi.org/10.1126/scitranslmed.aax7918.
McInnes, Leland, John Healy, and James Melville. 2018. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." *arXiv [stat.ML]*. arXiv. http://arxiv.org/abs/1802.03426.
McInnes, Leland, John Healy, Nathaniel Saul, and Lukas Großberger. 2018. "UMAP: Uniform Manifold Approximation and Projection." *Journal of Open Source Software*. https://doi.org/10.21105/joss.00861.
McKinney, Wes. 2017. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. "O'Reilly Media, Inc."

McMahan, Rachel H., Jennifer A. McWilliams, Kimberly R. Jordan, Steven W. Dow, Darcy B. Wilson, and Jill E. Slansky. 2006. “Relating TCR-Peptide-MHC Affinity to Immunogenicity for the Design of Tumor Vaccines.” *The Journal of Clinical Investigation* 116 (9): 2543–51.

McVeigh, Terri Patricia, Lauren M. Hughes, Nicola Miller, Margaret Sheehan, Maccon Keane, Karl J. Sweeney, and Michael J. Kerin. 2014. “The Impact of Oncotype DX Testing on Breast Cancer Management and Chemotherapy Prescribing Patterns in a Tertiary Referral Centre.” *European Journal of Cancer* 50 (16): 2763–70.

Meier, Joshua A., Feng Zhang, and Neville E. Sanjana. 2017. “GUIDES: sgRNA Design for Loss-of-Function Screens.” *Nature Methods* 14 (9): 831–32.

Meo, Pasquale De, Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, and Alessandro Provetti. 2011. “Generalized Louvain Method for Community Detection in Large Networks.” *2011 11th International Conference on Intelligent Systems Design and Applications*. https://doi.org/10.1109/isda.2011.6121636.

Metzger, Ross J., Ophir D. Klein, Gail R. Martin, and Mark A. Krasnow. 2008. “The Branching Programme of Mouse Lung Development.” *Nature* 453 (7196): 745–50.

Miller, Brian C., Debattama R. Sen, Rose Al Abosy, Kevin Bi, Yamini V. Virkud, Martin W. LaFleur, Kathleen B. Yates, et al. 2019a. “Subsets of Exhausted CD8+ T Cells Differentially Mediate Tumor Control and Respond to Checkpoint Blockade.” *Nature Immunology* 20 (3): 326–36.

———. 2019b. “Author Correction: Subsets of Exhausted CD8 T Cells Differentially Mediate Tumor Control and Respond to Checkpoint Blockade.” *Nature Immunology* 20 (11): 1556.

Miller, Leigh-Anne D., Susan E. Wert, Jean C. Clark, Yan Xu, Anne-Karina T. Perl, and Jeffrey A. Whitsett. 2004. “Role of Sonic Hedgehog in Patterning of Tracheal-Bronchial Cartilage and the Peripheral Lung.” *Developmental Dynamics: An Official Publication of the American Association of Anatomists* 231 (1): 57–71.

Minoo, Parviz, Guoshan Su, Hong Drum, Pablo Bringas, and Shioko Kimura. 1999. “Defects in Tracheoesophageal and Lung Morphogenesis inNkx2.1(−/−) Mouse Embryos.” *Developmental Biology*. https://doi.org/10.1006/dbio.1999.9234.

Mitsudomi, T., K. Nishioka, R. Maruyama, G. Saitoh, M. Hamatake, Y. Fukuyama, H. Yaita, T. Ishida, and K. Sugimachi. 1996. “Kinetic Analysis of Recurrence and Survival after Potentially Curative Resection of Nonsmall Cell Lung Cancer.” *Journal of Surgical Oncology* 63 (3): 159–65.

Mittrücker, Hans-Willi, Alexander Visekruna, and Magdalena Huber. 2014. “Heterogeneity in the Differentiation and Function of $CD8^+$ T Cells.” *Archivum Immunologiae et Therapiae Experimentalis* 62 (6): 449–58.

Moisés, Jorge, Alfons Navarro, Sandra Santasusagna, Nuria Viñolas, Laureano Molins, José Ramirez, Jeisson Osorio, et al. 2017. “NKX2–1 Expression as a Prognostic Marker in Early-Stage Non-Small-Cell Lung Cancer.” *BMC Pulmonary Medicine* 17 (1): 1–9.

Montesion, Meagan, Karthikeyan Murugesan, Dexter X. Jin, Radwa Sharaf, Nora Sanchez, Ameet Guria, Max Minker, et al. 2021. “Somatic HLA Class I Loss Is a Widespread Mechanism of Immune Evasion Which Refines the Use of Tumor Mutational Burden as a Biomarker of Checkpoint Inhibitor Response.” *Cancer Discovery* 11 (2): 282–92.

Morris, Samantha A., and George Q. Daley. 2013. "A Blueprint for Engineering Cell Fate: Current Technologies to Reprogram Cell Identity." *Cell Research* 23 (1): 33–48.

Moyer, Virginia A., and U.S. Preventive Services Task Force. 2014. "Screening for Lung Cancer: U.S. Preventive Services Task Force Recommendation Statement." *Annals of Internal Medicine* 160 (5): 330–38.

National Comprehensive Cancer Network. 2021. "Non-Small Cell Lung Cancer (Version 5.2021)." May 2021. https://www.nccn.org/professionals/physician_gls/pdf/nscl_blocks.pdf.

National Institutes of Health (US), and Biological Sciences Curriculum Study. 2007. *Understanding Cancer*. National Institutes of Health (US).

Negrini, Simona, Vassilis G. Gorgoulis, and Thanos D. Halazonetis. 2010. "Genomic Instability--an Evolving Hallmark of Cancer." *Nature Reviews. Molecular Cell Biology* 11 (3): 220–28.

Ng, Sheng Rong, William M. Rideout 3rd, Elliot H. Akama-Garren, Arjun Bhutkar, Kim L. Mercer, Jason M. Schenkel, Roderick T. Bronson, and Tyler Jacks. 2020. "CRISPR-Mediated Modeling and Functional Validation of Candidate Tumor Suppressor Genes in Small Cell Lung Cancer." *Proceedings of the National Academy of Sciences of the United States of America* 117 (1): 513–21.

Nigro, J. M., S. J. Baker, A. C. Preisinger, J. M. Jessup, R. Hostetter, K. Cleary, S. H. Bigner, N. Davidson, S. Baylin, and P. Devilee. 1989. "Mutations in the p53 Gene Occur in Diverse Human Tumour Types." *Nature* 342 (6250): 705–8.

"Nonlinear Oscillations in Chemical and Biological Systems." 1991. *Physica D. Nonlinear Phenomena* 51 (1-3): 152–60.

Nowell, P. C. 1976. "The Clonal Evolution of Tumor Cell Populations." *Science* 194 (4260): 23–28.

Nowell, Peter C. 1978. "Tumors as Clonal Proliferation." *Virchows Archiv. B: Cell Pathology* 29 (1): 145–50.

O'Hagan, R. C., and J. A. Hassell. 1998. "The PEA3 Ets Transcription Factor Is a Downstream Target of the HER2/Neu Receptor Tyrosine Kinase." *Oncogene* 16 (3). https://doi.org/10.1038/sj.onc.1201547.

Okita, Keisuke, Tomoko Ichisaka, and Shinya Yamanaka. 2007. "Generation of Germline-Competent Induced Pluripotent Stem Cells." *Nature* 448 (7151): 313–17.

Okubo, Tadashi, Paul S. Knoepfler, Robert N. Eisenman, and Brigid L. M. Hogan. 2005. "Nmyc Plays an Essential Role during Lung Development as a Dosage-Sensitive Regulator of Progenitor Cell Proliferation and Differentiation." *Development* 132 (6): 1363–74.

Olsen, L. F., and H. Degn. 1985. "Chaos in Biological Systems." *Quarterly Reviews of Biophysics* 18 (2): 165–225.

Orre, Lukas M., Maria Pernemalm, Johan Lengqvist, Rolf Lewensohn, and Janne Lehtiö. 2007. "Up-Regulation, Modification, and Translocation of S100A6 Induced by Exposure to Ionizing Radiation Revealed by Proteomics Profiling." *Molecular & Cellular Proteomics: MCP* 6 (12): 2122–31.

Ott, Patrick A., Zhuting Hu, Derin B. Keskin, Sachet A. Shukla, Jing Sun, David J. Bozym, Wandi Zhang, et al. 2017. "An Immunogenic Personal Neoantigen Vaccine for Patients with Melanoma." *Nature* 547 (7662): 217–21.

Pacheco, J. M., D. Gao, D. Smith, T. Purcell, M. Hancock, P. Bunn, T. Robin, et al. 2019. “Natural History and Factors Associated with Overall Survival in Stage IV ALK-Rearranged Non-Small Cell Lung Cancer.” *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer* 14 (4). https://doi.org/10.1016/j.jtho.2018.12.014.

Packer, Jonathan S., Qin Zhu, Chau Huynh, Priya Sivaramakrishnan, Elicia Preston, Hannah Dueck, Derek Stefanik, et al. 2019. “A Lineage-Resolved Molecular Atlas of C. Elegans Embryogenesis at Single-Cell Resolution.” *Science* 365 (6459). https://doi.org/10.1126/science.aax1971.

Paley, Michael A., Daniela C. Kroy, Pamela M. Odorizzi, Jonathan B. Johnnidis, Douglas V. Dolfi, Burton E. Barnett, Elizabeth K. Bikoff, et al. 2012. “Progenitor and Terminal Subsets of CD8+ T Cells Cooperate to Contain Chronic Viral Infection.” *Science* 338 (6111): 1220–25.

Papadaki, Chara, Maria Alexiou, Grace Cecena, Mihalis Verykokakis, Aikaterini Bilitou, James C. Cross, Robert G. Oshima, and George Mavrothalassitis. 2007. “Transcriptional Repressor Erf Determines Extraembryonic Ectoderm Differentiation.” *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.02237-06.

Paris, Andrew J., Katharina E. Hayer, Joseph H. Oved, Daphne C. Avgousti, Sushila A. Toulmin, Jarod A. Zepp, William J. Zacharias, et al. 2020. “STAT3–BDNF–TrkB Signalling Promotes Alveolar Epithelial Regeneration after Lung Injury.” *Nature Cell Biology*. https://doi.org/10.1038/s41556-020-0569-x.

Parish, Ian A., Jason Waithman, Gayle M. Davey, Gabrielle T. Belz, Justine D. Mintern, Christian Kurts, Robyn M. Sutherland, Francis R. Carbone, and William R. Heath. 2009. “Tissue Destruction Caused by Cytotoxic T Lymphocytes Induces Deletional Tolerance.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (10): 3901–6.

Parry, William. 1979. “The Lorenz Attractor and a Related Population Model.” In *Ergodic Theory*, 169–87. Springer Berlin Heidelberg.

Peinado, Héctor, Haiying Zhang, Irina R. Matei, Bruno Costa-Silva, Ayuko Hoshino, Goncalo Rodrigues, Bethan Psaila, et al. 2017. “Pre-Metastatic Niches: Organ-Specific Homes for Metastases.” *Nature Reviews. Cancer* 17 (5): 302–17.

Peng, Miaoguan, Guohong Wei, Yunjian Zhang, Hai Li, Yingrong Lai, Yan Guo, Yuxin Chen, et al. 2021. “Single-Cell Transcriptomic Landscape Reveals the Differences in Cell Differentiation and Immune Microenvironment of Papillary Thyroid Carcinoma between Genders.” *Cell & Bioscience* 11 (1): 39.

Pepicelli, C. V., P. M. Lewis, and A. P. McMahon. 1998. “Sonic Hedgehog Regulates Branching Morphogenesis in the Mammalian Lung.” *Current Biology: CB* 8 (19): 1083–86.

Perl, Anne-Karina T., Isamu Hokuto, Maria-Antonietta Impagnatiello, Gerhard Christofori, and Jeffrey A. Whitsett. 2003. “Temporal Effects of Sprouty on Lung Morphogenesis.” *Developmental Biology* 258 (1): 154–68.

Perl, Anne-Karina T., Susan E. Wert, Andras Nagy, Corrinne G. Lobe, and Jeffrey A. Whitsett. 2002. “Early Restriction of Peripheral and Proximal Cell Lineages during Formation of the Lung.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (16): 10482–87.

"Permanence and Extinction of a Stochastic Hybrid Model for Tumor Growth." 2019. *Applied Mathematics Letters* 94 (August): 66–72.

Peters, Bjoern, Morten Nielsen, and Alessandro Sette. 2020. "T Cell Epitope Predictions." *Annual Review of Immunology* 38 (April): 123–45.

Pfannenstiel, Lukas W., C. Marcela Diaz-Montero, Ye F. Tian, Joseph Scharpf, Jennifer S. Ko, and Brian R. Gastman. 2019. "Immune-Checkpoint Blockade Opposes CD8+ T-Cell Suppression in Human and Murine Cancer." *Cancer Immunology Research* 7 (3): 510–25.

Phloyphisut, Poomarin, Natapol Pornputtapong, Sira Sriswasdi, and Ekapol Chuangsuwanich. 2019. "MHCSeqNet: A Deep Neural Network Model for Universal MHC Binding Prediction." *BMC Bioinformatics* 20 (1): 270.

Plass, Christoph, Stefan M. Pfister, Anders M. Lindroth, Olga Bogatyrova, Rainer Claus, and Peter Lichter. 2013. "Mutations in Regulators of the Epigenome and Their Connections to Global Chromatin Patterns in Cancer." *Nature Reviews. Genetics* 14 (11): 765–80.

Plass, Mireya, Jordi Solana, F. Alexander Wolf, Salah Ayoub, Aristotelis Misios, Petar Glažar, Benedikt Obermayer, Fabian J. Theis, Christine Kocks, and Nikolaus Rajewsky. 2018. "Cell Type Atlas and Lineage Tree of a Whole Complex Animal by Single-Cell Transcriptomics." *Science* 360 (6391). https://doi.org/10.1126/science.aaq1723.

Platt, Randall J., Sidi Chen, Yang Zhou, Michael J. Yim, Lukasz Swiech, Hannah R. Kempton, James E. Dahlman, et al. 2014. "CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling." *Cell* 159 (2): 440–55.

Qiu, Xiaojie, Qi Mao, Ying Tang, Li Wang, Raghav Chawla, Hannah A. Pliner, and Cole Trapnell. 2017. "Reversed Graph Embedding Resolves Complex Single-Cell Trajectories." *Nature Methods* 14 (10): 979–82.

Raimondi, Cristina, Guido Carpino, Chiara Nicolazzo, Angela Gradilone, Walter Gianni, Alain Gelibter, Eugenio Gaudio, Enrico Cortesi, and Paola Gazzaniga. 2017. "PD-L1 and Epithelial-Mesenchymal Transition in Circulating Tumor Cells from Non-Small Cell Lung Cancer Patients: A Molecular Shield to Evade Immune System?" *Oncoimmunology* 6 (12): e1315488.

Ramalingam, S. S., J. Vansteenkiste, D. Planchard, B. C. Cho, J. E. Gray, Y. Ohe, C. Zhou, et al. 2020. "Overall Survival with Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC." *The New England Journal of Medicine* 382 (1). https://doi.org/10.1056/NEJMoa1913662.

Raphael, Benjamin J., Jason R. Dobson, Layla Oesper, and Fabio Vandin. 2014. "Identifying Driver Mutations in Sequenced Cancer Genomes: Computational Approaches to Enable Precision Medicine." *Genome Medicine* 6 (1): 5.

Raskov, Hans, Adile Orhan, Jan Pravsgaard Christensen, and Ismail Gögenur. 2020. "Cytotoxic CD8 + T Cells in Cancer and Cancer Immunotherapy." *British Journal of Cancer* 124 (2): 359–67.

Rawlins, Emma L., Cheryl P. Clark, Yan Xue, and Brigid L. M. Hogan. 2009. "The Id2+ Distal Tip Lung Epithelium Contains Individual Multipotent Embryonic Progenitor Cells." *Development* 136 (22): 3741–45.

Reck, M., D. Rodríguez-Abreu, A. G. Robinson, R. Hui, T. Csőszi, A. Fülöp, M. Gottfried, et al. 2019. "Updated Analysis of KEYNOTE-024: Pembrolizumab Versus

Platinum-Based Chemotherapy for Advanced Non-Small-Cell Lung Cancer With PD-L1 Tumor Proportion Score of 50% or Greater." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 37 (7). https://doi.org/10.1200/JCO.18.00149.

Regev, Aviv, Sarah A. Teichmann, Eric S. Lander, Ido Amit, Christophe Benoist, Ewan Birney, Bernd Bodenmiller, et al. 2017. "The Human Cell Atlas." *eLife* 6 (December). https://doi.org/10.7554/eLife.27041.

Regev, Aviv, Sarah Teichmann, Orit Rozenblatt-Rosen, Michael Stubbington, Kristin Ardlie, Ido Amit, Paola Arlotta, et al. 2018. "The Human Cell Atlas White Paper." *arXiv [q-bio.TO]*. arXiv. http://arxiv.org/abs/1810.05192.

Reinberg, Danny, and Lynne D. Vales. 2018. "Chromatin Domains Rich in Inheritance." *Science* 361 (6397): 33–34.

Révész, D., E. G. Engelhardt, J. J. Tamminga, F. M. N. H. Schramel, B. D. Onwuteaka-Philipsen, E. M. W. van de Garde, E. W. Steyerberg, E. P. Jansma, H. C. W. De Vet, and V. M. H. Coupé. 2017. "Decision Support Systems for Incurable Non-Small Cell Lung Cancer: A Systematic Review." *BMC Medical Informatics and Decision Making* 17 (1): 144.

Rizvi, Naiyer A., Matthew D. Hellmann, Alexandra Snyder, Pia Kvistborg, Vladimir Makarov, Jonathan J. Havel, William Lee, et al. 2015. "Mutational Landscape Determines Sensitivity to PD-1 Blockade in Non–small Cell Lung Cancer." *Science* 348 (6230): 124–28.

Robert, Caroline. 2020. "A Decade of Immune-Checkpoint Inhibitors in Cancer Therapy." *Nature Communications* 11 (1): 3801.

Roberts, J. A. G., and G. R. W. Quispel. 1992. "Chaos and Time-Reversal Symmetry. Order and Chaos in Reversible Dynamical Systems." *Physics Reports* 216 (2): 63–177.

Rogers, Zoë N., Christopher D. McFarland, Ian P. Winters, Jose A. Seoane, Jennifer J. Brady, Stephanie Yoon, Christina Curtis, Dmitri A. Petrov, and Monte M. Winslow. 2018. "Mapping the in Vivo Fitness Landscape of Lung Adenocarcinoma Tumor Suppression in Mice." *Nature Genetics* 50 (4): 483–86.

"Role for ETS Domain Transcription Factors Pea3/Erm in Mouse Lung Development." 2003. *Developmental Biology* 261 (1): 10–24.

Romero, Rodrigo, Francisco J. Sánchez-Rivera, Peter M. K. Westcott, Kim L. Mercer, Arjun Bhutkar, Alexander Muir, Tania J. González Robles, et al. 2020. "Keap1 Mutation Renders Lung Adenocarcinomas Dependent on Slc33a1." *Nature Cancer* 1 (6): 589–602.

Rosenberg, Steven A., and Nicholas P. Restifo. 2015. "Adoptive Cell Transfer as Personalized Immunotherapy for Human Cancer." *Science* 348 (6230): 62–68.

Rozenblatt-Rosen, Orit, Michael J. T. Stubbington, Aviv Regev, and Sarah A. Teichmann. 2017. "The Human Cell Atlas: From Vision to Reality." *Nature* 550 (7677): 451–53.

Ruan, Haoyu, Yihang Zhou, Jie Shen, Yue Zhai, Ying Xu, Linyu Pi, Ruofan Huang, et al. 2020. "Circulating Tumor Cell Characterization of Lung Cancer Brain Metastases in the Cerebrospinal Fluid through Single-Cell Transcriptome Analysis." *Clinical and Translational Medicine* 10 (8): e246.

Ruf-Zamojski, Frederique, Yongchao Ge, Venugopalan Nair, Michel Zamojski, Hanna

Pincas, Chirine Toufaily, Jessica Tome-Garcia, et al. 2018. “Single-Cell Stabilization Method Identifies Gonadotrope Transcriptional Dynamics and Pituitary Cell Type Heterogeneity.” *Nucleic Acids Research* 46 (21): 11370–80.

Rye, Inga H., Anne Trinh, Anna B. Saetersdal, Daniel Nebdal, Ole Christian Lingjaerde, Vanessa Almendro, Kornelia Polyak, et al. 2018. “Intratumor Heterogeneity Defines Treatment-Resistant HER2+ Breast Tumors.” *Molecular Oncology* 12 (11): 1838–55.

Sade-Feldman, Moshe, Keren Yizhak, Stacey L. Bjorgaard, John P. Ray, Carl G. de Boer, Russell W. Jenkins, David J. Lieb, et al. 2019. “Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma.” *Cell* 176 (1-2): 404.

Safford, Meredith, Samuel Collins, Michael A. Lutz, Amy Allen, Ching-Tai Huang, Jeanne Kowalski, Amanda Blackford, et al. 2005. “Egr-2 and Egr-3 Are Negative Regulators of T Cell Activation.” *Nature Immunology* 6 (5): 472–80.

Sahin, Ugur, Evelyna Derhovanessian, Matthias Miller, Björn-Philipp Kloke, Petra Simon, Martin Löwer, Valesca Bukur, et al. 2017. “Personalized RNA Mutanome Vaccines Mobilize Poly-Specific Therapeutic Immunity against Cancer.” *Nature* 547 (7662): 222–26.

Salwig, Isabelle, Birgit Spitznagel, Ana Ivonne Vazquez-Armendariz, Keynoosh Khalooghi, Stefan Guenther, Susanne Herold, Marten Szibor, and Thomas Braun. 2019. “Bronchioalveolar Stem Cells Are a Main Source for Regeneration of Distal Lung Epithelia in Vivo.” *The EMBO Journal* 38 (12). https://doi.org/10.15252/embj.2019102099.

Sánchez-Rivera, Francisco J., Thales Papagiannakopoulos, Rodrigo Romero, Tuomas Tammela, Matthew R. Bauer, Arjun Bhutkar, Nikhil S. Joshi, et al. 2014. “Rapid Modelling of Cooperating Genetic Events in Cancer through Somatic Genome Editing.” *Nature* 516 (7531): 428–31.

Sanjana, Neville E., Ophir Shalem, and Feng Zhang. 2014. “Improved Vectors and Genome-Wide Libraries for CRISPR Screening.” *Nature Methods* 11 (8): 783–84.

Sanson, Kendall R., Ruth E. Hanna, Mudra Hegde, Katherine F. Donovan, Christine Strand, Meagan E. Sullender, Emma W. Vaimberg, et al. 2018. “Optimized Libraries for CRISPR-Cas9 Genetic Screens with Multiple Modalities.” *Nature Communications* 9 (1): 5416.

Santos-Rosa, Helena, Robert Schneider, Andrew J. Bannister, Julia Sherriff, Bradley E. Bernstein, N. C. Tolga Emre, Stuart L. Schreiber, Jane Mellor, and Tony Kouzarides. 2002. “Active Genes Are Tri-Methylated at K4 of Histone H3.” *Nature* 419 (6905): 407–11.

Schatz, David G., and Yanhong Ji. 2011. “Recombination Centres and the Orchestration of V (D) J Recombination.” *Nature Reviews. Immunology* 11 (4): 251–63.

Schenkel, Jason, Rebecca Herbst, David Canner, Michelle Hillman, Amy Li, Olivia C. Smith, Jonathan Y. Kim, et al. 2020. “Increased Dysfunction of tumor-specific TCF-1+ CD8+ T Cells Associated with Increase in Non-Cycling TCF-1+ SlamF6- Cells Subsets Revealed by Single Cell Analysis.” https://doi.org/10.2139/ssrn.3707275.

Schiebinger, Geoffrey, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, et al. 2019. “Optimal-Transport Analysis of

Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming." *Cell* 176 (4): 928–43.e22.
Schietinger, Andrea, and Philip D. Greenberg. 2014. "Tolerance and Exhaustion: Defining Mechanisms of T Cell Dysfunction." *Trends in Immunology* 35 (2): 51–60.
Schmale, Hartwig, and Casimir Bamberger. 1997. "A Novel Protein with Strong Homology to the Tumor Suppressor p53." *Oncogene* 15 (11): 1363–67.
Schneider, Günter, Marc Schmidt-Supprian, Roland Rad, and Dieter Saur. 2017. "tissue specific Tumorigenesis: Context Matters." *Nature Reviews. Cancer* 17 (4): 239–53.
Schreiber, H., T. H. Wu, J. Nachman, and W. M. Kast. 2002. "Immunodominance and Tumor Escape." *Seminars in Cancer Biology* 12 (1): 25–31.
Schreiber, Robert D., Lloyd J. Old, and Mark J. Smyth. 2011. "Cancer Immunoediting: Integrating Immunity's Roles in Cancer Suppression and Promotion." *Science* 331 (6024): 1565–70.
Schulte-Schrepping, Jonas, Humberto J. Ferreira, Adem Saglam, Emily Hinkley, and Joachim L. Schultze. 2020. "Advances in Single-Cell Epigenomics of the Immune System." *Epigenetics of the Immune System*. https://doi.org/10.1016/b978-0-12-817964-2.00008-3.
Schumacher, Ton N., Wouter Scheper, and Pia Kvistborg. 2019. "Cancer Neoantigens." *Annual Review of Immunology* 37 (April): 173–200.
Sell, Stewart. 2010. "On the Stem Cell Origin of Cancer." *The American Journal of Pathology* 176 (6): 2584.
Serna, Leidy Y., Miguel A. Mañanas, Alher M. Hernández, and Roberto A. Rabinovich. 2018. "An Improved Dynamic Model for the Respiratory Response to Exercise." *Frontiers in Physiology* 9. https://doi.org/10.3389/fphys.2018.00069.
Sgouras, D. N., M. A. Athanasiou, G. J. Beal Jr, R. J. Fisher, D. G. Blair, and G. J. Mavrothalassitis. 1995. "ERF: An ETS Domain Protein with Strong Transcriptional Repressor Activity, Can Suppress Ets-Associated Tumorigenesis and Is Regulated by Phosphorylation during Cell Cycle and Mitogenic Stimulation." *The EMBO Journal* 14 (19): 4781–93.
Shackleton, Mark, Elsa Quintana, Eric R. Fearon, and Sean J. Morrison. 2009. "Heterogeneity in Cancer: Cancer Stem Cells versus Clonal Evolution." *Cell* 138 (5): 822–29.
Sharma, A., and J. M. Sen. 2013. "Molecular Basis for the Tissue Specificity of β-Catenin Oncogenesis." *Oncogene* 32 (15): 1901–9.
Sharma, Shikhar, Theresa K. Kelly, and Peter A. Jones. 2010. "Epigenetics in Cancer." *Carcinogenesis* 31 (1): 27–36.
Shaw, A. T., G. J. Riely, Y. J. Bang, D. W. Kim, D. R. Camidge, B. J. Solomon, M. Varella-Garcia, et al. 2019. "Crizotinib in ROS1-Rearranged Advanced Non-Small-Cell Lung Cancer (NSCLC): Updated Results, Including Overall Survival, from PROFILE 1001." *Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO* 30 (7). https://doi.org/10.1093/annonc/mdz131.
Shemesh, Colby S., Joy C. Hsu, Iraj Hosseini, Ben-Quan Shen, Anand Rotte, Patrick Twomey, Sandhya Girish, and Benjamin Wu. 2021. "Personalized Cancer Vaccines: Clinical Landscape, Challenges, and Opportunities." *Molecular Therapy: The Journal of the American Society of Gene Therapy* 29 (2): 555–70.

Shim, Woo Jun, Enakshi Sinniah, Jun Xu, Burcu Vitrinel, Michael Alexanian, Gaia Andreoletti, Sophie Shen, et al. 2020. “Conserved Epigenetic Regulatory Logic Infers Genes Governing Cell Identity.” *Cell Systems* 11 (6): 625–39.e13.

Shi, Yuankai, Joseph Siu-Kie Au, Sumitra Thongprasert, Sankar Srinivasan, Chun-Ming Tsai, Mai Trong Khoa, Karin Heeroma, Yohji Itoh, Gerardo Cornelio, and Pan-Chyr Yang. 2014. “A Prospective, Molecular Epidemiology Study of EGFR Mutations in Asian Patients with Advanced Non-Small-Cell Lung Cancer of Adenocarcinoma Histology (PIONEER).” *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer* 9 (2): 154–62.

Shlush, Liran I., and Dov Hershkovitz. 2015. “Clonal Evolution Models of Tumor Heterogeneity.” *American Society of Clinical Oncology Educational Book. American Society of Clinical Oncology. Annual Meeting*, e662–65.

Shu, Weiguo, Susan Guttentag, Zhishan Wang, Thomas Andl, Philip Ballard, Min Min Lu, Stefano Piccolo, et al. 2005. “Wnt/β-Catenin Signaling Acts Upstream of N-Myc, BMP4, and FGF Signaling to Regulate Proximal–distal Patterning in the Lung.” *Developmental Biology* 283 (1): 226–39.

Siddiqui, Imran, Karin Schaeuble, Vijaykumar Chennupati, Silvia A. Fuertes Marraco, Sandra Calderon-Copete, Daniela Pais Ferreira, Santiago J. Carmona, et al. 2019. “Intratumoral Tcf1PD-1CD8 T Cells with Stem-like Properties Promote Tumor Control in Response to Vaccination and Checkpoint Blockade Immunotherapy.” *Immunity* 50 (1): 195–211.e10.

Siegel, Rebecca L., Kimberly D. Miller, and Ahmedin Jemal. 2018. “Cancer Statistics, 2018.” *CA: A Cancer Journal for Clinicians*. https://doi.org/10.3322/caac.21442.

Silverstein, Arthur M. 2001. *Paul Ehrlich’s Receptor Immunology: The Magnificent Obsession*. Elsevier.

Singh, A., and J. Settleman. 2010. “EMT, Cancer Stem Cells and Drug Resistance: An Emerging Axis of Evil in the War on Cancer.” *Oncogene* 29 (34): 4741–51.

Singh, Indrabahadur, Aditi Mehta, Adriana Contreras, Thomas Boettger, Gianni Carraro, Matthew Wheeler, Hector A. Cabrera-Fuentes, et al. 2014. “Hmga2 Is Required for Canonical WNT Signaling during Lung Development.” *BMC Biology* 12 (March): 21.

Sizemore, Gina M., Jason R. Pitarresi, Subhasree Balakrishnan, and Michael C. Ostrowski. 2017. “The ETS Family of Oncogenic Transcription Factors in Solid Tumours.” *Nature Reviews. Cancer* 17 (6): 337–51.

Sklar, Lawrence. 1976. “Space, Time, and Spacetime.” https://doi.org/10.1525/9780520340701.

Smith-Garvin, Jennifer E., Gary A. Koretzky, and Martha S. Jordan. 2009. “T Cell Activation.” *Annual Review of Immunology* 27: 591–619.

Smolander, Johannes, Sini Junttila, Mikko S. Venäläinen, and Laura L. Elo. 2021. “ILoReg: A Tool for High-Resolution Cell Population Identification from Single-Cell RNA-Seq Data.” *Bioinformatics* 37 (8): 1107–14.

Snyder, Eric L., Hideo Watanabe, Margaret Magendantz, Sebastian Hoersch, Tiffany A. Chen, Diana G. Wang, Denise Crowley, et al. 2013. “Nkx2-1 Represses a Latent Gastric Differentiation Program in Lung Adenocarcinoma.” *Molecular Cell* 50 (2): 185–99.

Sparano, Joseph A., Robert J. Gray, Della F. Makower, Kathleen I. Pritchard, Kathy S. Albain, Daniel F. Hayes, Charles E. Geyer Jr, et al. 2018. “Adjuvant Chemotherapy

Guided by a 21-Gene Expression Assay in Breast Cancer." *The New England Journal of Medicine* 379 (2): 111–21.
Speiser, Daniel E., Ping-Chih Ho, and Grégory Verdeil. 2016. "Regulatory Circuits of T Cell Function in Cancer." *Nature Reviews. Immunology* 16 (10): 599–611.
Stanta, Giorgio, and Serena Bonin. 2018. "Overview on Clinical Relevance of Intra-Tumor Heterogeneity." *Frontiers of Medicine* 5 (April): 85.
Stearns, Vered. 2018. "TAILORing Adjuvant Systemic Therapy for Breast Cancer." *The New England Journal of Medicine*.
Stehelin, D., H. E. Varmus, J. M. Bishop, and P. K. Vogt. 1976. "DNA Related to the Transforming Gene(s) of Avian Sarcoma Viruses Is Present in Normal Avian DNA." *Nature*. https://doi.org/10.1038/260170a0.
Stoeckius, Marlon, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K. Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. 2017. "Simultaneous Epitope and Transcriptome Measurement in Single Cells." *Nature Methods* 14 (9): 865–68.
Strickler, John H., Brent A. Hanks, and Mustafa Khasraw. 2021. "Tumor Mutational Burden as a Predictor of Immunotherapy Response: Is More Always Better?" *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 27 (5): 1236–41.
Strønen, Erlend, Mireille Toebes, Sander Kelderman, Marit M. van Buuren, Weiwen Yang, Nienke van Rooij, Marco Donia, et al. 2016. "Targeting of Cancer Neoantigens with Donor-Derived T Cell Receptor Repertoires." *Science* 352 (6291): 1337–41.
Stuart, Tim, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M. Mauck 3rd, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. 2019. "Comprehensive Integration of Single-Cell Data." *Cell* 177 (7): 1888–1902.e21.
Sugimoto, Hikaru, Thomas M. Mundel, Mark W. Kieran, and Raghu Kalluri. 2006. "Identification of Fibroblast Heterogeneity in the Tumor Microenvironment." *Cancer Biology & Therapy* 5 (12): 1640–46.
Sun, Shiquan, Jiaqiang Zhu, Ying Ma, and Xiang Zhou. 2019. "Accuracy, Robustness and Scalability of Dimensionality Reduction Methods for Single-Cell RNA-Seq Analysis." *Genome Biology* 20 (1): 269.
Sutherland, Kate D., Ji-Ying Song, Min Chul Kwon, Natalie Proost, John Zevenhoven, and Anton Berns. 2014. "Multiple Cells-of-Origin of Mutant K-Ras-Induced Mouse Lung Adenocarcinoma." *Proceedings of the National Academy of Sciences of the United States of America* 111 (13): 4952–57.
Swann, Jeremy B., and Mark J. Smyth. 2007. "Immune Surveillance of Tumors." *The Journal of Clinical Investigation* 117 (5): 1137–46.
Szczurek, Ewa, Tyll Krüger, Barbara Klink, and Niko Beerenwinkel. 2020. "A Mathematical Model of the Metastatic Bottleneck Predicts Patient Outcome and Response to Cancer Treatment." *PLoS Computational Biology* 16 (10): e1008056.
Tabassum, Doris P., and Kornelia Polyak. 2015. "Tumorigenesis: It Takes a Village." *Nature Reviews. Cancer* 15 (8): 473–83.
Tammela, Tuomas, Francisco J. Sanchez-Rivera, Naniye Malli Cetinbas, Katherine Wu, Nikhil S. Joshi, Katja Helenius, Yoona Park, et al. 2017. "A Wnt-Producing Niche

Drives Proliferative Potential and Progression in Lung Adenocarcinoma." *Nature* 545 (7654): 355–59.

Tarabichi, Maxime, Iñigo Martincorena, Moritz Gerstung, Armand M. Leroi, Florian Markowetz, PCAWG Evolution and Heterogeneity Working Group, Paul T. Spellman, et al. 2018. "Neutral Tumor Evolution?" *Nature Genetics*.

Tchernitsa, Oleg I., Christine Sers, Johannes Zuber, Bernd Hinzmann, Martin Grips, Anja Schramme, Per Lund, Anke Schwendel, André Rosenthal, and Reinhold Schäfer. 2004. "Transcriptional Basis of KRAS Oncogene-Mediated Cellular Transformation in Ovarian Epithelial Cells." *Oncogene* 23 (26): 4536–55.

Tefft, J. D., M. Lee, S. Smith, M. Leinwand, J. Zhao, P. Bringas Jr, D. L. Crowe, and D. Warburton. 1999. "Conserved Function of mSpry-2, a Murine Homolog of Drosophila Sprouty, Which Negatively Modulates Respiratory Organogenesis." *Current Biology: CB* 9 (4): 219–22.

Temel, Jennifer S., Joseph A. Greer, Alona Muzikansky, Emily R. Gallagher, Sonal Admane, Vicki A. Jackson, Constance M. Dahlin, et al. 2010. "Early Palliative Care for Patients with Metastatic Non-Small-Cell Lung Cancer." *The New England Journal of Medicine* 363 (8): 733–42.

Thaiss, Christoph A., Verena Semmling, Lars Franken, Hermann Wagner, and Christian Kurts. 2011. "Chemokines: A New Dendritic Cell Signal for T Cell Activation." *Frontiers in Immunology* 2 (August): 31.

Thangavelu, Govindarajan, Christa Smolarchuk, and Colin C. Anderson. 2010. "Co-Inhibitory Molecules: Controlling the Effectors or Controlling the Controllers?" *Self/nonself* 1 (2): 77–88.

"The Edwin Smith Surgical Papyrus. Published in Faesimile and Hieroglyphie Transliteration, with Translation and Commentary. By James Henry Breasted. Demy 4to. Vol. I. Pp. 596 + Xxiv, Illustrated; Vol. II. 22 Plates. 1930. Chicago: The Univesity of Chicago Press. London: Cambridge University Press. 90s." 2005. *The British Journal of Surgery* 18 (70): 341–42.

Thommen, Daniela S., and Ton N. Schumacher. 2018. "T Cell Dysfunction in Cancer." *Cancer Cell* 33 (4): 547–62.

Thrane, E. V., R. Becher, M. Låg, M. Refsnes, H. S. Huitfeldt, and P. E. Schwarze. 1997. "Differential Distribution and Increased Levels of Ras Proteins during Lung Development." *Experimental Lung Research* 23 (1): 35–49.

Thul, Peter J., Lovisa Åkesson, Mikaela Wiking, Diana Mahdessian, Aikaterini Geladaki, Hammou Ait Blal, Tove Alm, et al. 2017. "A Subcellular Map of the Human Proteome." *Science* 356 (6340). https://doi.org/10.1126/science.aal3321.

Timmerman, J. M., and R. Levy. 2000. "The History of the Development of Vaccines for the Treatment of Lymphoma." *Clinical Lymphoma* 1 (2): 129–39; discussion 140.

Tippimanchai, Darinee D., Kyle Nolan, Joanna Poczobutt, Gregory Verzosa, Howard Li, Hannah Scarborough, Jing Huang, et al. 2018. "Adenoviral Vectors Transduce Alveolar Macrophages in Lung Cancer Models." *Oncoimmunology* 7 (6): e1438105.

Toh, Tan Boon, Jhin Jieh Lim, and Edward Kai-Hua Chow. 2017. "Epigenetics in Cancer Stem Cells." *Molecular Cancer* 16 (1): 29.

Toker, Daniel, Friedrich T. Sommer, and Mark D'Esposito. 2020. "A Simple Method for Detecting Chaos in Nature." *Communications Biology* 3 (January): 11.

Traag, V. A., L. Waltman, and N. J. van Eck. 2019. "From Louvain to Leiden:

Guaranteeing Well-Connected Communities.” *Scientific Reports*. https://doi.org/10.1038/s41598-019-41695-z.

Tran, Eric, Paul F. Robbins, Yong-Chen Lu, Todd D. Prickett, Jared J. Gartner, Li Jia, Anna Pasetto, et al. 2016. “T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer.” *The New England Journal of Medicine* 375 (23): 2255–62.

Tran, Hoa Thi Nhu, Kok Siong Ang, Marion Chevrier, Xiaomeng Zhang, Nicole Yee Shin Lee, Michelle Goh, and Jinmiao Chen. 2020. “A Benchmark of Batch-Effect Correction Methods for Single-Cell RNA Sequencing Data.” *Genome Biology* 21 (1): 12.

Trapnell, Cole, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J. Lennon, Kenneth J. Livak, Tarjei S. Mikkelsen, and John L. Rinn. 2014. “The Dynamics and Regulators of Cell Fate Decisions Are Revealed by Pseudotemporal Ordering of Single Cells.” *Nature Biotechnology* 32 (4): 381–86.

Travaglini, Kyle J., Ahmad N. Nabhan, Lolita Penland, Rahul Sinha, Astrid Gillich, Rene V. Sit, Stephen Chang, et al. 2020. “A Molecular Cell Atlas of the Human Lung from Single-Cell RNA Sequencing.” *Nature* 587 (7835): 619–25.

Travis, William D., Elisabeth Brambilla, Allen P. Burke, Alexander Marx, and Andrew G. Nicholson. 2015. “Introduction to The 2015 World Health Organization Classification of Tumors of the Lung, Pleura, Thymus, and Heart.” *Journal of Thoracic Oncology*. https://doi.org/10.1097/jto.0000000000000663.

Travis, William D., Elisabeth Brambilla, Andrew G. Nicholson, Yasushi Yatabe, John H. M. Austin, Mary Beth Beasley, Lucian R. Chirieac, et al. 2015. “The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification.” *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer* 10 (9): 1243–60.

Tsallis, Constantino. 1988. “Possible Generalization of Boltzmann-Gibbs Statistics.” *Journal of Statistical Physics* 52 (1): 479–87.

Tsuyuzaki, Koki, Hiroyuki Sato, Kenta Sato, and Itoshi Nikaido. 2020. “Benchmarking Principal Component Analysis for Large-Scale Single-Cell RNA-Sequencing.” *Genome Biology* 21 (1): 9.

Tulp, Nicolaas. 1716. “Nicolai Tulpii Amstelodensis Exconsulis Observationes Medicæ.” https://doi.org/10.5962/bhl.title.119776.

“Tumor Functional Heterogeneity Unraveled by scRNA-Seq Technologies.” 2020. *Trends in Cancer Research* 6 (1): 13–19.

Turk, Matthew, and Alex Pentland. 1991. “Eigenfaces for Recognition.” *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn.1991.3.1.71.

Uhlén, Mathias, Linn Fagerberg, Björn M. Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, et al. 2015. “Proteomics. Tissue-Based Map of the Human Proteome.” *Science* 347 (6220): 1260419.

Utzschneider, Daniel T., Mélanie Charmoy, Vijaykumar Chennupati, Laurène Pousse, Daniela Pais Ferreira, Sandra Calderon-Copete, Maxime Danilo, et al. 2016. “T Cell Factor 1-Expressing Memory-like CD8+ T Cells Sustain the Immune Response to Chronic Viral Infections.” *Immunity* 45 (2): 415–27.

Van Bellegem, Sébastien, and Rainer von Sachs. 2004. “Forecasting Economic Time Series with Unconditional Time-Varying Variance.” *International Journal of*

Forecasting 20 (4): 611–27.
Van Nostrand, Jeanine L., Margot E. Bowen, Hannes Vogel, Maria Barna, and Laura D. Attardi. 2017. "The p53 Family Members Have Distinct Roles during Mammalian Embryonic Development." *Cell Death and Differentiation* 24 (4): 575–79.
Viola, A., and A. Lanzavecchia. 1996. "T Cell Activation Determined by T Cell Receptor Number and Tunable Thresholds." *Science*. https://doi.org/10.1126/science.273.5271.104.
Viré, Emmanuelle, Carmen Brenner, Rachel Deplus, Loïc Blanchon, Mario Fraga, Céline Didelot, Lluis Morey, et al. 2006. "The Polycomb Group Protein EZH2 Directly Controls DNA Methylation." *Nature* 439 (7078): 871–74.
Visvader, Jane E. 2011. "Cells of Origin in Cancer." *Nature* 469 (7330): 314–22.
Voelkl, Bernhard, Naomi S. Altman, Anders Forsman, Wolfgang Forstmeier, Jessica Gurevitch, Ivana Jaric, Natasha A. Karp, et al. 2020. "Reproducibility of Animal Research in Light of Biological Variation." *Nature Reviews. Neuroscience* 21 (7): 384–93.
Vogelstein, Bert, Nickolas Papadopoulos, Victor E. Velculescu, Shibin Zhou, Luis A. Diaz Jr, and Kenneth W. Kinzler. 2013. "Cancer Genome Landscapes." *Science* 339 (6127): 1546–58.
Vousden, Karen H., and David P. Lane. 2007. "p53 in Health and Disease." *Nature Reviews. Molecular Cell Biology* 8 (4): 275–83.
Wager-Smith, Karen, and Steve A. Kay. 2000. "Circadian Rhythm Genetics: From Flies to Mice to Humans." *Nature Genetics* 26 (1): 23–27.
Waldman, Alex D., Jill M. Fritz, and Michael J. Lenardo. 2020. "A Guide to Cancer Immunotherapy: From T Cell Basic Science to Clinical Practice." *Nature Reviews. Immunology* 20 (11): 651–68.
Wang, Cheng, Rong Yin, Juncheng Dai, Yayun Gu, Shaohua Cui, Hongxia Ma, Zhihong Zhang, et al. 2018. "Whole-Genome Sequencing Reveals Genomic Signatures Associated with the Inflammatory Microenvironments in Chinese NSCLC Patients." *Nature Communications* 9 (1): 2054.
Wang, Dingzhi, and Raymond N. DuBois. 2015. "Immunosuppression Associated with Chronic Inflammation in the Tumor Microenvironment." *Carcinogenesis* 36 (10): 1085–93.
Wang, Jin, Kun Zhang, Li Xu, and Erkang Wang. 2011. "Quantifying the Waddington Landscape and Biological Paths for Development and Differentiation." *Proceedings of the National Academy of Sciences of the United States of America* 108 (20): 8257–62.
Wang, Lily Yan, Panwen Wang, Mulin Jun Li, Jing Qin, Xiaowo Wang, Michael Q. Zhang, and Junwen Wang. 2011. "EpiRegNet: Constructing Epigenetic Regulatory Network from High Throughput Gene Expression Data for Humans." *Epigenetics: Official Journal of the DNA Methylation Society* 6 (12): 1505–12.
Wang, Ting, Suoli Han, and Ge Du. 2021. "S100A6 Represses Calu-6 Lung Cancer Cells Growth via Inhibiting Cell Proliferation, Migration, Invasion and Enhancing Apoptosis." *Cell Biochemistry and Function*, May. https://doi.org/10.1002/cbf.3639.
Wang, Xiliang, Yao He, Qiming Zhang, Xianwen Ren, and Zemin Zhang. 2021. "Direct Comparative Analyses of 10X Genomics Chromium and Smart-seq2." *Genomics, Proteomics & Bioinformatics*, March. https://doi.org/10.1016/j.gpb.2020.02.005.

Wang, Xuyi, Evan R. Simpson, and Kristy A. Brown. 2015. “p53: Protection against Tumor Growth beyond Effects on Cell Cycle and Apoptosis.” *Cancer Research* 75 (23): 5001–7.

Wang, Yifan, Weiye Deng, Nan Li, Shinya Neri, Amrish Sharma, Wen Jiang, and Steven H. Lin. 2018. “Combining Immunotherapy and Radiotherapy for Cancer Treatment: Current Challenges and Future Directions.” *Frontiers in Pharmacology* 9 (March): 185.

Wang, Yong, and Nicholas E. Navin. 2015. “Advances and Applications of Single-Cell Sequencing Technologies.” *Molecular Cell* 58 (4): 598–609.

Wang, Zheng, Xiang Li, and Ying Xu. 2021. “Prediction of Overall Survival and Response to Immune Checkpoint Inhibitors: An Immune-Related Signature for Gastric Cancer.” *Translational Oncology* 14 (6): 101082.

Warburton, David, Saverio Bellusci, Stijn De Langhe, Pierre-Marie Del Moral, Vincent Fleury, Arnaud Mailleux, Denise Tefft, Mathieu Unbekandt, Kasper Wang, and Wei Shi. 2005. “Molecular Mechanisms of Early Lung Specification and Branching Morphogenesis.” *Pediatric Research* 57 (5 Pt 2): 26R – 37R.

Warburton, David, Ahmed El-Hashash, Gianni Carraro, Caterina Tiozzo, Frederic Sala, Orquidea Rogers, Stijn De Langhe, et al. 2010. “Lung Organogenesis.” *Current Topics in Developmental Biology* 90: 73–158.

Ward, J. F. 1986. “Ionizing Radiation Induced DNA Damage: Identities and DNA Repair.” *Mechanisms of DNA Damage and Repair*. https://doi.org/10.1007/978-1-4615-9462-8_14.

Warren, Iyan, Mitchell Maloy, Daniel Guiggey, Ogechi Ogoke, Theodore Groth, Tala Mon, Saber Meamardoost, et al. 2020. “Foxa1 and Foxa2 Together Control Developmental Gene Regulatory Networks, and Differentiation Genes, in Both Human Stem-Cell Derived Liver Progenitors and in a Human Liver Cell Line: Evidence of a Collapse of Human Liver Differentiation.” *bioRxiv*. https://doi.org/10.1101/2020.06.01.128108.

Waters, Andrew M., and Channing J. Der. 2018. “KRAS: The Critical Driver and Therapeutic Target for Pancreatic Cancer.” *Cold Spring Harbor Perspectives in Medicine* 8 (9). https://doi.org/10.1101/cshperspect.a031435.

Weaver, M., N. R. Dunn, and B. L. Hogan. 2000. “Bmp4 and Fgf10 Play Opposing Roles during Lung Bud Morphogenesis.” *Development* 127 (12): 2695–2704.

Weaver, M., J. M. Yingling, N. R. Dunn, S. Bellusci, and B. L. Hogan. 1999. “Bmp Signaling Regulates Proximal-Distal Differentiation of Endoderm in Mouse Lung Development.” *Development* 126 (18): 4005–15.

“Website.” n.d. Accessed July 18, 2021. Beauchemin, Kyle Jaymes. 2016. “Characterization of the Murine Lung Transcriptome during Pre- and Post-Natal Development in Three Inbred Strains of Laboratory Mice.” Edited by Carol J. Bult and Viravuth P. Yin. Ann Arbor, United States: The University of Maine. https://www.proquest.com/dissertations-theses/characterization-murine-lung-transcriptome-during/docview/1844967052/se-2.

Weinberg, Robert A. 2013. *The Biology of Cancer*. Garland Science.

Wei, Spencer C., Colm R. Duffy, and James P. Allison. 2018. “Fundamental Mechanisms of Immune Checkpoint Blockade Therapy.” *Cancer Discovery* 8 (9): 1069–86.

Wells, Daniel K., Marit M. van Buuren, Kristen K. Dang, Vanessa M. Hubbard-Lucey, Kathleen C. F. Sheehan, Katie M. Campbell, Andrew Lamb, et al. 2020. “Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction.” *Cell* 183 (3): 818–34.e13.

Welten, Suzanne P. M., Alexander Yermanos, Nicolas S. Baumann, Franziska Wagen, Nathalie Oetiker, Ioana Sandu, Alessandro Pedrioli, et al. 2020. “Tcf1+ Cells Are Required to Maintain the Inflationary T Cell Pool upon MCMV Infection.” *Nature Communications* 11 (1): 2295.

Westcott, Peter M. K., Kyle D. Halliwill, Minh D. To, Mamunur Rashid, Alistair G. Rust, Thomas M. Keane, Reyno Delrosario, et al. 2015. “The Mutational Landscapes of Genetic and Chemical Models of Kras-Driven Lung Cancer.” *Nature* 517 (7535): 489–92.

Wherry, E. John, Sang-Jun Ha, Susan M. Kaech, W. Nicholas Haining, Surojit Sarkar, Vandana Kalia, Shruti Subramaniam, Joseph N. Blattman, Daniel L. Barber, and Rafi Ahmed. 2007. “Molecular Signature of CD8+ T Cell Exhaustion during Chronic Viral Infection.” *Immunity* 27 (4): 670–84.

Wherry, E. John, and E. John Wherry. 2011. “T Cell Exhaustion.” *Nature Immunology*. https://doi.org/10.1038/ni.2035.

Wherry, E. John, E. John Wherry, and Makoto Kurachi. 2015. “Molecular and Cellular Insights into T Cell Exhaustion.” *Nature Reviews Immunology*. https://doi.org/10.1038/nri3862.

Whiteside, T. L. 2008. “The Tumor Microenvironment and Its Role in Promoting Tumor Growth.” *Oncogene* 27 (45): 5904–12.

Williams, Marc J., Benjamin Werner, Chris P. Barnes, Trevor A. Graham, and Andrea Sottoriva. 2016. “Identification of Neutral Tumor Evolution across Cancer Types.” *Nature Genetics* 48 (3): 238–44.

Williams, Mary C. 2003. “Alveolar Type I Cells: Molecular Phenotype and Development.” *Annual Review of Physiology* 65: 669–95.

Willis, Amy, Eun Joo Jung, Therese Wakefield, and Xinbin Chen. 2004. “Mutant p53 Exerts a Dominant Negative Effect by Preventing Wild-Type p53 from Binding to the Promoter of Its Target Genes.” *Oncogene* 23 (13): 2330–38.

Wilpe, Sandra van, Sofie H. Tolmeijer, I. Jolanda M. de Vries, Rutger H. T. Koornstra, and Niven Mehra. 2021. “LDH Isotyping for Checkpoint Inhibitor Response Prediction in Patients with Metastatic Melanoma.” *Immuno* 1 (2): 67–77.

Winslow, Monte M., Talya L. Dayton, Roel G. W. Verhaak, Caroline Kim-Kiselak, Eric L. Snyder, David M. Feldser, Diana D. Hubbard, et al. 2011. “Suppression of Lung Adenocarcinoma Progression by Nkx2-1.” *Nature* 473 (7345): 101–4.

Wong, A. J., J. M. Ruppert, S. H. Bigner, C. H. Grzeschik, P. A. Humphrey, D. S. Bigner, and B. Vogelstein. 1992. “Structural Alterations of the Epidermal Growth Factor Receptor Gene in Human Gliomas.” *Proceedings of the National Academy of Sciences of the United States of America* 89 (7): 2965–69.

Wong, G. H., and I. Clark-Lewis. 1983. “Interferon-Gamma Induces Enhanced Expression of Ia and H-2 Antigens on B Lymphoid, Macrophage, and Myeloid Cell Lines.” *The Journal of*. https://www.jimmunol.org/content/131/2/788.short?casa_token=T_AMa0yxFLUAAAAA:e5OW_sCkFazpug4MLx4WnrlFx1QueKNd9X6ZC2CGsb59adxVYIRSEKMc4o

mIu4Mo569fyGaugIZ3Nw.
Wu, Annie A., Virginia Drake, Huai-Shiuan Huang, Shihchi Chiu, and Lei Zheng. 2015. “Reprogramming the Tumor Microenvironment: Tumor-Induced Immunosuppressive Factors Paralyze T Cells.” *Oncoimmunology* 4 (7): e1016700.
Wu, Junjie, and David J. Waxman. 2018. “Immunogenic Chemotherapy: Dose and Schedule Dependence and Combination with Immunotherapy.” *Cancer Letters* 419 (April): 210–21.
Wu, Xiaotian, Hao Wu, and Zhijin Wu. 2021. “Penalized Latent Dirichlet Allocation Model in Single-Cell RNA Sequencing.” *Statistics in Biosciences*, March, 1–20.
Wu, Zhijin, and Hao Wu. 2020. “Accounting for Cell Type Hierarchy in Evaluating Single Cell RNA-Seq Clustering.” *Genome Biology* 21 (1): 1–14.
Xu, Linda W., Kevin K. H. Chow, Michael Lim, and Gordon Li. 2014. “Current Vaccine Trials in Glioblastoma: A Review.” *Journal of Immunology Research* 2014 (April): 796856.
Yang, James C. 2015. “Toxicities Associated With Adoptive T-Cell Transfer for Cancer.” *Cancer Journal* 21 (6): 506–9.
Yang, Jing, Sendurai A. Mani, Joana Liu Donaher, Sridhar Ramaswamy, Raphael A. Itzykson, Christophe Come, Pierre Savagner, Inna Gitelman, Andrea Richardson, and Robert A. Weinberg. 2004. “Twist, a Master Regulator of Morphogenesis, Plays an Essential Role in Tumor Metastasis.” *Cell* 117 (7): 927–39.
Yang, Yang, Hongmei Zheng, Yuting Zhan, and Songqing Fan. 2019. “An Emerging Tumor Invasion Mechanism about the Collective Cell Migration.” *American Journal of Translational Research* 11 (9): 5301–12.
Yang, Y., D. L. Smith, K. D. Keating, D. B. Allison, and T. R. Nagy. 2014. “Variations in Body Weight, Food Intake and Body Composition after Long-Term High-Fat Diet Feeding in C57BL/6J Mice.” *Obesity* 22 (10). https://doi.org/10.1002/oby.20811.
Yasukawa, Motoaki, Noriyoshi Sawabata, Takeshi Kawaguchi, Norikazu Kawai, Tokiko Nakai, Chiho Ohbayashi, and Shigeki Taniguchi. 2018. “Histological Grade: Analysis of Prognosis of Non-Small Cell Lung Cancer After Complete Resection.” *In Vivo* 32 (6): 1505–12.
Yeung, K. Y., and W. L. Ruzzo. 2001. “Principal Component Analysis for Clustering Gene Expression Data.” *Bioinformatics* 17 (9): 763–74.
Yewdell, Jonathan W. 2006. “Confronting Complexity: Real-World Immunodominance in Antiviral CD8+ T Cell Responses.” *Immunity* 25 (4): 533–43.
Youngblood, Ben, J. Scott Hale, Haydn T. Kissick, Eunseon Ahn, Xiaojin Xu, Andreas Wieland, Koichi Araki, et al. 2017. “Effector CD8 T Cells Dedifferentiate into Long-Lived Memory Cells.” *Nature* 552 (7685): 404–9.
Yuan, B., C. Li, S. Kimura, R. T. Engelhardt, B. R. Smith, and P. Minoo. 2000. “Inhibition of Distal Lung Morphogenesis in Nkx2.1(-/-) Embryos.” *Developmental Dynamics: An Official Publication of the American Association of Anatomists* 217 (2): 180–90.
Yu, Jian, and Lin Zhang. 2003. “No PUMA, No Death: Implications for p53-Dependent Apoptosis.” *Cancer Cell* 4 (4): 248–49.
Zamprogno, Pauline, Simon Wüthrich, Sven Achenbach, Giuditta Thoma, Janick D. Stucki, Nina Hobi, Nicole Schneider-Daum, et al. 2021. “Second-Generation Lung-on-a-Chip with an Array of Stretchable Alveoli Made with a Biological Membrane.” *Communications Biology* 4 (1): 1–10.

Zappia, Luke, Belinda Phipson, and Alicia Oshlack. 2018. “Exploring the Single-Cell RNA-Seq Analysis Landscape with the scRNA-Tools Database.” *PLoS Computational Biology* 14 (6): e1006245.

Zhang, Allen W., Andrew McPherson, Katy Milne, David R. Kroeger, Phineas T. Hamilton, Alex Miranda, Tyler Funnell, et al. 2018. “Interfaces of Malignant and Immunologic Clonal Dynamics in Ovarian Cancer.” *Cell* 173 (7): 1755–69.e22.

Zhang, Huihua, Xiaojia Chen, Jiakang Wang, Wenhua Guang, Wei Han, Hang Zhang, Xuan Tan, and Yong Gu. 2014. “EGR1 Decreases the Malignancy of Human Non-Small Cell Lung Carcinoma by Regulating KRT18 Expression.” *Scientific Reports* 4 (July): 5416.

Zhang, Jiaxue, Tong Lyu, Yaming Cao, and Hui Feng. 2021. “Role of TCF-1 in Differentiation, Exhaustion, and Memory of CD8+ T Cells: A Review.” *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* 35 (5): e21549.

Zhang, Jingfeng, Sifeng Hu, and Yansen Li. 2019. “KRT18 Is Correlated with the Malignant Status and Acts as an Oncogene in Colorectal Cancer.” *Bioscience Reports* 39 (8). https://doi.org/10.1042/BSR20190884.

Zhang, Lin, Jose R. Conejo-Garcia, Dionyssios Katsaros, Phyllis A. Gimotty, Marco Massobrio, Giorgia Regnani, Antonis Makrigiannakis, et al. 2003. “Intratumoral T Cells, Recurrence, and Survival in Epithelial Ovarian Cancer.” *The New England Journal of Medicine* 348 (3): 203–13.

Zhang, Sheng, Hongjun Zhang, and Jiandong Zhao. 2009. “The Role of CD4 T Cell Help for CD8 CTL Activation.” *Biochemical and Biophysical Research Communications*. https://doi.org/10.1016/j.bbrc.2009.04.134.

Zhang, Zhen, Shasha Liu, Bin Zhang, Liang Qiao, Yi Zhang, and Yi Zhang. 2020. “T Cell Dysfunction and Exhaustion in Cancer.” *Frontiers in Cell and Developmental Biology* 8 (February): 17.

Zhang, Zhen, Kim Newton, Sarah K. Kummerfeld, Joshua Webster, Donald S. Kirkpatrick, Lilian Phu, Jeffrey Eastham-Anderson, et al. 2017. “Transcription Factor Etv5 Is Essential for the Maintenance of Alveolar Type II Cells.” *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1621177114.

Zhang, Zhen, Jamie M. Verheyden, John A. Hassell, and Xin Sun. 2009. “FGF-Regulated Etv Genes Are Essential for Repressing Shh Expression in Mouse Limb Buds.” *Developmental Cell* 16 (4): 607–13.

Zhao, Junfei, Feixiong Cheng, and Zhongming Zhao. 2017. “Tissue specific Signaling Networks Rewired by Major Somatic Mutations in Human Cancer Revealed by Proteome-Wide Discovery.” *Cancer Research* 77 (11): 2810.

Zheng, Grace X. Y., Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, et al. 2017. “Massively Parallel Digital Transcriptional Profiling of Single Cells.” *Nature Communications* 8 (January): 14049.

Zhou, Xinyuan, Shuyang Yu, Dong-Mei Zhao, John T. Harty, Vladimir P. Badovinac, and Hai-Hui Xue. 2010. “Differentiation and Persistence of Memory CD8(+) T Cells Depend on T Cell Factor 1.” *Immunity* 33 (2): 229–40.