A broadened HLA ligandome uncovers new immunotherapy targets for pancreatic cancer
+
A prime editor mouse to model a broad spectrum of somatic mutations

By
Zackery A. Ely

B.A. Molecular and Cellular Biology
Minors: Chemistry, Literary Studies
Vanderbilt University, 2018

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2023

© 2023 Massachusetts Institute of Technology. All rights reserved.

Signature of author: ____
Zackery Ely
Department of Biology
September 6, 2022

Certified by: ____
Tyler Jacks
Professor of Biology
Thesis Supervisor

Accepted by: ____
Mary Gehring
Professor of Biology
Chair, Biology Graduate Committee

A broadened HLA ligandome uncovers new immunotherapy targets for pancreatic cancer
+
A prime editor mouse to model a broad spectrum of somatic mutations

by
Zackery A. Ely

Submitted to the Department of Biology on September 6, 2022 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in Biology

ABSTRACT

Pancreatic cancer is a lethal malignancy recalcitrant to immune checkpoint blockade and other immunotherapies. A subset of tumors is computationally predicted to harbor potentially immunogenic peptides for MHC class I (MHC-I) presentation, but the nature, expression, and immunogenicity of these peptides has yet to be determined. The only prior study of the pancreatic cancer immunopeptidome focused on profiling MHC-I-associated peptides (MAPs) from canonical proteins in bulk tumor samples; however, non-malignant cell populations comprise most of the pancreatic tumor mass, obscuring the identity of MAPs that derive specifically from cancer cells. In the second chapter of this thesis, I resolve this challenge through extensive profiling of patient-derived organoids with whole-genome sequencing, RNA sequencing, and immunopeptidomics. These data enable a proteogenomics approach that tailors MAP identification to each individual patient sample. Harnessing this platform, my colleagues and I uncovered a diverse cohort of MAPs derived from somatic mutations and transcript isoforms that are functionally unexpressed in most or all healthy tissues. These include MAPs derived from novel, unannotated open reading frames (nuORFs) present within long noncoding RNAs, processed transcripts, and 5' and 3' untranslated regions. We found that cytotoxic T cells specific to nuORF-derived MAPs can be readily generated from peripheral blood mononuclear cells of healthy donor individuals. This result highlights the immunogenicity of nuORF-derived MAPs and establishes them as promising targets for immunotherapies in pancreatic cancer.

In Chapter 3, I report the development of a genetically engineered mouse model (GEMM) for performing prime editing *in vivo*. This system represents a rapid alternative to traditional cancer mouse models, which often take months or years to develop. Through a Cre-inducible prime editor enzyme encoded in the mouse germline, prime editor GEMMs can mediate rapid and precise engineering of most cancer mutations, including many that are challenging or infeasible to achieve with other CRISPR technology. We demonstrate the utility of this system by mediating secondary *Kras* mutations and common *Trp53* hotspot mutations in model-derived pancreatic organoids. Finally, we model lung and pancreatic cancer *in vivo* using lentiviral delivery of prime editing guide RNAs or orthotopic transplantation of prime edited organoids. We anticipate that prime editing GEMMs will accelerate preclinical functional studies of cancer-associated alleles that are challenging to model by traditional approaches.

Thesis Advisor: Tyler Jacks
Title: Professor of Biology

CURRICULUM VITAE

Zackery A. Ely

EDUCATION

MASSACHUSETTS INSTITUTE OF TECHNOLOGY — CAMBRIDGE, MA

Biology, Ph.D. Candidate — **August 2018-September 2022**

- Graduate student in the laboratory of Dr. Tyler Jacks
- Thesis project: "A broadened HLA ligandome uncovers new immunotherapy targets for pancreatic cancer"
- National Institutes of Health F31 Predoctoral Research Fellowship
- Courses: Principles of Drug Development, Strategies in Drug Discovery, Science and Business of Biotechnology
- Activities: MIT Ashdown House Committee Chair, teaching assistant (The Hallmarks of Cancer), undergraduate research mentor

VANDERBILT UNIVERSITY — NASHVILLE, TN

Molecular and Cellular Biology, B.S. — **January 2016-May 2018**

- Phi Beta Kappa Honor Society
- Honors Thesis: "Evolutionary and Structural Analysis of Galectin Genes Involved in Immunity"
- Minors in Literary Studies and Chemistry
- Activities: Undergraduate research, Vanderbilt Student Government (Chief Justice), volunteer work in Accra, Ghana

UNIVERSITY OF ROCHESTER — ROCHESTER, NY

Biochemistry — **January 2014-December 2015**

- Activities: Community Service Network (President), NROTC

RESEARCH EXPERIENCE

KOCH INSTITUTE FOR INTEGRATIVE CANCER RESEARCH — CAMBRIDGE, MA

Graduate Student, Jacks Lab — **July 2019-Present**

- Research areas: cancer immunotherapy, genome editing, computational genomics, genetically engineered mouse models
- Constructed CRISPR-based mouse model as a preclinical system for rapid modeling of cancer driver mutations and testing of targeted therapies
- Developed a computational pipeline to discover and target novel sources of cancer antigens in patients with pancreatic cancer
- Lead collaboration with laboratories at the Broad Institute and Dana-Farber Cancer Institute to generate T cells specific to antigens in pancreatic cancer
- Profiled T cell exhaustion programs in pancreatic cancer through single-cell RNA-Seq analysis of tumor-infiltrating T cells
- Laboratory experience: CRISPR-mediated gene-editing, prime and base editing, organoid derivation and culture, *ex vivo* activation and expansion of antigen-specific T cells, *in vivo* cancer modeling, ELISPOT assays, molecular cloning, microscopy, etc.
- Programming/computational experience: UNIX, R, multimodal single-cell RNA-Seq data analysis, whole-genome and RNA sequencing analysis, somatic variant detection, neoantigen prediction, etc.

VANDERBILT UNIVERSITY — NASHVILLE, TN

Undergraduate Researcher, Rokas Lab — **May 2016-October 2018**

- Research areas: computational genomics, structural biology, phylogenetics
- Studied evolutionary history of galectin genes across mammals and humans using computational genomics and phylogenetics
- Applied population genomics and structural modeling to model a novel Galectin-8 protein structure encoded by genetic variants linked to autoimmune disease
- Programming/computational experience: protein structural modeling (Rosetta), population-level genome sequence analyses, multiple sequence alignment and analysis, phylogenetic tree reconstruction, evolutionary rate analysis, etc.

RESEARCH PUBLICATIONS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY CAMBRIDGE, MA

- **Ely Z.A.**, Mathey-Andrews N., Naranjo S., Rideout W. M., Cabana C.M., Holland K., *et al*. (2022). A prime editor mouse model for modeling a broad spectrum of somatic mutations. *Manuscript in review*.
- **Ely Z.A.,** Freed-Pastor W.A., Verzani E., Kulstad Z., Abelin J., Clauser K., *et al*. (2022). A broadened HLA ligandome uncovers new immunotherapy targets for pancreatic cancer. *Manuscript in preparation*.
- Freed-Pastor W.A., Lambert L.J., **Ely Z.A.,** Pattada N. B., Bhutkar A., Eng G., *et al*. (2021). The CD155/TIGIT axis promotes and maintains immune evasion in neoantigen-expressing pancreatic cancer. *Cancer Cell.* 39: 1342-1360.
- Westcott P.M.K., Sacks N., Schenkel J.M., **Ely Z.A**., Smith O., Hauck H., *et al*. (2021). Low neoantigen expression and poor T cell priming underlie early immune escape in cancer. *Nature Cancer*. 2: 1071-1085.
- Westcott P.M.K., Muyas F., Smith O., Hauck H., Sacks, N., **Ely Z.A.**, *et al*. (2021). Mismatch repair deficiency is not sufficient to increase tumor immunogenicity. *Manuscript submitted for review*.
- Canner D., Schenkel J.M., Park J. K, **Ely Z.A.**, Hillman M., Herbst R.H., *et al*. (2021). CRISPR/Cas9 screen in a model of lung adenocarcinoma identifies *Arid1a* as a novel regulator of T cell dysfunction. *Manuscript submitted*.
- Hwang W.L., Jagadeesh K.A., Guo J.A., Hoffman H.I., Yadollahpour P., Reeves J., [*et al*., including **Ely Z.A.**] (2021). Single-nucleus and spatial multi-omics of frozen pancreatic cancer reveals multi-compartment reprogramming and enrichment of a neuronal-like phenotype after neoadjuvant treatment. *Nature Genetics*. In press.
- **Ely Z.A.**, Moon J.Y., Sliwoski G.R., Sangha A.K., Shen X., Labella A.L., *et al*. (2019). The impact of natural selection on the evolution and function of placentally expressed galectins. *Genome Biology and Evolution*. 11: 2574-2592.

AWARDS AND HONORS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY CAMBRIDGE, MA

- 2022 National Institutes of Health F31 Ruth L. Kirschstein Predoctoral Fellowship
- 2021-2022 Ludwig Koch Institute Fellowship
- 2020 Margaret A. Cunningham Immune Mechanisms of Cancer Research Fellowship

VANDERBILT UNIVERSITY NASHVILLE, TN

- 2018 Outstanding Undergraduate Research Award in the Biological Sciences
    - Awarded annually to a single senior undergraduate for outstanding research in the department
- 2018 Phi Beta Kappa Tellinghuisen Award
    - Awarded annually to a single individual out of a graduating class of 1,500 for outstanding undergraduate research
- 2017-2018 Beckman Scholar Award
    - Awarded annually to approximately 30 national scholars ($26,000 in research support funds)
- 2018 Beckman Symposium Invited Speaker
    - Selected out of 60 national scholars to give an oral presentation to over 100 scientists in attendance
- 2018 Society for Molecular Biology and Evolution Annual Conference
    - Selected to present poster for annual conference
- 2017 Best Presentation Award at Vanderbilt Undergraduate Research Fair (>140 participants)
- 2016 Nichols Humanitarian Fund Award
    - Scholarship supporting a volunteer project in Accra, Ghana (nonprofit: Unite for Sight)

ABSTRACTS AND PRESENTATIONS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY CAMBRIDGE, MA

- **Ely Z.A**., Mathey-Andrews N., Naranjo S., Rideout W. M., Cabana C.M., Holland K., *et al*. A prime editor mouse for modeling a broad spectrum of somatic mutations. Invited Poster Presentation at the Vallee Foundation Research Symposium. Cascais, Portugal. June 2022.

- Westcott P.M.K., Remolar F.M., Smith O., Hauck H., Sacks N.J., **Ely Z.A.,** *et al*. (2022). Mismatch repair deficiency is not sufficient to increase tumor immunogenicity. Cancer Immunol Res. (10) (1 Supplement). P023.
- **Ely Z.A.***, Guo J.A.*, Shiau C.*, Chung H., Magee E., Delorey T.M., *et al*. (2021). Developmental Lineages and Microenvironmental Architecture of Pancreatic Neuroendocrine Tumors at a Single-Cell Resolution. 52nd Annual Meeting of the American Pancreatic Association. *These authors contributed equally.
- Freed-Pastor W. A., Lambert, L., **Ely Z.A.**, Pattada N.B., Bhutkar A., Jaeger A.M., *et al*. (2021). Functional interrogation of immune escape in neoantigen-expressing pancreatic cancer identifies a critical role for the CD155/TIGIT axis. Cancer Res. (6) (9 Supplement). PO-063.
- Westcott P.M.K., Sacks N., Smith O., Schenkel J., **Ely Z.A.**, *et al*. (2020). T cell antigen expression levels govern progression and therapy response in a novel model of colon cancer. Cancer Res. (80) (16 Supplement). 3447.
- **Ely Z.A.** Discovery of shared cancer antigens guided by bioinformatics. Oral Presentation at the Koch Institute Immunotherapy Supergroup Meeting. October 2021.
- **Ely Z.A.** Evolutionary and Structural Analysis of Galectin Genes Involved in Immunity. Invited Oral Presentation at the Beckman Symposium. Irvine, California. August 2018.

ACKNOWLEDGMENTS

I am fortunate to have met many excellent colleagues, collaborators, and friends over the past three years. Dozens of people worked with me in many ways large and small.

I will start by acknowledging two of my main collaborators: Will Freed-Pastor and Nick Mathey-Andrews.

Will has been a great mentor, collaborator, and friend since my first day in the lab. I learned most of what I know about pancreatic cancer through our collaboration and his mentorship. I look forward to continuing our work together.

Nick contributed immensely to the project described in Chapter 3. He has been a great thought partner, a good friend, and the best collaborator one could hope for.

I am grateful to several postdoctoral scientists who all took time to teach me the science of cancer biology: William Hwang, Peter Westcott, Jason Schenkel, Megan Burger, Carla Concepcion, and Alex Jaeger.

Two other mentors, Santiago Naranjo and James Angstman, taught me the science of CRISPR-based genome editing and the art of using a pipette.

I am thankful to the friends I have made in graduate school: Christina Cabana, Nick Mathey-Andrews, Sean-Luc Shanahan, Vidit Bhandarkar, Grissel Cervantes Jaramillo, Colin Fowler, and Emma Dawson. I am lucky to have these bright people as my peers and friends.

I am extremely grateful to Kim Mercer, whose efforts with the prime editor model made Chapter 3 possible. She has been a kind friend and colleague throughout my time in the lab.

Britt Rideout supplied indispensable expertise and skills during the development of the prime editor model.

Jennifer Su provided important technical skills for some of the work described in Chapter 2.

I also acknowledge several excellent collaborators whom I have worked with from outside lab: Greg Newby, Samuel Gould, Eva Verzani, Jenn Abelin, Susan Klaeger, Steven Carr, David Liu, and Karl Clauser. These scientists contributed significantly to the work described in Chapter 2 and Chapter 3.

Francisco Sánchez-Rivera is another important collaborator. His pSECC paper was the first Jacks Lab paper I ever read, and it inspired many of my research interests. He has been a great mentor to me in the last year as we worked to complete the manuscript described in Chapter 3.

Judy Teixeira and Karen Yee supported my career in more ways than I could count. I appreciate Karen for all the time and effort she spent reviewing my fellowship applications and

coordinating certain technical aspects of my work. I am grateful to Judy for her constant patience, administrative support, bright personality, and friendship.

Kate Anderson and Margaret Magendantz also provided important support to my work and the lab at large.

I acknowledge Doug Lauffenburger and Stefani Spranger for serving on my thesis committee for the past three years. I am also grateful to Catherine Wu for agreeing to serve with them during my thesis defense.

I have worked with two excellent students who contributed significantly to both of projects: Katie Holland and Miles Agus.

None of the work and learnings I report here would have been possible without my advisor, Tyler Jacks. Working in his lab has been the most fulfilling experience of my life. Tyler has been an insightful advisor, a supportive mentor, and an inspiring role model since the day I started my rotation in the Jacks Lab. I cannot begin to express the extent of my gratitude for him and the opportunity he gave me. Thank you for everything, Tyler.

The Jacks Lab at large is a community I am enormously proud to be a part of. I am grateful to every technician, graduate student, postdoc, and administrator who made the Jacks Lab a ceaselessly stimulating and always exciting place for me to be. I have sincerely enjoyed going to work every day (and night!) for the past three years.

Beyond the lab, I acknowledge Adam Frank. His Cosmic Origins of Life was the best course I ever took. I'll never forget the hours spent learning about astronomy in his office.

I am also thankful to teachers who encouraged me early on, well before college: Mark Jennings, Donna Spain, Julie Holt, and Melanie Fangman.

I will always be grateful for Antonis Rokas, who gave me my first research opportunity and made graduate school possible for me. From him I learned the science of computational genomics, which has fueled all my research ever since. He was an excellent mentor to me as an undergraduate.

I acknowledge Xing-Xing Shen and Jacob Steenwyk, also from the Rokas Lab. They were early role models for me as I started along this career.

I also acknowledge Jens Meiler and Tony Capra, professors who supported my research career at Vanderbilt.

I also acknowledge my oldest friend, Kel Kellerman, and his dad, Keith. Kel has been a steadfast friend throughout my life.

I am grateful to my other best friend, Nick Morgante, and his dad, Dan. Nick has inspired me more than anyone since the day we joined forces in Crosby 024. His wisdom and constant friendship have seen me through the best and worst of times.

My thanks to T.A. Morris, another great friend and mentor.

My uncle, Don Ely, encouraged a love for science for as long as I can remember. He inspired me to pursue a life of learning. I am grateful for everything he has taught me.

I thank my brother, Josh Ely. He has always supported me with laughter, friendship, and wisdom. I look forward to watching the great things he will continue to do with his life and career.

Mom and Dad, thank you for making all of this possible.

I am deeply grateful to my wife, Morgane Ledoux. She made everything about the past four years better and more meaningful than they would have been without her. Her patience, wisdom, and humor are constant sources of joy and encouragement to me.

I dedicate this thesis to my grandmother, Judith Ely. Judy was one my closest friends. She was caring, compassionate, and unconditionally kind. Above all, she was a ceaseless optimist. She relentlessly believed in a brighter future, both for everyone she loved and for people around the world. Her memory inspires me, and it reminds me that the pursuit of new knowledge is the key to unlocking the better future she always believed in.

TABLE OF CONTENTS

ABSTRACT 2

CURRICULUM VITAE 3

ACKNOWLEDGMENTS 6

Table of Contents 9

CHAPTER 1: Introduction 11

Chapter 1, Part 1: Enabling the immune system to target pancreatic cancer. 11

Chapter 1, Part 2: Modeling the cancer genome with genome editing technology. 50

References 71

CHAPTER 2: A broadened HLA ligandome uncovers new immunotherapy targets for pancreatic cancer 81

Abstract 82

Introduction 83

Results 85

Discussion 105

Supplementary Figures and Legends 109

Methods 122

References 141

CHAPTER 3: A prime editor mouse to model a broad spectrum of somatic mutations *in vivo* 145

Abstract 146

Introduction 146

Results 149

Discussion 167

Supplementary Figures and Legends 173

Methods 192

References 209

CHAPTER 4: Discussion 214

References 224

APPENDIX I: The CD155/TIGIT axis promotes and maintains immune evasion in neoantigen-expressing pancreatic cancer 226

Abstract 227

Introduction 228

Results 229
Discussion 259
Methods.. 267
References 292

APPENDIX II: Low neoantigen expression and poor T cell priming underlie early immune escape in cancer 299

Abstract 300
Introduction.. 300
Results 302
Discussion 322
Methods.. 326
References 341

CHAPTER 1: INTRODUCTION

Chapter 1, Part 0.

I have been fortunate to explore several research areas during my graduate studies in the Jacks Lab. After joining the lab, I initially focused on modeling cancer mutations through applying genome editing technology, but I soon developed an additional interest in cancer immunology. This chapter will introduce projects in both areas. Part 1 will introduce a background for pancreatic cancer and immunotherapy, and Part 2 will describe a history of cancer mouse models and genome editing techniques. These projects inhabit different research areas, and thus Parts 1 and 2 will separately motivate distinct projects described in Chapters 2 and 3. Though disparate, both projects share a focus on pancreatic cancer, a complex, deadly, and fascinating disease. This thesis will describe efforts to both model the disease and target it with immunotherapy.

Chapter 1, Part 1: Enabling the immune system to target pancreatic cancer.

Diagnosis, detection, and treatment of the disease.

Of the nearly 500,000 patients diagnosed with pancreatic cancer worldwide each year, [1] more than 80% will die within a year of diagnosis [2]. More than 90% will die within two. Most patients receive their diagnosis after the disease has already metastasized to distant organs [3], precluding the possibility of curative surgery. Prior to this late stage, patients are often asymptomatic, or they exhibit only nonspecific symptoms, such as weight loss or nausea. Without obvious, specific symptoms, it is challenging to identify the disease early. Early diagnosis is further hindered by a lack of cost-effective, noninvasive screening techniques [4].

Despite these challenges, a subset of patients (~10-20%) are diagnosed without detectable metastases [5]. Such cases of localized pancreatic cancer may be eligible for the Whipple procedure, a sometimes curative, oftentimes debilitating surgery that resects part of the pancreas.

This procedure is currently the only curative treatment for the disease, yet more than 80% of patients who undergo it will relapse and die within five years [5]. Nevertheless, the ~20% five-year survival rate among surgically resected patients motivates many researchers to improve early detection of the disease and increase the fraction of patients eligible for surgery.

Early detection methods for pancreatic cancer include endoscopy, clinical imaging methods like computed tomography, and assays for blood-based biomarkers, including DNA methylation patterns, the cancer-associated protein mesothelin, and mutant DNA shed by tumor cells [5,6]. Unfortunately, these approaches are neither broadly applicable nor sufficiently sensitive. For example, plasma-based detection of *KRAS* mutations, which occur in almost all pancreatic tumors, identified only ~45% of patients from a cohort of those with surgically resectable disease [6]. Similarly, methylation-based assays are currently limited to a sensitivity of ~63% for early-stage pancreatic cancer [7]; however, these assays exhibit a remarkable specificity >99%. This could enable broad application without incurring a significant tradeoff in false-positive diagnoses, a critical consideration for early detection methods. Continued efforts from biotechnology companies that incorporate multi-variable readouts may someday yield an assay sufficiently sensitive and specific to motivate broader screening for pancreatic cancer.

Early detection has reduced mortality from breast and colon cancer [8], but some researchers argue that early detection may not sway the future prognoses of pancreatic cancer. These doubts rest upon two disparate, seminal discoveries from human tumor genome sequencing and mouse models. In mice, Rhim *et al.* (2012) developed a model where pancreatic cancer cells are fluorescently labeled, enabling tracking of the disease and its metastatic spread [4]. These investigators found that cancer cells can migrate beyond the pancreas early in tumor development, invading the bloodstream and seeding distant organs before the primary tumor is detectable [4]. In patients, Notta *et al*. (2016) used cancer genome sequencing to infer the evolutionary trajectory of tumors. They found that severe chromosomal instability can facilitate the simultaneous loss of multiple pancreatic tumor suppressor genes, potentially enabling premalignant lesions to immediately acquire highly invasive properties in a subset of patients [9]. Both studies support a rapid model of pancreatic cancer progression and predict that only a small window may exist between the formation of an observable lesion and its subsequent metastatic spread.

These studies conflict with the traditional paradigm of pancreatic cancer evolution, which proposes a stepwise evolution that spans several years, even up to a decade [10]. In this model, formalized by Hruban *et al*. in 2000, initial activation of the *KRAS* oncogene triggers the formation of early-stage neoplasms. This is followed by loss of *CDKN2A* in low-grade neoplastic lesions and then deletion of tumor suppressor genes *TP53* and/or *SMAD4* in high-grade lesions that then progress to adenocarcinoma [11]. Molecular analyses of histologically defined precursor lesions supported this model, with several studies reporting *KRAS*-activating mutations across all grades and others identifying inactivating mutations in tumor suppressor genes only in later-stage lesions [11–13].

While this model remains a popular paradigm, the characterization of whole pancreatic cancer genomes by Notta *et al*. provides some nuance. The stepwise evolution of pancreatic cancer probably at least co-exists with cases of the ‘catastrophic’ model of tumor evolution. In this model, simultaneous driver gene alterations enable rapid transformation of early pancreatic neoplasms into invasive carcinomas. Which model predominates remains an open question, and its answer has important implications for the feasibility of early detection for pancreatic cancer.

If the disease is not diagnosed early, then it must be treated late. But nearly every class of anti-cancer treatments has been trialed in metastatic pancreatic cancer to little avail. Despite decades of clinical studies, no treatment other than surgery yielded a significant clinical benefit until 1997 [14]. In a Phase III clinical trial, gemcitabine, a nucleoside analog inhibiting DNA synthesis, met clinical endpoints, including pain reduction and prevention of weight loss in 23.8% of treated patients. While the median survival rate in gemcitabine-treated patients remained shorter than six months, gemcitabine became the standard of care for metastatic pancreatic cancer.

It took another fourteen years for gemcitabine to be displaced as standard first-line therapy. In 2011, French clinicians demonstrated that combination chemotherapy including folinic acid (leucovorin), 5-fluorouracil, irinotecan and oxaliplatin, collectively termed FOLFIRINOX, nearly doubled the overall survival of patients relative to gemcitabine [15]. However, over 50% of patients treated with FOLFIRINOX experience debilitating grade 3 or grade 4 adverse events, including neutropenia, anemia, fatigue, and vomiting. FOLFIRINOX has thus been primarily limited to patients younger than 76 years old, who can better tolerate the treatment and its considerable side effects.

Many other chemotherapeutic regimens, targeted therapies, radiotherapies, and immunotherapies have been tested, but all have been met with dismal results [5]. There is the occasional success, such as a 6% increase in one-year survival when gemcitabine is combined with the epidermal growth factor inhibitor, erlotinib [16], but no therapy has outperformed FOLFIRINOX without significant toxicity. And modern treatments rarely extend life beyond a couple months compared to treatment in the pre-gemcitabine era. For most patients today, pancreatic cancer remains a death sentence.

Pancreatic cancer treatment has failed for several reasons. Many researchers point to pancreatic cancer's dense tumor stroma, a notorious hallmark of the disease. Stromal cells, such as fibroblasts, uniquely dominate pancreatic tumors, often comprising more than 80% of the entire tumor volume [17]. This stroma is highly desmoplastic, featuring thick networks of collagen and extracellular matrix deposited by cancer-associated fibroblasts. The stroma represents a physical barrier for drug entry in animal models of pancreatic cancer, creating a delivery challenge thought to contribute to the lack of efficacy of many drugs in patients [5]. To better understand the challenges and opportunities for drug delivery, the next section will explore the origins of pancreatic tumors and their microenvironment.

Cellular origins of the disease

The pancreas resides deep within the abdomen, behind the stomach. The pancreatic duct traverses the body of the pancreas to its head, where it joins with the common bile duct and empties into the duodenum. The pancreatic duct is the primary vessel of the so-called pancreatic juice, a cocktail of enzymes and bicarbonate secreted by pancreatic exocrine cells to facilitate digestion. Most of the pancreatic juice is produced by enzyme-secreting acinar cells and bicarbonate-secreting ductal cells, the major epithelial cells of the pancreas. Acinar cells comprise >85% of the volume of the pancreas [18], and the less populous ductal cells form the lining of the pancreatic duct.

Pancreatic ductal adenocarcinoma (PDAC) is the most common and lethal manifestation of pancreatic cancer, often simply called PDAC. PDAC is distinct from a rarer subset of pancreatic neuroendocrine tumors (PNETs) that originate from the endocrine cells of the pancreas. These cells include alpha and beta cells that regulate secretion of glucagon and insulin. While a small subset of all cancers in the pancreas (< 3%) originate from endocrine cells [19], more than 90 percent derive from exocrine cell types, primarily acinar and ductal cells.

Both acinar and ductal cells are debated as a cell type of origin for human PDAC [20]. While primary PDACs localize to the duct and often resemble ductal cells histologically, transformed acinar cells can generate PDAC in genetically engineered mouse models [21]. In light of this finding, the histologic resemblance of PDAC to ductal cells is explained by a cell type transformation, acinar-ductal metaplasia (ADM), wherein acinar cells lose markers of cell identity and adopt a ductal phenotype after activation of oncogenic *KRAS* and/or in the presence of inflammation [20,22]. In humans, a subset of ADM lesions reside adjacent to *bona fide* pancreatic neoplasms bearing identical *KRAS* mutations [23], reinforcing the possibility of a lineage by which acinar cells can give rise to neoplastic cells with ductal characteristics.

Like acinar cells, ductal cells are also an empirically validated cell of origin for PDAC in mouse models [24]. For example, Boj and colleagues showed that ductal organoids, three-dimensional cell culture models derived directly from pancreatic ducts, can recapitulate classic stages of PDAC progression after transplantation into murine hosts [25]. Thus, there is substantial evidence for distinct modes of pancreatic tumorigenesis. While some studies suggest that cell of origin can dramatically affect the phenotype of the ultimate malignant tumor [20], this remains to be clinically validated.

Genomic origins of the disease

While the cell-of-origin is debated, the genetic roots are well-accepted. Like almost all cancer, PDAC is caused by somatic mutations in genes that promote cellular proliferation (oncogenes) or constrain it (tumor suppressor genes), leading to accelerated and unchecked cell division that fuels tumor growth. Since the advent of next-generation sequencing, several large-scale consortium efforts, including The Cancer Genome Atlas (TCGA), have profiled the entire genomes or exomes of thousands of cancer patients' tumors, including PDAC [17]. The mutational landscapes of many cancers, such as lung adenocarcinoma, are defined by heterogenous oncogenes and tumor suppressor genes driving cancer in different subsets of patients, but the mutational landscape of PDAC is much more constrained. Strikingly, pancreatic tumors almost uniformly harbor somatic mutations in *KRAS*, a proto-oncogene that drives many cancer types but frequently in less than 50% of patients with a given disease [26]. Mutations in other oncogenes, such as *EGFR* or *PIK3CA*, are seldom to never observed in pancreatic tumors.

Several studies have sequenced pancreatic intraepithelial neoplasia (PanIN) lesions, the benign precursors to pancreatic adenocarcinoma. In autopsy-derived samples, over 70% of PanIN lesions harbor *KRAS* mutations [23]. Other studies of primary and advanced PDAC have identified *KRAS* mutations in more than 90% of cases [17,27]. While these findings suggest *KRAS* mutations are necessary events for PDAC development, these mutations alone may be insufficient. Strikingly, *KRAS* mutations are found in PanINs of ~38% of individuals who have no evidence of pancreatic cancer [28]. Reflecting this, current guidelines do not recommend any clinical action for incidental discovery of low-grade PanINs in otherwise healthy individuals [29]. This is because low-grade PanINs are unlikely to progress to adenocarcinoma, and they are likely present in the majority of healthy older individuals [30]. Collectively, these studies support the notion that *KRAS* mutations, while fundamental and early events in disease development, are insufficient to drive complete progression to adenocarcinoma.

This is further supported by experiments showing that some cell types may possess a 'fail-safe' mechanism that senses aberrant signaling driven by oncogenic KRAS. In response to overactive KRAS signaling, the cell cycle can be arrested in the G1 growth phase, eventually leading to cellular senescence [31]. In some studies, this has been linked to the accumulation of proteins encoded by *TP53* and *CDKN2A*, tumor suppressor genes that regulate cell cycle progression [32]. For example, Morton and colleagues used a mouse model of pancreatic cancer to show that premalignant pancreatic lesions formed by oncogenic KRAS exhibit markers of senescence and growth arrest, including upregulation of p53 and its target, p21^{CIP1}, compared to normal ductal cells. p21^{CIP1} is activated by p53 and serves as a potent cyclin-dependent kinase inhibitor that can induce cell cycle arrest. In mice, Morton *et al.* found that lesions formed by oncogenic KRAS only progressed to adenocarcinoma after they deleted both copies of *Trp53*.

It is unsurprising, then, that other genetic alterations necessary for PDAC development include the loss of one or more tumor suppressor genes. Only three tumor suppressor genes—*TP53*, *CDKN2A*, and *SMAD4*—are frequently inactivated in PDAC. These genes are mutated or otherwise inactivated in more than 50-80% of patients [5,9,17,27]. While mutations can occur in other tumor suppressor genes, such as *ARID1A*, these events are limited to fewer than 10% of patients. The high frequency of *CDKN2A*, *TP53*, and *SMAD4* mutations and their regular co-occurrence indicates a functional role in driving disease progression, which has been widely reviewed across numerous studies [5].

As discussed, the stepwise model of PDAC progression posits that pancreatic neoplasms acquire driver mutations sequentially. Oncogenic *KRAS* mutations occur first, followed by loss of *CDKN2A* in low-grade PanINs and then loss of *TP53* and/or *SMAD4* in high-grade lesions and adenocarcinomas [33]. The chronology of these mutations is inferred from targeted and whole-exome sequencing of pancreatic neoplasms spanning low-grade PanINs to invasive PDAC within and across individuals with the disease. But this paradigm is complicated by recent studies that look beyond somatic mutations like single nucleotide variants [9,17,34]. Using whole-genome and/or DNA methylation sequencing, these studies have revealed rampant chromosomal rearrangement patterns and recurrent DNA methylation events that frequently inactivate tumor suppressor genes or amplify oncogenic *KRAS*. It remains to be seen whether these events also follow the sequential ordering proposed for traditional somatic mutations in *KRAS*, *CDKN2A*, *TP53*, and *SMAD4*, though, as discussed, some of these events can occur simultaneously in a subset of tumors.

Regardless of their ordering, these genetic alterations cooperate to enable excessive cell division uncoupled from the cell cycle checkpoints governing the behavior of most normal cells [35]. As pancreatic tumors evolve, malignant cells lose normal cell lineage identity and acquire various phenotypes that facilitate tumor growth. PDAC driver genes operate critical cell signaling nodes, and these mutations consequently dysregulate hundreds of genes in pancreatic cancer cells [36]. In addition to cell autonomous effects, these mutations can effect changes in the local microenvironment.

Desmoplasia in the PDAC Microenvironment

The substantial tissue remodeling characteristic of PDAC has been linked, in part, to oncogenic RAS through the Hedgehog signaling pathway. Hedgehog signaling is critical during embryogenesis, and often becomes reactivated in several cancers as tumors dedifferentiate and lose lineage identity. Oncogenic KRAS may promote the ectopic expression of Sonic Hedgehog (SHH) [37,38], a critical ligand absent in normal pancreatic epithelia but secreted by cancer cells [39]. SHH acts in a paracrine fashion, binding its receptor on local fibroblast and stellate cells. This interaction stimulates fibroblast differentiation into a myofibroblast subtype, signified by expression of α-smooth muscle actin [40]. This active subtype of cancer-associated fibroblasts is causally linked to and representative of the strong desmoplastic response signature to pancreatic tumors. The link between SSH and desmoplasia is supported by pharmacological studies in animal models showing that inhibition of the Hedgehog pathway decreases the proliferation of α-smooth muscle actin-expressing fibroblasts, corresponding to a depletion of stromal tissue [41]. Hedgehog, along with other pathways, can thus promote PDAC's notorious desmoplastic microenvironment.

The densely packed nature of tumor-associated fibroblasts and production of a rich extracellular matrix renders an unusually hypovascularized environment [42]. The extracellular matrix wrought by cancer-associated fibroblasts generates high interstitial pressure, reducing perfusion into the local microenvironment [43]. The tumor stroma thus both physically barricades cancer cells and stifles the delivery route of drugs into the tumor. This is widely cited as a contributor to the failure of traditional therapeutics in pancreatic cancer [5,42,43]. However, while traditional small molecules and biologics may be excluded from the PDAC microenvironment, numerous studies support the notion that *immune cells* routinely traffic in and out of the tumor stroma, prompting an interest in cell-based or cell-mediated therapies as an alternative treatment avenue [44–46].

Immunotherapy

Given the challenges posed by the microenvironment for delivery of traditional drugs, immunotherapy presents a promising alternative. Cancer immunotherapies potentiate pre-existing immune responses, or induce *de novo* responses, that can destroy tumor cells with high specificity. To understand the immune system's role in treating cancer, we will focus on cytotoxic T lymphocytes, commonly known as $CD8^+$ T cells, for the CD8 surface protein that uniquely marks their cellular membrane. Like all leukocytes, T cells originate in the bone marrow. But unlike other major immune cell subsets, T cells traffic to the thymus for their development into specialized cell lineages (this is the origin of their name— "T" for thymus). T cell function is ultimately directed by the T cell receptor (TCR), a surface protein that recognizes short peptides bound by major histocompatibility complex (MHC) molecules. MHC class I is presented on the surface of almost all mammalian cell types [47]. Through their TCRs, $CD8^+$ T cells recognize peptides (8-11 amino acids long) bound by MHC class I molecules, while $CD4^+$ T cells recognizes peptides (13-25 amino acids long) bound by MHC class II molecules. In $CD8^+$ T cells, the TCR:MHC class I interaction is stabilized by the CD8 co-receptor [47].

In the thymus, undifferentiated T cells are molded primarily, though not exclusively, into either $CD4^+$ 'helper' T cells or $CD8^+$ cytotoxic T cells. During this time, all T cells undergo thymic selection, the process by which immature T cells are selected for peptide:MHC binding ability in the thymic cortex and are culled for autoreactivity in the thymic medulla. In the latter stage, called negative selection, autoreactive T cells with TCRs that recognize an organism's natural peptide:MHC complexes with high affinity are destroyed, thereby skewing the repertoire of TCRs toward a "self"-tolerant phenotype and protecting the organism from T cell attacks on healthy cells.

T cells that escape thymic selection are released into the blood stream, and the TCRs expressed by these T cells are collectively known as the TCR repertoire. The TCR repertoire is enormously vast, made possible by V(D)J gene rearrangement, a process of somatic recombination that reshuffles DNA sequences in the genes encoding components of the TCR [47]. This largely random process is thought to be capable of generating over 10^{15} structurally distinct TCRs before thymic selection [48]. Thymic selection then reduces this vast number into the much smaller TCR repertoire active in the human circulatory system.

After thymic selection, naïve T cell lineages defined by a unique TCR and descended from a single T cell persist for years in humans [48]. Sequencing-based efforts suggest that a relatively constant pool of up to 100 million TCRs is maintained throughout a human lifetime, and the repertoire size is continually balanced by thymic output and by events that cause T cell death or persistence. Critically, the constrained TCR repertoire dictates what antigens T cells can respond to in the setting of infection or cancer.

Antigen recognition depends on TCR interaction with MHC molecules, encoded by the most polymorphic gene family in the human genome [47]. In humans, there are three MHC class I genes: human leukocyte antigen (*HLA*)-*A*, *HLA-B*, and *HLA-C*. Collectively, over 6,000 MHC class I alleles exist in the human population, often encoding structurally distinct proteins. This variation primarily occurs in the extracellular domain of the protein, especially in the peptide binding groove. This gives rise to distinct, though often overlapping, peptide binding preferences among different MHC class I alleles. From an evolutionary perspective, the polymorphic nature of MHC class I alleles reflects both the rich diversity of geographic lineages in the human species and a competitive edge against pathogens. Indeed, much of this variation is thought to result from pathogen-driven diversifying selection. In this process, MHC alleles are preferentially maintained in heterozygous combinations that afford individuals a greater breadth of peptide candidates for MHC binding [49].

Once outside the thymus, CD8$^+$ T cells remain inactive and persist at low frequency until they encounter an antigen recognized by their TCR. Prior to antigen engagement, T cells are classified as naïve. T cells are thought to remain in a naïve state until they encounter their cognate antigen presented by an antigen-presenting cell (APC) in a secondary lymphoid organoid, such as a lymph node. While several cell types can serve as APCs, dendritic cells act as professional APCs critical for the initial stimulation and activation of naïve T cells. Dendritic cells reside in tissues throughout the body, either in a resting state or actively sampling their local environments for foreign proteins in response to signs of an infection [50]. As phagocytes, dendritic cells can ingest large amounts of extracellular debris, including mutant proteins released from cancer cells, though a variety of processes. Once internalized, proteins are degraded and traffic through antigen processing pathways for presentation by MHC class I on the surface of dendritic cells [51]. As activated dendritic cells migrate to secondary lymphoid organoids, they encounter and present their antigens to naïve T cells.

Antigen presentation by dendritic cells to a cognate T cell is called T cell “priming.” Priming is a critical point at which T cells differentiate from a naïve state to an effector state. The TCR interaction with cognate antigen, coupled with costimulatory interactions, triggers signaling pathways that induce dramatic increases in T cell proliferation. Proliferation results in an expanded clonotype of many T cells bearing the same TCR [47]. In addition to dendritic cells and TCR:antigen binding, $CD4^+$ T cells and inflammatory stimuli can also promote $CD8^+$ T cell priming, and these secondary cues may be required to elicit full T cell activation, including the development of memory populations [52]. Once primed, effector $CD8^+$ T cells exit secondary lymphoid organs and circulate broadly throughout the body, in search of signals indicating a site of infection or cancer. When engaging cells presenting cognate antigen, effector $CD8^+$ T cells unleash a barrage of effector proteins, including perforins and granzymes that puncture the target cell membrane (perforins) and trigger apoptosis (granzymes).

Though the development, lineages, and functions of T cells extend and diversify well beyond this description, the process described above encompasses most of the core components of the cancer immunity cycle. The final stages involve T cell infiltration into the tumor microenvironment and subsequent interaction with antigen-presenting cancer cells. These interactions can lead to cancer cell death, but they can also lead to T cell exhaustion, a cell state that severely restrains the overall anti-tumor immune response. T cell exhaustion is fundamentally targeted by several successful immunotherapies that will be described later in the next section.

Brief overview of immunotherapy

Physicians and scientists have long speculated that the immune system could be induced to eradicate cancer. This notion can be traced back as early as the 1890s, when surgeon William Coley became perplexed by the loss of his first cancer patient, a young girl who succumbed to sarcoma despite a radical surgery performed by Coley [53]. Poring over medical records at the New York Hospital (now Weill Cornell), Coley stumbled upon an older case of another patient with sarcoma. Stunningly, this patient had possessed an inoperable sarcoma that regressed after a streptococcal infection of his skin. Coley investigated further and found that the patient was now clear of disease. This spurred Coley on to a yearslong development of the earliest reported cancer immunotherapy, eventually called Coley's toxins, which included extract from heat-killed *Streptococcus pyogenes* and gram-negative *Serratia marcescens*. By 1940, more than 20 patients of 104 treated with Coley's toxins experienced permanent and complete disease regression. Over the decades, Coley observed that this treatment had little effect on carcinomas, and his toxins were eventually superseded by chemotherapy, which demonstrated broader applicability across different types of cancer. Of note, though, Coley himself wrote that while his toxins did not elicit a remarkable effect on most carcinomas, he speculated that other vaccine formulations may elicit strong responses in other tissues. This has yet to be achieved for pancreatic carcinomas, though it is worth noting that the Bacillus Calmette–Guérin (BCG) vaccine is among the standard of care for the treatment of modern-day bladder cancer patients. It is widely thought that BCG promotes an effective anti-tumor immune response in these patients, though the precise mechanism remains under investigation [54].

After Coley, numerous other immunotherapies were tested with little success. Progress was probably hampered by a poor understanding of the immune system—even Coley's efforts were done in an era when the germ theory of disease was still in its infancy and was not broadly accepted by the medical community. Furthermore, the primary agents of the adaptive immune system— B and T lymphocytes—were not in common immunology vernacular until the late 1950s [55]. Even the TCR, arguably the most important functional molecule in 21st-century immunotherapy, was not discovered until 1982 [56]. However, these discoveries prompted exponential growth in our understanding of the immune system, which has revolutionized the development of biologically informed immunotherapies.

In this vein, pioneering work by Steven Rosenberg and others led to the first broadly applicable immunotherapy approved by the Food and Drug Administration (FDA): a cytokine therapy based on interleukin-2 (IL-2). Like Coley, Rosenberg's career in immunotherapy was motivated by a case of spontaneous tumor regression, this time in a patient with gastric cancer that had metastasized to the liver and lymph nodes [57,58]. The patient's only treatment had been palliative surgery that left most of the tumor intact. In the excised portion, physicians noted an abundance of tumor-infiltrating T cells, indicating a potential anti-tumor immune response. This prompted an interest in therapeutics that could enhance the anti-tumor activity of T cells, and Rosenberg's case report was followed a few years later by the discovery of IL-2 and its capacity to stimulate and maintain T cell growth *in vitro* [55].

In mice and patients with melanoma, Rosenberg found that IL-2 could elicit robust tumor regression [59]. In mice, this effect was dependent on the presence of lymphocytes, as mice depleted of lymphocytes by irradiation prior to treatment showed no therapeutic benefit. These data prompted large-scale clinical trials, which ultimately led to approval of IL-2 therapy for melanoma patients in 1992. This was a landmark event in the history of immunotherapy, representing the first approval of a broadly applicable cancer treatment that targets cells of the adaptive immune system. While IL-2 therapy elicits unquestionable clinical benefit for some patients, it leads to significant toxicity and has been superseded by immune checkpoint inhibitors, arguably the most important clinical development in cancer treatment since the advent of chemotherapy.

IL-2 therapy stimulates T cell activation and expansion, promoting their ability to infiltrate and attack pockets of antigen-presenting target cells. In turn, cancer cells may respond with additional defense mechanisms, including the upregulation of surface co-inhibitory factors such as programmed death-ligand 1 (PD-L1). PD-L1's primary function is to regulate T cell responses, preventing the immune system from targeting normal healthy tissue. PD-L1 is typically expressed on the surface of many immune cell types and in many non-hematopoietic tissues (e.g., epithelia, endothelial cells) [60]. PD-L1 binds to PD-1, a receptor expressed by many T cells subtypes under certain conditions, including after initial T cell activation and in the setting of chronic antigen exposure [61]. This binding interaction triggers downstream signaling events, including the recruitment of SH2-containing phosphatase 2 [62], in PD-1-expressing T cells. These events serve to constrain outputs from TCR and CD28 signaling and result in reduced T cell proliferation and cytokine production [63]. Collectively, this can promote a state of T cell dysfunction, rendering T cells unable to mount an effector response upon cognate antigen recognition. While helpful in regulating initial T cell activation and tempering an overactive immune response, this immune checkpoint can be co-opted by tumor cells that overexpress PD-L1 to shield themselves from T cell attack.

Another key checkpoint involves cytotoxic T-lymphocyte associated protein 4 (CTLA-4), which functions in a less direct manner to negatively regulate T cell responses. CTLA-4 can be expressed in both cytotoxic T cells and regulatory T cells. In cytotoxic T cells, CTLA-4 competes with a co-stimulatory receptor, CD28, for binding to CD80 and CD86 expressed by antigen-presenting cells [64]. CD28, when bound, normally amplifies the activation of T cells, encouraging growth and the transition to an effector state [47]; however, when sufficiently expressed, CTLA-4 outcompetes CD28 for CD80/86 binding, stifling overall T cell activation [65]. In the context of chronic antigen exposure, as is the case with tumor antigens, T cells may become exhausted and unable to execute effector functions, due in part to expression of CTLA-4 on their surface. Regulatory T cells also express CTLA-4 and can actively suppress $CD8^+$ T cell activation by stripping antigen-presenting cells of CD80/86 through transendocytosis [66]. Many of the mechanistic insights into CTLA4 were made in the 1980s and 1990s [67,68]. For example, James Allison and colleagues hypothesized that this molecule may restrain the immune system from a sustained attack on tumor cells. In 1996, they showed that antagonistic antibodies specific to CTLA-4 reduced tumor growth in mouse models bearing antigen-expressing colon carcinoma transplanted beneath the skin [69]. 15 years later, the FDA approved Ipilimumab, an anti-CTLA-4 monoclonal antibody. After a series of clinical trials demonstrating robust and durable efficacy, Ipilimumab became part of the standard of care for patients with advanced stage IV, inoperable melanoma [70]. Soon after, additional monoclonal antibodies targeting PD-1 (Nivolumab) and PD-L1 antibodies were approved for advanced melanoma as well. Furthermore, the combination of both nivolumab and ipilimumab is now among first line therapy for metastatic melanoma and some additional cancers [71]. Strikingly, these immune checkpoint inhibitors (ICIs) completely eradicate widely metastatic tumors in a subset of patients, providing durable cures for over a decade in some individuals [72]. In a Phase III trial of advanced melanoma, 52% of patients treated with this combination survived more than five years after treatment. Though highly effective, it should be noted that this combination can still incur substantial toxicities.

While ICIs yield stunning effects in a subset of patients with melanoma and lung adenocarcinoma, clinical trials in PDAC have completely failed to demonstrate efficacy. Most notably, two trials tested single agent anti-CTLA-4 and anti-PD-L1 antibodies in a total of 41 PDAC patients [73,74]. Unfortunately, not a single therapeutic response was observed in either study. Potential reasons for this will be explored in the next section.

Beyond ICIs, cancer vaccines and adoptive cell therapies seek to circumvent the endogenous immune system's limited capacity to sustainably recognize and destroy cancer cells. Adoptive cell therapy (ACT) often involves the transfer of T cells obtained from the patient's body, such as through extraction of tumor-infiltrating lymphocytes—TILs—from resected tumors. These cells are then expanded *ex vivo* and reinfused into the patient. ACT can also involve T cells obtained outside the patient's body, e.g., from a donor individual. These T cells naturally possess or are engineered to express a TCR specific to a known tumor antigen [75].

With ACT, clinicians aim to eradicate the cancer by exogenous supply of T cells already known to target an antigen. Some companies are also pursuing synthetic enhancement of these T cells (e.g., by deleting the gene encoding PD-1) prior to delivery, thus potentially endowing T cells with enhanced properties, such as resistance to dysfunction. Though numerous ACT modalities have been tested, with some limited success, chimeric-antigen receptor T cells (CAR-T) are currently the only FDA-approved ACT. CAR-T cells are engineered to target a surface protein rather than a peptide:MHC molecule. Unfortunately, the clinical benefit of CAR-T is currently limited to a subset of patients with hematopoietic malignancies [76]. CAR-T trials targeting mesothelin, a protein highly overexpressed in some pancreatic tumors relative to most normal tissues, have shown minimal, rather underwhelming clinical benefit in pancreatic cancer, but additional studies are ongoing [77,78].

Other types of ACT, such as those targeting the MART-1 (melanoma antigen recognized by T cells-1) peptide:MHC complex, have struggled in part due to unanticipated TCR cross-reactivity with off-target complexes [79]. This has led to severe adverse events, including treatment-related fatalities in the worst examples. By contrast, there are more promising examples of impressive efficacy without severe adverse events. For example, Tran and colleagues infused a colon cancer patient with TIL-derived cells that recognized a neoantigen derived from the $KRAS^{G12D}$ mutation [80]. This patient experienced a dramatic regression of all liver metastases that was sustained for at least nine months. A similar result was recently reported for a pancreatic cancer patient treated with T cells specific to the same KRAS mutation. In this patient, autologous T cells were engineered *ex vivo* to present a TCR specific to the neoantigen. Like the colon cancer case, multiple metastases regressed, a response sustained up to at least six months, when the study was published [78]. These examples inspire continued optimism for the potential of ACT as a cancer treatment modality. With continued identification of cancer-specific antigens and increased knowledge of T cell receptor reactivities, researchers may eventually identify ACT strategies that effectively treat solid malignancies.

Like ACT, cancer vaccines aim to mount an anti-tumor immune response where one may not previously exist. While traditional vaccines for infectious disease are often prophylactic, cancer vaccines are often intended to treat pre-existing disease. (There are clinically approved examples of prophylactic cancer vaccines for certain virus-driven cancers, such as human papillomavirusdriven cervical cancer.) Cancer vaccine strategies vary and include applications like injection of the patient with proteins representing known cancer-specific antigens or with genetical material encoding such antigens [81]. However, the only FDA-approved, non-prophylactic cancer vaccine targeting a known tumor antigen is Provenge, approved for the treatment of hormone refractory prostate cancer. Provenge works through vaccination outside the patient's body: dendritic cells are isolated and pulsed *ex vivo* with prostatic acid phosphatase (PAP), an antigen often overexpressed in prostate cancer. These dendritic cells are then reinfused to the patient to effectively prime T cells and induce an anti-tumor response. This therapy elicits a modest clinical benefit, with a median increase in overall survival of ~4 months versus patients treated with placebo [82]. Beyond Provenge, many cancer vaccine strategies are currently in trial, and optimism exists for greater clinical benefit based on the advent of new vaccination modalities and strategies. These include mRNA-based vaccines, as well as improved bioinformatic prediction of cancer-specific antigens [83].

The vast landscape of immunotherapy stretches well beyond those described, encompassing CD40 agonists that aim to enhance T cell priming *in vivo*, and emerging ICIs such as antibodies targeting T cell immunoreceptor with Ig and ITIM domain (TIGIT) and many other classes still [45]. Beyond modulators of immune cell function, other emerging approaches engineer synthetic proteins that target predefined peptide:MHC complexes. For example, Yarmarkovich *et al.* introduced an approach for developing novel CARs that recognize peptide:MHC class I complexes via single chain fragment variables (scFvs). These were optimized through an iterative process of panning the target complex against a phage library of more than $2*10^{10}$ structures [84]. After integration of optimized scFvs into a CAR construct, these researchers demonstrated antigen-specific killing of tumor cells by the engineered CAR-T cells *in vitro* and *in vivo*. Using a similar phage panning approach, Han-Chung Hsiue and colleagues developed a bi-specific single-chain diabody consisting of an scFv that binds a peptide derived the common $p53^{R175H}$ point mutation in complex with the common HLA-A*02:01 allele [85]. The scFv is fused to a traditional antibody that binds to a constant region of the TCR-CD3 complex on T cells. This second binding event triggers downstream TCR signaling, even if the TCR is not specific to the target antigen. This promotes a T cell effector response as if the T cell had recognized its cognate antigen, redirecting a killing response against the nearby tumor cell. Like Yarmarkovich *et al*., these researchers demonstrated antigen-specific killing of tumor cells *in vitro* and *in vivo*. Both studies highlight an improving ability to engineer novel protein structures that can redirect T cells to prespecified cancer antigens. As these technologies improve, they will expand the repertoire of cancer-specific targets amenable to immunotherapy and potentially broaden the cohort of cancer types that can be effectively treated.

Methods to identify immunotherapy targets in pancreatic cancer

As discussed, immunotherapies do not currently benefit most patients suffering from PDAC. Many studies have posited that this is due to an ‘immunologically cold’ tumor microenvironment, in which cytotoxic T cells are scarce, immunosuppressive cells are abundant, and a low mutational burden precludes productive immune recognition. In principle, a low mutational burden should give rise to few or no mutated peptides (neopeptides) that could be bound by MHC class I molecules and presented on the tumor surface for recognition by $CD8^+$ T cells. Neoantigens efficiently bound and presented by MHC are an empirically validated target of anti-tumor T cells and many T cell-based immunotherapies, including ICI, cytokine therapy, and ACT [80,86]. Their supposed absence in PDAC patients that fail to respond to ICI is often contrasted with the success of ICIs in a small cohort (<2%) of PDAC patients whose tumors bear defects in the mismatch repair (MMR) system. Such defects include loss-of-function mutations in an MMR gene like *MSH2*. MMR deficiencies promote errors during DNA replication, leading to a higher mutational burden, more than five or ten times that of a typical PDAC tumor [87]. A higher mutational burden, in theory, should produce more neoantigens and is typically predictive of a better ICI response within and across many tumor types. Accordingly, most patients with MMR-deficient PDAC tumors experience a significant clinical benefit when treated with ICI [88]. The low mutational burden in MMR-proficient PDAC is widely cited to explain why patients do not respond to ICI [88–90]. However, emerging evidence indicates the primary reason may involve defects in the immunosurveillance of pancreatic tumors [44]. Specifically, Hedge and colleagues showed that dendritic cells are both scarce and dysfunctional in PDAC tumors of both patients and mouse models. This was drastically different than the microenvironment of lung adenocarcinoma, a cancer subtype that often responds to ICI and is frequently infiltrated with $CD8^+$ T cells bearing markers of exhaustion. In the lung, tumors were typically infiltrated with functional dendritic cells presenting co-stimulatory molecules that promote T cell priming. This suggests that dendritic cells efficiently survey the lung tumor microenvironment for antigens, leading to efficient T cell priming and a neoantigen-specific response. By contrast, the scarcity of dendritic cells in PDAC tumors prevents adequate immunosurveillance. This would preclude neoantigen-specific T cell priming and subsequent effector cell exhaustion. Without these steps, PDAC patients would not benefit from ICI, since they lack effector T cells that recognize tumor antigens to begin with.

Furthermore, recent computational studies have challenged the notion that PDAC tumors lack neoantigens, predicting dozens of neoantigens in the typical PDAC patient [45,91–93]. This suggests that neoantigens play a previously underappreciated role in PDAC immunology and has prompted continued clinical interest in immunotherapy for this disease. Most recently, a landmark Phase I trial this year reported positive results for neoantigen-based vaccines in pancreatic cancer patients [81]. In this trial, 16 patients received mRNA-based vaccines informed by bioinformatic neoantigen prediction based on somatic mutations detected in resected samples. They also received FOLFIRINOX and anti-PD-L1 antibody treatment. Eight patients demonstrated a response, exhibiting a longer relapse-free survival than non-responders, with the median unreached in the responder cohort after more than 15 months. While these samples sizes are insufficient to make final conclusions, they suggest that neoantigen-based vaccines may soon play a role in prolonging patient survival after surgery. Mechanistically, researchers also observed a robust expansion of neoantigen-specific T cells in these patients, supporting the expected therapeutic mechanism and underscoring the potential of this approach.

To date, the field has primarily analyzed pancreatic cancer neoantigens derived from missense mutations that are typically unique to individual patients. Within each tumor, many missense neoantigens are predicted to reside in only a fraction of tumor cells, limiting therapeutic potential. However, neoantigens can derive from a much broader class of mutations and other, non-mutational sources [94–98]. Each of these represents a potentially untapped pool of immunotherapy targets in pancreatic cancer. The next section will review several such examples.

Frameshift mutations

Many early papers predicting neoantigens focused solely on those derived from missense mutations that alter a single amino acid. For example, seminal work by Gubin *et al*. (2014) and Campbell *et al*. (2016) predicted and validated missense-derived neoantigens in both mouse models and human tissue, respectively. However, both papers limited their analysis to single nucleotide substitutions, but somatic insertions/deletions (indels) are also common mutations in human cancer. Indels in protein-coding regions can create frameshift mutations that eliminate protein function by shifting the reading frame, thereby disrupting the entire downstream protein sequence [99].

Such frameshift mutations can thus generate long, novel peptide sequences that, in principle, could recur across different cancer patients, as different frameshifts in the same gene can yield overlapping alternative reading frames. Frameshift mutations in tumor suppressor genes (TSGs), especially *TP53*, *SMAD4*, and *CDKN2A*, are common drivers of pancreatic cancer. I hypothesize that frameshift mutations in TSGs may generate recurrent neoantigens capable of stimulating tumor-specific cytotoxic T cell responses. Given their intrinsic link to a tumor-initiating mutation, I also predict that these neoantigens would be clonal or expressed by most tumor cells. Some bioinformatic evidence and a single clinical example in a melanoma patient support the possibility of TSG-derived frameshift neoantigens [100,101], but the existence of TSG frameshift-derived neoantigens has not been investigated in pancreatic cancer. If confirmed in pancreatic cancer, these neoantigens could represent powerful targets for immunotherapies, particularly adoptive T cell therapy and neoantigen vaccines, even in patients with a low mutational burden, who may otherwise lack potent neoantigens.

Frameshift neoantigens are also interesting in the context of thymic selection. During negative selection, autoreactive T cells with TCRs that recognize an organism's natural peptide:MHC complexes too strongly undergo apoptosis, thereby skewing the repertoire of TCRs toward a "self"-tolerant phenotype. This may explain why the majority of predicted missense neoantigens are not immunogenic in validation studies [102], as cross-reactive T cells that recognize both the neoantigen and its wild-type analog are removed during negative selection, biasing the TCR repertoire against recognition of neopeptides that differ from their wild-type counterpart by only one amino acid. The exceptions would be neopeptides that are sufficiently divergent from self-peptides or those that substantially enhance MHC class I binding affinity. Since frameshift mutations generate entirely novel amino acid sequences, neopeptides derived from frameshifts in TSGs may also be frequently immunogenic.

While TSG frameshift-derived neoantigens could be immunogenic, clonal, and recurrent among patients, they may be less likely to be translated into protein. Frameshift mutations generally induce a premature stop codon (PTC) encoded within the alternative reading frame. PTCs can reduce transcript expression by triggering nonsense-mediated decay (NMD), a pathway that recognizes and degrades PTC-bearing transcripts. Though NMD can certainly target frameshift-bearing transcripts, Litchfield and colleagues showed that many frameshifts frequently escape NMD in cancer, depending partly on the position of the indel within the transcript [103]. Furthermore, Roudko *et al*. (2020) showed that frameshift mutations are translated into protein in human MMR-deficient endometrial, colorectal, and stomach cancers.

Finally, while NMD may target transcripts for degradation, thereby reducing protein levels, some level of peptide is still translated during the pioneering round of translation, before NMD takes place, as so-called defective ribosomal products. These products are preferentially shuttled to the MHC class I presentation pathway [104], potentially enhancing the likelihood of antigen presentation despite an overall loss in protein production. Thus, while NMD may reduce expression of potential neoantigens derived from frameshift mutations, many are likely still translated and represent a source of MHC class I-associated peptides.

Retained introns

In addition to mutations in protein coding genes, variation in the transcriptome may represent a source of cancer-specific antigens. Like the genome, the transcriptome also becomes dysregulated and distorted as cancer cells evolve and de-differentiate from normal cell lineages, leading to global changes in gene expression patterns [105]. This can affect the expression of genes encoding splicing factors, promoting errors in splicing regulation and the consequential production of novel mRNA transcripts derived from retained introns (RIs), exon skipping or shuffling, other splicing events [106]. Some splicing factors also serve as frequently mutated cancer driver genes. For example, *SF3B1*, an RNA splicing factor that regulates intron excision during mRNA processing, is often mutated in uveal melanoma and other cancers [104]. Mutations in *SF3B1* can lead to the production of more than 1,000 novel splicing junctions, including the retention of introns in mRNA. These mis-spliced transcripts have been shown to generate novel protein sequences and immunogenic peptide:MHC complexes in human patients [104]. Alternative splicing events like these are thought to be cancer-specific, since the responsible mutations (e.g., in *SF3B1*) are limited to the cancer cells. Thus, dysregulated alternative splicing can represent another class of cancer-specific epitopes not presented by normal tissues nor incorporated into thymic selection.

In 2018, Smart *et al*. published the first demonstration of retained introns encoding peptides presented by MHC class I in cancer. Using a novel RNA-Seq analysis pipeline, they identified hundreds of transcripts retaining introns in melanoma cells. Many of the same introns were absent in a cohort of normal skin tissue controls, suggesting possible cancer specificity. With mass spectrometry, they then profiled MHC class I-bound peptides on the surface of human melanoma, B cell lymphoma, and leukemia cell lines, identifying one or two peptides derived from cancer-specific retained introns in every cell line. These findings established proof-of-concept for retained introns as an additional source of cancer epitopes, though broader expression of retained introns in tissues throughout the body was not examined.

Like frameshift mutations, retained introns (RIs) are underexplored as a source of cancer-restricted antigens in PDAC. While a recent study predicted retained introns in a cohort of PDAC samples from The Cancer Genome Atlas [107], these researchers only considered bulk tumor samples. This represents a confounding variable, as bulk pancreatic tumor samples notoriously possess more stromal cells than they do cancer cells [17]. Retained introns are primarily detected by calculating the ratio of sequencing reads that span an exon-intron junction over the those that span the annotated exon-exon junction [97]. Thus, in bulk tumor samples, the intron retention ratio may be diluted by sequencing reads derived from stromal cells unaffected by dysregulated splicing. Furthermore, it remains unclear whether retained introns reported in this study are present in normal tissues, as the investigators only considered normal pancreas samples as healthy tissue controls. This limitation will be explored further in Chapter 2.

Novel or unannotated open reading frames (nuORFs)

Only 1-2% of the human genome content is predicted to encode a protein sequence [108]. For decades, the remaining genome was dismissed by many as “Junk” DNA without an obvious function [109]. However, it is increasingly appreciated that the noncoding genome performs a variety of functions. For example, many non-genic sequences encode long noncoding RNAs or microRNAs that fine-tune the expression levels of canonical protein-coding genes [110]. Other regions encode gene promoters and enhancers that interact with transcription factors to orchestrate gene expression programs, including the use of alternative isoforms [105]. Consortium efforts like ENCODE have even established that most of the human genome is pervasively transcribed into RNA [109], yet most of this RNA is not thought to encode protein. Noncoding RNAs, by definition, lack open reading frames (ORFs), or they possess ORFs too short to give rise to a functional protein [111]. Most noncoding RNA sequences have not been conserved throughout mammalian evolution, suggesting they would not encode proteins with important function. Thus, noncoding RNAs are often ignored as potential sources of cancer antigens.

However, in 2015, pioneering work from the laboratories of Aviv Regev and Kevin Struhl revealed that many noncoding RNAs are bound by ribosomes, the organelle machinery responsible for mass production of cellular proteins. While this has long been established for mRNAs, Ji *et al*. (2015) found that ORFs deriving from as many as 40% of lncRNAs and pseudogenes can be translated by ribosomes into protein sequences. Ji and colleagues further found that ~9% of these novel protein-coding ORFs are evolutionarily conserved between mice and humans, suggesting a selective pressure to maintain critical functions. These results show that the mammalian proteome is broader and more complex than initially appreciated. Ouspenskaia *et al*. (2022) later showed that protein sequences encoded by these novel or unannotated ORFs (nuORFs) are processed for MHC class I presentation in a variety of both normal and cancer cell types. This was validated through extensive ribosome sequencing (Ribo-Seq), a technique that identifies mRNAs undergoing active translation. With Ribo-Seq profiling, the researchers empirically identified thousands of translated nuORFs contained within long noncoding RNAs (lncRNAs), pseudogenes, and 5' and 3' untranslated regions. These data facilitated construction of an augmented proteome search space for the detection of peptides eluted from MHC class I molecules. This enabled a second layer of empirical validation, and the results provided direct evidence of nuORF-derived peptides bound to MHC class I molecules. Notably, many of these peptides derived from transcripts that were not detectably expressed in normal tissues, suggesting that nuORFs could represent yet another class of cancer-restricted epitopes. Furthermore, Ouspenskaia and colleagues identified dozens of somatic mutations predicted to alter the protein sequences of nuORFs, thus furnishing an additional source of mutation-derived neoepitopes. nuORFs remain unexamined as a source of cancer-restricted epitopes in pancreatic cancer.

Other sources

Though they will not be examined in the present thesis work, other classes of cancer-specific antigens should be mentioned. For example, alternative splicing events beyond intron retention could give rise to cancer antigens. These include exon skipping, exon shuffling, and the formation of alternative 3' or 5' splice sites [106]. Lu *et al*. (2021) found that pharmacological modulation splicing factors amplifies the production of such events, increasing the load of alternative splicing-derived peptides presented by MHC class I [112]. Beyond splicing, dysregulated RNA editing by double-stranded RNA-specific adenosine deaminase (ADAR) could induce novel amino acid substitutions, and alternative polyadenylation events could lead to the translation of RNA sequences that are not normally read by the ribosome [106].

Prediction of cancer-specific MHC-associated peptides

Prediction of cancer-restricted antigens is informed by bioinformatic analysis of genomic or transcriptomic data. Next-generation RNA sequencing (RNA-Seq), whole exome sequencing (WES), and whole genome sequencing (WGS) are the basis of numerous high-profile papers predicting and/or detecting cancer-specific antigens in patients [113–115]. To date, neoantigen prediction has largely emphasized missense mutations, which are predicted using a common computational framework. In general, this framework operates by 1) identification of patient-specific HLA alleles with HLA genotyping algorithms, 2) somatic variant calling, 3) mutant peptide sequence annotation, and 4) the prediction of binding affinity between patient-specific MHC class I alleles and all possible mutation-derived peptides. Most neoepitope prediction follows this general workflow, though dozens of additional parameters can be incorporated, including RNA expression levels and the predicted likelihood of peptide processing and stability [83]. Candidate neoepitopes can be further filtered or ranked according to several variables, such as the strength of predicted MHC binding affinity, DNA or RNA variant allele frequency, and the ratio of mutant peptide binding affinity for MHC versus that of the cognate wild-type peptide, among others.

The ultimate utility of neoepitope prediction pipelines depends on the accuracy of each of the above steps. The first step, HLA genotyping, is extremely accurate. In a benchmarking comparison of five popular HLA typing software tools, OptiType achieved the highest accuracy in calling donor-specific class I alleles, across both WGS (71% accuracy) and WES (98%) datasets. This was benchmarked against a set of 1,000 HLA genotypes validated with polymerase chain reactions [116].

Somatic variant calling is similarly accurate and highly sensitive: approximately 95% of single nucleotide variants (SNVs) and insertions/deletions (indels) with a variant allele frequency of at least 15% are reliably detected, assuming 50X coverage (i.e., the number of sequencing reads that map to a defined nucleotide in the genome) [117]. Variant annotation tools, such as Ensembl's Variant Effect Predictor [118], are also highly reliable; however, their ability to predict a mutant peptide sequence depends on underlying functional annotations of each genomic sequence in a reference database. For SNVs and indels affecting coding sequences, these tools are extremely accurate and even predict downstream frameshifted protein sequences.

While these upstream components are extremely accurate, peptide:MHC binding affinity predictions are more variable. One of the leading programs in the field is the NetMHCpan-4.1 prediction algorithm [119], an artificial neural network. NetMHCpan-4.1 is trained by data derived from both peptide:MHC binding affinity assays and mass spectrometry profiling of peptides eluted from MHC complexes. The training dataset underlying NetMHCpan-4.1 includes peptides from at least 250 protein-coding MHC class I alleles, including several high frequency alleles in the human population. While allele-specific accuracy depends on the size of the respective training set [120], NetMHCpan-4.1 is extraordinarily sensitive. When assessed for performance on biochemically-validated MHC binding peptides, NetMHCpan-4.1 predicted binding of ~97% of candidate ligands to the corresponding MHC allele [121]. While highly sensitive, NetMHCpan-4.1 has a considerable false positive rate, though estimates of this vary [120,121].

In practice, NetMHCpan-4.1 outputs various metrics, including a predicted MHC binding affinity (IC_{50}) and an allele-specific percentile binding rank (i.e., the ranking of the peptide's predicted affinity relative to a set of random natural peptides). Traditionally, researchers establish an IC_{50} threshold (typically <500 nanomolar) to designate neoepitope candidates, though it is increasingly common to defer to the allele-specific percentile rank and designate a cutoff of 2%. This has recently been shown to promote the highest balance between sensitivity and specificity [119,121].

The high sensitivity of these tools makes them valuable for antigen discovery efforts. This is well highlighted in work by Lu *et al*. (2021). Using an earlier version of NetMHCpan (4.0), they demonstrate that out of thousands of murine MHC class I-bound peptides identified from mass spectrometry, more than 80% were predicted to bind to mouse MHC class I alleles. In contrast, less than 1% of peptides selected from random positions in the proteome were predicted to bind to MHC class I. These results highlight that most peptides discovered with empirical assays can also be predicted bioinformatically.

The accuracy and broad utility of neoepitope prediction pipelines is evidenced in many empirically validated preclinical and clinical studies [81,113–115,122]. For example, studies by Sahin *et al*. (2017) and Keskin *et al*. (2019) harnessed neoantigen prediction algorithms to conduct Phase I clinical trials in patients with melanoma and glioblastoma, respectively [113,122]. Both studies assembled neoepitope-based cancer vaccines for delivery to patients. These vaccines elicited *de novo* CD8$^+$ T cell responses against patient-specific neoepitopes in both melanoma and glioblastoma. This corresponded to increased tumor infiltration of neoantigen-specific T cells in glioblastoma and sustained progression-free survival in melanoma. Strikingly, both studies and other clinical trials report an unexpected yet robust CD4$^+$ T cell response against vaccinating neoantigens [114]. The reasons for this remain unclear and warrant further investigation.

Beyond the clinic, neoepitope predictions enabled the discovery of hundreds of empirically validated neoepitopes in human samples and cancer mouse models. For example, seminal work by Gubin and colleagues harnessed neoepitope prediction to identify natural neoepitopes derived from missense mutations in *Lama4* and *Alg8* in a 3-methylcholanthrene-induced (MCA) mouse sarcoma cell line. This enabled the first empirical demonstration that mutation-derived neoepitopes are targeted by exhausted, PD-1$^+$ CD8$^+$ T cells in an animal model *in vivo*. This provided key mechanistic insights regarding anti-tumor CD8$^+$ T cells following therapeutic reinvigoration with immune checkpoint inhibitors. These two neoepitopes now form the basis of many preclinical murine systems that enable researchers to probe the interaction of cancer and the immune system *in vivo*. This is particularly important as classic genetically engineered mouse models of cancer accumulate far fewer mutations than human cancers [45,123].

Despite their tremendous utility, peptide:MHC binding algorithms do not account for critical upstream cellular processes that can dictate which peptides ultimately get presented on the surface of cells [120]. Differences in protein turnover rates, cellular localization, and proteasomal processing patterns influence which proteins are degraded into the specific 8-11mer peptide sequences required by MHC class I alleles. Finally, processes that affect mRNA stability and degradation influence protein abundance, an important consideration as peptide substrate concentrations must be sufficient for meaningful interaction with MHC molecules and engagement with cognate T cells [123]. These caveats represent one reason why many predicted neoepitopes are not empirically validated by mass spectrometry.

For non-mutation-based cancer antigens, epitope prediction often relies on RNA-Seq data, rather than whole exome or whole genome sequencing. For example, RNA-Seq data is required to identify cancer-specific splice isoform variants and their encoded epitopes [106]. This is necessary in part to help restrict the search space of downstream empirical assays. For example, mass spectrometry profiling of MHC-bound ligands, commonly called immunopeptidomics, can only identify peptide sequences that are presented within a defined proteome search space. In most studies, this search space simply includes the reference proteome and thus excludes novel splice isoforms.

The confidence in mapping spectra back to the originating protein is quantified by a false discovery rate (FDR), i.e., the proportion of identified peptides passing thresholds that are not actually the correct spectral mapping [95]. Several studies demonstrate that the size of a protein search space affects the FDR in immunopeptidomics experiments [95,112,124], and excessively large search spaces promote an unacceptably high FDR [124]. Due to these limitations, an augmented search space comprised of all possible alternative splicing events is impractical. Instead, RNA sequencing data can be used to identify aberrant transcripts to construct a smaller, informed search space. These principles apply to other sources of translated products, including transcripts from the non-coding genome, which could generate a theoretically vast number of potential peptides.

Candidate epitopes computationally predicted or empirically detected are often assessed for immunogenicity, i.e., their ability to stimulate a T cell response. There are numerous variations of T cell immunogenicity assays. In general, antigen-presenting cells are stimulated with a candidate peptide (or a genetic sequence encoding the peptide). These cells are cocultured with $CD8^+$ T cells, which are assessed for a functional T cell response. This is often signified by the release of interferon gamma, granzyme B, or other functional molecules released by effector $CD8^+$ T cells upon antigen recognition [96,102]. Other cellular assays employ customized peptide:MHC tetramers, complexes comprised of four biotinylated, peptide-loaded MHC protein monomers bound to a streptavidin core, conjugated to a fluorophore. These reagents bind TCRs, enabling identification and isolation of antigen-specific T cells via flow cytometry.

In studies of cancer, both types of T cell-based assays are often performed using tumor-infiltrating lymphocytes (hereafter, TILs). In principle, identification of antigen-specific $CD8^+$ TILs demonstrates that a candidate neoepitope elicited an anti-tumor immune response in a patient or animal from which the epitope was predicted (or detected). However, TIL-based screening validates very few neoantigens, often fewer than 2% of nominated neoepitope candidates. This number reaches only as high as 6% in consortium-based efforts that optimize neoepitope prediction [83,102].

Some investigators argue that TIL-based assays suffer from inherent biases that underestimate the true cancer epitope burden [125]. For example, TIL-based neoantigen validation assumes the subject being studied successfully mounted an anti-tumor immune response against every presented neoepitope. It further presumes that the fraction of TILs obtained at the time of extraction represents the full life history of the anti-tumor immune response. It is conceivable, however, that cancer-specific TILs are variably present, despite continued expression of cognate antigen.

Expounding on this concept, Strønen *et al*. (2016) demonstrated that healthy donor peripheral blood mononuclear cells (PBMCs) enable *in vitro* generation of neoepitope-specific T cells from undetectable naïve TCR repertoires [125]. This study generated antigen-specific T cells for 11/57 (19%) tested neoepitopes nominated by a prediction pipeline analysis of melanoma patient tumors. Critically, these donor-derived T cells could recognize and elicit an effector response against cultured tumor cells from which the neoepitopes had been predicted. Furthermore, patient-derived TILs were found to recognize only one of the eleven validated neoepitopes, suggesting that patient T cells frequently failed to respond to *bona fide* neoantigens. These results support the notions that (1) the host immune system does not necessarily mount a response against every neoepitope presented by cancer cells and (2) that TIL-based screening may underestimate the presence of candidate neoepitopes nominated by prediction pipelines. The true cancer-specific epitope burden in tumors likely falls between the number predicted by bioinformatic pipelines and the number validated by empirical assays. In the future, as computational and empirical practices improve, researchers will more accurately identify presented neoepitopes, enabling more informed immunotherapies.

While neoantigen prediction and validation has been well documented in melanoma and lung cancer, the cancer-specific epitope landscape in PDAC remains vastly understudied. Prior efforts in PDAC are primarily limited to missense mutations, including an elegant study by Balachandran *et al*. (2017). These researchers demonstrated that tumors from long-term survivors of PDAC are enriched with missense-derived neoepitopes that bear high sequence homology to viral peptides, which may render them highly immunogenic. Accordingly, the authors identified frequent T cell responses to these neoepitopes in patient samples, suggesting a successful anti-tumor immune response is responsible for the sustained tumor regressions in this cohort of long-term survivors. In a similar study, Łuksza, Sethna, and Rojas *et al*. (2022) later showed that missense neoepitopes in the *TP53* tumor suppressor gene can elicit immune responses in PDAC patients. Beyond these studies of missense mutations, the broader neoepitope landscape remains largely understudied in PDAC.

Empirical investigations are also lacking. To date, only one immunopeptidomics study in human PDAC has been published [126]. This study primarily analyzed bulk tumor tissue, limiting the ability to distinguish MHC class I-bound peptides derived from cancer cells from those derived from stromal cells. Furthermore, these investigators only scanned the reference proteome, preventing identification of epitopes derived from mutations, alternative splicing events, and nuORFs.

These observations highlight a critical gap in our knowledge of pancreatic cancer. Few efforts have been made to characterize the landscape of antigens in this deadly disease, despite abundant computational and empirical evidence of diverse cancer antigens in other malignancies. Emerging clinical data, such as the neoepitope vaccination trial discussed previously, underscore the potential for effective PDAC immunotherapies informed by knowledge of the neoepitopes borne by patient tumors. This potential is further reinforced by work in mouse models of PDAC, including work from our laboratory and others that have devised effective immunotherapy strategies in the preclinical setting.

In the context of PDAC, the KPC (*Kras*$^{LSL-G12D/}$;· *Trp53*$^{LSL-R172H/+}$;*Pdx1-Cre*) mouse model is considered the gold-standard because of its fidelity to salient traits of the human disease, such as a fibroblast- and macrophage-dominant tumor microenvironment, similar neoplastic tissue morphology, and propensity to metastasize to tissues such as the liver, lung, and lymph nodes. Though these are desirable attributes, Evans *et al*. (2016) reported that KPC tumors have mutation burdens far lower than the typical human cancer and lack predicted neoantigens. Thus, the KPC model is not ideal for studying the relationship between neoantigens and the immune system. To circumvent this, my colleagues William Freed-Pastor and Laurens Lambert developed a new system based on the orthotopic transplantation of genetically defined $Kras^{LSL\text{-}G12D/+}$; $Trp53^{flox/flox}$;$H11^{neoantigen}$ pancreatic organoids, which express a defined neoepitope: the high affinity MHC class I-restricted $CD8^+$ T cell antigen $OVA_{257–264}$ [SIINFEKL], or missense-derived neoepitopes from *Lama4* or *Alg8*. Defined neoantigens enable us to precisely track antigen-specific $CD8^+$ T cell responses *in vivo*. Importantly, we showed that, like the KPC model, pancreatic adenocarcinomas evolving from these transplants preserve key histopathological features of the human disease.

After profiling neoepitope-specific TILs in these mice and studying other TILs derived from PDAC patients, we found that TIGIT and PD-1, both exhaustion markers, were frequently co-expressed on the surface of TILs from our model and from patients. We then devised a triple combination immunotherapy based on CD40 agonism, PD-1 inhibition, and TIGIT inhibition. Notably, we observed a 46% objective response rate and, most importantly, a complete tumor regression in 23% of the 48 animals tested with this combination in our preclinical trial (Appendix I). In contrast, all possible single-agent and dual-agent combinations failed to elicit a powerful response (at best an 11% ORR with no CRs using CD40 agonism as a monotherapy). Notably, anti-PD-1 monotherapy did not elicit a response in any animal (0% ORR; n = 9), corroborating prior work in the KPC mouse model [127]. Our results with TIGIT/PD-1 co-blockade plus CD40 agonism provided a powerful preclinical proof that carefully designed immunotherapy can elicit a therapeutic response *in vivo*, even in mouse models that exhibit immunosuppressive hallmarks of the human disease.

Motivated by these preclinical findings, as well as the profound success of immunotherapy in other tumor types, I sought to address the gap in our understanding of the neoepitope landscape in human PDAC. As I will describe in Chapter 2, this work began with the development of a suite of bioinformatic pipelines. Collectively, these enable prediction of cancer-restricted epitopes derived from missense and frameshift mutations, retained introns, and nuORFs. These tools were applied to whole genome sequencing (WGS) and RNA-Seq data obtained from pancreatic cancer patient-derived organoids, provided by colleagues at the Dana-Farber Cancer Institute. Organoids are three-dimensional cultures of cells that bear a greater phenotypic resemblance to the architecture and biology of cell types *in vivo* than traditional two-dimensional cell lines [25]. Importantly, organoids also represent pure cancer cell populations, enriched by culture conditions that favor pancreatic tumor cells but not the stromal compartment. Since organoids are pure tumor cell populations, we are thus able to study the mutational landscape, transcriptome, and immunopeptidome of human pancreatic cancer cells without the confounding influence of stromal cells.

With a broadened bioinformatic lens focused on these patient-derived samples, we found that pancreatic cancer cells bear a larger somatic mutation burden and predicted neoepitope load than previously estimated by prior studies of bulk tumor samples. We also predicted shared cancer-restricted epitopes derived from frameshifts in tumor suppressor genes and retained introns. Using immunopeptidomics, we empirically validated the presentation of epitopes derived from retained introns, lncRNAs, 5’ UTRs, missense mutations, a frameshift in the tumor suppressor gene *SMAD4*, and other sources. We identified a large subset of these as PDAC-specific by scanning the transcriptome and immunopeptidome of hundreds of normal tissue specimens derived from over 30 different tissues and finding no evidence of functional expression levels in essential normal tissues. Finally, we demonstrate the recognition of three of these PDAC-specific MHC class I-bound peptides by $CD8^+$ T cells derived from the blood of healthy donors, indicating they can elicit an endogenous immune response in HLA-matched individuals. Collectively, this work elucidates a broadened class of immunotherapy targets for pancreatic cancer, pinpointing new clinical directions for this notoriously recalcitrant disease.

Synopsis and Outlook

While the history of pancreatic cancer is bleak, new technologies and biological insights continue to emerge and reveal new ways of targeting cancer and rationally combining novel treatments. These include updated views of the mutational landscape, transcriptome, and proteome of pancreatic cancer cells, revealed in part by new cancer cell culturing systems, next-generation sequencing techniques, and increasingly sensitive immunopeptidomics [128].

My graduate work has focused on the discovery of new targets for this disease, which I hope will be applied towards new approaches in immunotherapy for treating pancreatic cancer. Preclinical work and certain clinical data, such as the approval of checkpoint inhibitors for mismatch repair deficient pancreatic cancers and the recent phase I trial employing neoantigen vaccines, suggest that next-generation immunotherapies possess great potential for treating this disease. The targets identified in Chapter 2 could be amenable to several of these approaches, including adoptive cell therapy, cancer vaccines, and peptide:MHC centric CARs and bi-specific antibodies described earlier in this chapter. In Chapter 4, I will outline several of these applications in greater detail.

Chapter 1, Part 2: Modeling the cancer genome with genome editing technology.

The most important goal of all cancer research is to find new ways to diagnose, detect, or treat cancer. Before entering clinical trials, therapeutics must be evaluated through preclinical testing with animal models. These models enable scientists to ethically evaluate the benefits and side effects of a new therapeutic before risking administration to human patients.

In cancer research, mice have long been the gold-standard animal model. Many of the most popular mouse models of cancer are genetically engineered so that they closely approximate features of the human disease. This is commonly done by engineering mutations that drive cancer in humans into the cancer-related genes of mice. By using these models in a controlled fashion, investigators can induce cancer in specific tissues (e.g., lung or pancreas) of mice and thereby study how the disease originates, evolves, and responds to experimental therapeutics. While valuable, genetically engineered mouse models take months or years to develop. Thus, the generation of new mouse models struggles to keep pace with the discovery of new cancer driver mutations occurring on a regular basis (see Chapter 3). These mutations, alone and in diverse combinations, produce dramatic differences in the biology and therapeutic vulnerabilities of tumors in different patients. Generating mouse models that recapitulate these effects is a critical step toward better understanding them.

In this section, I will describe a brief history of genetically engineered mouse models and genome editing technology. I will then introduce the work presented in Chapter 3, in which my colleagues and I developed a new genetically engineered system capable of installing a broad spectrum of different cancer mutations into various tissues of adult mice. This technology enables modeling of most human cancer-associated mutations, without requiring the time-consuming construction of a new mouse for each mutation of interest.

Cancer as a genetic disease

Human cancer is driven by sporadic genetic alterations, including single nucleotide variants, insertions, deletions, chromosomal losses and gains, and chromosomal rearrangements. In solid epithelial malignancies, driver mutations often co-occur in sets of two or more, with each set typically conferring both a gain-of-function in an oncogene and a loss-of-function in a tumor suppressor gene. Major consortium efforts like The Cancer Genome Atlas (TCGA) and the Pan-Cancer Atlas of Whole Genomes (PCAWG) have performed deep sequencing of cancer exomes and genomes across thousands of cancer patients and dozens of cancer types. These efforts continuously identify new cancer driver mutations, a catalog that now numbers in the thousands [129].

Driven by the insights of next-generation sequencing, precision oncology paradigms have emerged to tailor patient treatment options according to the specific genotype of each patient's tumor. This paradigm is best exemplified by the use of targeted therapies. These drugs work by directly interfering with the products of driver mutations and related cell signaling pathways. A famous example is the landmark drug imatinib, a small molecule engineered to specifically inhibit the protein encoded by the *BCR-ABL* gene fusion, a driver event present in more than 90% of cases of chronic myelogenous leukemia (CML). CML was among the most lethal cancer types prior to the advent of imatinib. Now, more than 90% of patients will survive five years after diagnosis without tumor recurrence, a stunning therapeutic success.

Since the resounding success of imatinib, dozens of drugs targeting mutant oncogenes have been approved as first-line therapies by the Food and Drug Administration. These include EGFR inhibitors like osimertinib, KRAS inhibitors like sotorasib, and others. While most targeted therapies are not nearly as effective as imatinib, they still often extend patient survival and therapeutic response rates relative to preexisting standards of care.

Despite the clinical benefit of targeted therapies, solid tumors frequently develop resistance to these treatments. The source of resistance is often genetic, and next-generation sequencing efforts have unveiled hundreds of secondary mutations that can diminish a targeted drug's effect. Epigenetic changes and copy number alterations are also known to promote resistance [130], but these are less frequently responsible than secondary mutations in the targeted oncogene or other genes [131]. Interestingly, secondary mutations are often specific to tumors treated with targeted therapies and may never be observed in pre-treated primary cancers. For example, the T315I mutation in *BCR-ABL* is responsible for ~11% of clinical resistance to imatinib [132], and ~50% of non-small cell lung cancer tumors that acquire resistance to EGFR inhibitors bear the secondary T790M mutation in *EGFR*. The discovery of these mutations has prompted the development of second-generation inhibitors, like *BCR-ABL*-targeted ponatinib [133]. Molecules like ponatinib can successfully inhibit or prevent the development of oncogenes bearing the secondary mutations, thereby prolonging patient survival even further [134].

While rationally informed by human genetics, these targeted therapies require preclinical systems to evaluate their efficacy and safety before administration to human patients. In the next section, I will discuss systems for preclinical modeling of primary and secondary mutations that facilitate the testing of new therapies.

Brief historical note

While genetically engineered mouse models are the focus of this thesis, there are many other methods for modeling cancer in mice. For example, cell-derived xenograft (CDX) models, popular in the pharmaceutical industry, work by culturing patient tumor cells *in vitro* and then transplanting them into a mouse host. These models are typically fast, with tumors often developing within a matter of days, depending on the transplanted tissue and tumor type. The resulting tumor is comprised of human cancer cells, reflecting the cancer driver and passenger mutations of the human disease. This affords investigators the option of rapidly testing therapies targeted to tumors with specific genetic alterations (e.g., a certain driver mutation, gene fusion, etc.). However, CDX models require immunodeficient hosts, mice that have had their immune system disabled. This is necessary to enable engraftment and tumor outgrowth, as the immune system would otherwise mount a response against the plethora of human protein sequences not incorporated into murine central tolerance. Thus, CDX models preclude the testing of most immunotherapies and their combinations.

Immunocompromised animals further impede the discovery of unanticipated therapeutic mechanisms. For example, researchers studying KRAS^{G12C} inhibitors unexpectedly discovered that mutant KRAS inhibition promotes reshaping of the immune milieu (e.g., increase in CD8+ T cell infiltration, reduction in immunosuppressive populations). This effect was highly synergistic with immune checkpoint inhibitors in an immunocompetent mouse model of colon cancer [135]. KRAS^{G12C} inhibition and ICIs are now being pursued as combination treatments in clinical studies [136], highlighting the importance of capturing the tumor-immune system interplay in response to therapy.

In another class of mouse cancer models, researchers expose mice to carcinogens, such as tobacco smoke or other chemical compounds, to induce somatic mutations. Carcinogen-induced models thereby enable direct evaluation of substances hypothesized to promote the development of cancer. They can also recapitulate tumor phenotypes that may be specifically linked to a carcinogen-specific exposure, such as an increase in G:C to T:A transversions that are associated with a number of carcinogens. These mice also develop a significantly higher mutational burden than genetically engineered mice [137], which tend to have dramatically lower mutational burden than human tumors. However, carcinogen-induced models do not enable preselection of the driver mutation, challenging efforts to study therapeutic vulnerabilities related to specific driver genes and/or mutations.

Modeling cancer with genetically engineered mice

In contrast to the sporadic nature of carcinogen-induced models, genetically engineered mouse models (GEMMs) enable the study of specific driver gene mutations in an immunocompetent system. This is typically accomplished by encoding driver mutations within genetic modules present in the mouse germline. Hundreds of these germline alleles exist, and they collectively enable cancer initiation across a wide variety of genetic and physiological contexts. These alleles can be classified as either conditional or constitutive.

Constitutive alleles encode mutations that inactivate the gene of interest in every cell of the mouse. In a classic example, Donehower and colleagues developed a constitutive allele by encoding a germline deletion of *Trp53* [138]. These investigators observed that the resulting mice developed normally, even for a few months after birth. However, the mice exhibited a high propensity for spontaneous tumor development in tissues throughout the body, reflective of p53's critical role in regulating the cell cycle of many tissues. Most of these mice developed tumors by six months, an age at which tumors are rarely found in wild-type animals. This study thus provided early, convincing proof of p53's role as a tumor suppressor *in vivo*.

However, the utility of constitutive alleles is limited, as is their fidelity to the evolutionary paths experienced by adult human tumors. Whereas human tumors are largely thought to originate from sporadic mutation in a specific adult tissue [35], constitutive alleles initiate cancer early during mouse development, affecting a far broader swath of cell types and organs than the human analog. This necessarily limits studies of tissue-specific gene function in a specified adult tissue [139].

In contrast, conditional alleles require a secondary event for activation of the mutation. This enables researchers to control the location and timing of tumor development. Two prominent conditional alleles include the Lox-STOP-Lox-*Kras*^{G12D} allele (*Kras*$^{LSL\text{-}G12D/+}$) and the *Trp53* floxed allele (*Trp53*$^{flox/+}$) [140]. Both alleles are controlled by the Cre-lox system. This system has two central components: Cre recombinase and loxP DNA sequences, both derived from bacteriophages. In these viruses, Cre maintains genomic integrity and function by recognizing loxP sequences present within the viral genome. These sequences have not been detected in any other known genome beyond the P1 bacteriophage [141]. When installed in the mouse genome, loxP sequences enable Cre-mediated excision, inversion, or translocation of genetic sequences. The positioning and orientation of the two loxP sites dictate which of these events will occur.

In the *Kras*$^{LSL\text{-}G12D/+}$ allele, a “lox-stop- lox” (LSL) cassette places loxP sites on either side of a “STOP” module located upstream of the first exon in *Kras*, which has been modified to encode the oncogenic G12D mutation [142]. This module prematurely terminates transcription of *Kras*^{G12D}, but not the wild-type allele. Cre-mediated recombination excises the sequence between the loxP sites, enabling expression of the mutant oncoprotein. In contrast, the *Trp53*$^{flox/+}$ allele contains loxP sites flanking exons 2-10 of *Trp53*. Thus, functionally wild-type p53 is expressed until exposure to Cre. Cre then deletes either one or two *Trp53* alleles, depending on whether the mice are heterozygous (*Trp53*$^{flox/+}$) or homozygous (*Trp53*$^{flox/flox}$). In mice with these alleles, Cre can be supplied exogenously through viral delivery.

In the classical “KP” mouse model, mice harbor both *Kras*$^{LSL\text{-}G12D/+}$ and *Trp53*$^{flox/flox}$, and Cre is often delivered to the lung via lentivirus or adenovirus vectors. Intratracheal injection enables viral infection of mouse lung cells, followed by lentiviral integration into the host genome and subsequent transcription and translation of Cre recombinase. These mice develop dozens to hundreds of lung tumors, bearing classical histopathological features and evolutionary progression of lung adenocarcinoma seen in humans. There are other inducible systems beyond Cre-lox, such as Tet-on and Tet-off. In these systems, expression of a transgene is regulated by a transcriptional activator that can be switched on or off by exposure to tetracycline or a similar derivative.

Cre can also be generated endogenously. For example, in the KPC (*Kras*$^{LSL\text{-}G12D/+}$*; Trp53*$^{LSL\text{-}R172H/+}$*; Pdx1-Cre*) mouse model, Cre is encoded within the mouse genome. Cre expression is regulated by the pancreatic and duodenal homeobox 1 (Pdx1) promoter [143,144]. The Pdx1 promoter is bound by the Pdx1 transcription factor, which regulates the transcription of a variety of host genes essential for the embryonic development of the mammalian pancreas. In KPC mice, the pancreatic-specific expression of Pdx1 enables formation of adenocarcinoma in the pancreas but not other tissues.

In the KPC model, activation of the heterozygous *Trp53*$^{LSL\text{-}R172H/+}$ is uniformly followed by loss of heterozygosity (LOH) of the other allele, which is sufficient to drive tumorigenesis cooperatively with *Kras*$^{LSL\text{-}G12D}$. Canonically, tumor suppressor genes like *Trp53* require loss or functional inactivation of both alleles to sufficiently promote tumorigenesis. However, studies using the KPC model have shown that certain point mutations like *Trp53*^{R172H} can also exercise a dominant-negative or even gain-of-function effects. Studies comparing traditional KPC mice to *Kras*$^{LSL\text{-}G12D/+}$*; Trp53*$^{flox/+}$*; Pdx1-Cre* mice found that only KPC mice developed metastases [32].

This suggests that p53 point mutations can confer neomorphic properties that promote tumor development beyond the simple loss of cell cycle checkpoints. This result is consistent with similar observations comparing p53 point mutations to p53 deletions in other mouse models of cancer [145].

Other mouse models of PDAC incorporate *Cdkn2a*$^{-}$ alleles. *Cdkn2a* is an unusual gene that encodes for two different proteins with completely distinct amino acid sequences. This is made possible by alternative transcriptional start sites. These generate two distinct transcripts, p16(INK4A) and p14(ARF), that are distinguished by different first exons but that share the same downstream exons [146]. p16(INK4A) and p14(ARF) encode different reading frames, enabling the production of distinct protein products. Both p16(INK4A) and p14(ARF) function as tumor suppressors: p16(INK4A) functions as a cyclin-dependent kinase inhibitor that can suppress cell cycle progression, and p14(ARF) inhibits the mouse double minute homolog 2 (MDM2), a negative regulator of *Trp53* [35]. *CDKN2A* mutations in pancreatic cancer patients typically inactive both isoforms, though a small subset have been noted that only inactivate p16(INK4A). *CDKN2A* is thought to be inactivated in >90% of pancreatic tumors, by way of mutation, gene deletion, or promoter hypermethylation, which frequently silences *CDKN2A* expression in pancreatic cancer [5].

Kras$^{LSL\text{-}G12D/+}$*;Trp53*$^{flox/+}$*; p16*$^{Ink4a\text{-}/+}$*; Pdx1-Cre* mice developed tumors much more quickly (15 weeks compared to 22 weeks) than mice with only the *Kras*$^{LSL\text{-}G12D/+}$*;Trp53*$^{flox/+}$ alleles [147]. This indicates that loss of both *Trp53* and *p16*^{INK4A} accelerates tumor development, consistent with the finding that both genes are inactivated in the majority of pancreatic cancers [5].

The *Kras*$^{LSL\text{-}G12D/+}$ and *Pdx1-Cre* alleles have together been combined with *Trp53*$^{flox/flox}$, *p16*$^{Ink4a\text{-}flox/flox}$, or *Smad4*$^{flox/flox}$ alleles in different models. Interestingly, these show distinct phenotypes.

For example, models based on either and *p16*$^{Ink4a\text{-}flox/flox}$ or *Trp53*$^{flox/flox}$ succumb to disease within two or three months, respectively. Only *Trp53*$^{flox/flox}$ animals show evidence of metastases, and the primary tumors also adopt a more cystic phenotype [148]. In contrast, similar models based on *Smad4*$^{flox/flox}$, the other major tumor suppressor gene frequently inactivated in human PDAC, survive for more than nine months. These mice also exhibit a higher rate of intraductal papillary mucinous neoplasms, neoplastic lesions distinct from PanINs that can precede the development of pancreatic adenocarcinoma [149]. Together, these foundational studies suggest that distinct combinations of driver gene alterations dictate differences in the outcome and phenotype of pancreatic cancers.

All of the described alleles in principle can be combined through breeding programs, thus enabling complex combinations of cancer mutations. Such breeding programs, however, are time-consuming both to conduct and to maintain. The combined process of establishing breeding pairs and subsequent murine gestation takes months. These programs also require regular experiments to confirm the genotype of each individual mouse. Furthermore, existing alleles encoding a particular mutation may not be amenable to the scientific question of interest. For the KPC model and its variants, most pancreatic cells in the mice are subject to activation of oncogenic *Kras* and deletion of a tumor suppressor gene. Accordingly, *Kras*$^{LSL\text{-}G12D/+}$*;Trp53*$^{flox/flox}$*; p16*$^{Ink4aflox/flox}$*; Pdx1-Cre* mice develop widespread tumors throughout their pancreas and succumb to this tumor burden before later-stage events like metastases can be examined. With this model, it is therefore impossible to examine how the temporal ordering of *Trp53* and *p16*^{Ink4a} inactivation influences tumor progression and patterns of metastasis.

While GEMMs have deepened our knowledge of cancer biology and enabled testing of novel therapies, their success also underscores the need for fine-tuned selection of driver mutations.

For example, all 20 GEMMs of pancreatic cancer reviewed by Westphalen and Olive (2012) were based on *Kras*^{G12D}. Models based on *Kras*^{G12C} and *Kras*^{G12R} were developed only recently [150]. While *KRAS*^{G12D} is the most abundant PDAC mutation (affecting 30-40% of patients), different *Kras* alleles have varying oncogenic propensities, rooted in differences in intrinsic GTPase activity and differential mutant protein affinity to the RAF kinases [151].

To this point, Zafra *et al*. (2020) examined differences among G12D, G12C, and G12R in the setting of *Kras*$^{LSL\text{-}mutation/+}$; *p48-Cre* (*KC*) models. (p48, like Pdx1, is exclusively expressed in the pancreata of mice, though its expression is thought to be limited to acinar cells.) In *KC-G12D* mice, more than 90% of pancreatic tissue exhibited a PanIN phenotype by 12 weeks, with less than 10% of the organ defined by normal acinar tissue. In contrast, *KC-G12C* mice exhibited a PanIN burden of only ~50% by 12 weeks. *KC-G12R* mice showed *no* evidence of PanINs, only exhibiting lesions arrested in the acinar-to-ductal metaplasia (ADM) state, evidenced by co-expression of the acinar cell marker CPA1 and the ductal marker KRT19. ADM is thought to precede the formation of PanINs for tumor lineages that originate from acinar cells [21]. This difference in *KC-G12R* mice endured even after a year. These results elegantly highlight how mutational differences in driver genes, even those as subtle as distinct alterations in the same codon, dictate dramatic differences in oncogenic potential. Such differences can only be captured by a comprehensive study of these mutations *in vivo*.

As another example, most pancreatic cancer patients bear point mutations in *TP53* other than the two mutations (*TP53*^{R175H} and *TP53*^{R273H}) represented in current GEMMs. The other mutations remain understudied in murine models of the disease. This gap is critical, especially in light of research showing that many genes, including *TP53*, often harbor point mutations that confer different properties relative to loss-of-function truncations, including gain-of-function effects [152].

For example, Schulz-Heddergott and colleagues demonstrated that *TP53*^{R248Q} exhibits a gain-of-function effect by hyperactivating the JAK2/STAT3 pathway, leading to more aggressive tumor progression in models of colon cancer [153]. This mutation is also common in pancreatic cancer, but these *in vivo* observations remain untested due to a lack of pancreatic cancer mouse models and human cell lines harboring the mutation.

While traditional GEMMs have shed tremendous light on the nature of several cancer mutations, the technology struggles to keep pace with the thousands of driver mutations identified from next-generation sequencing efforts. This impediment prompted the adoption of a new cancer modeling paradigm, one driven by the advent of targeted genome editing technologies based on CRISPR-Cas9. This approach enables rapid, combinatorial generation of cancer driver mutations in somatic cells, without the need for a germline allele. These models dramatically accelerate our ability to study novel cancer mutations.

Engineering the cancer genome with CRISPR-Cas9

CRISPR-Cas9 is a powerful genome-editing technology that originates from microbes. The discovery and development of this technology has been extensively reviewed in numerous articles [154,155], but the seminal advancements are worth describing here. In the 1980s, researchers observed arrays of repeated DNA sequences in the genomes of *Escherichia coli* [156]. These mysterious repeats did not receive dedicated attention until the 1990s, when Dr. Francisco Mojica reported the discovery of clusters of palindromic repeats in the genome of *Haloferax mediterranei*, an archaebacteria native to the marshes of Spain where Mojica conducted his research [157]. Mojica later identified these repeats, now known as CRISPR (clustered regularly interspersed short palindromic repeats), in thousands of organisms throughout bacterial and archaebacterial phyla. Critically, these sequences were present in the genomes of viruses that infect these species, suggesting transmission of genetic sequences among microbes and pathogens [158]. Across species, certain genes appeared physically linked to the presence of CRISPR arrays, often residing as adjacent loci [159]. The function of these CRISPR-associated genes (*Cas*) remained obscure for a few years.

In 2007, Barrangou *et al*. showed that CRISPR-associated *cas* genes encode specialized nucleases that facilitate excision of short genetic sequences from viral pathogens [160]. They also showed that Cas nucleases can incorporate these sequences into the bacterial host genome, establishing them as the repeat sequences observed in CRISPR arrays. Upon re-infection, the integrated repeats are transcribed as short RNA sequences that bind with additional Cas proteins, forming complexes that target and destroy the invading viral genome. CRISPR thus functions as an adaptive immune system for vast numbers of microbial species.

Harnessing this fundamental knowledge, pioneering work from the laboratories of Emmanuelle Charpentier, Feng Zhang, Jennifer Doudna, and George Church later demonstrated re-engineering of these components for intentional modification of the genomes of bacteria and mammals [161–163]. Critically, Jinek *et al*. (2012) engineered the first single guide RNA (sgRNA, or, gRNA) from two CRISPR RNA components and demonstrated programmable targeting of the bacterial genome using Cas9 [161], a CRISPR-associated endonuclease. Cas9 cleaves DNA into a double-stranded break upon binding to a NGG motif adjacent to the protospacer targeted by the gRNA. (This protospacer-adjacent motif is called a PAM.) This technique was adapted by contemporaneous publications that demonstrated targeted genome editing in mammalian cells, including human cells [162,163]. Collectively, these studies ushered in a revolution for biological research, based on a technology that enabled facile and broadly applicable targeting of the genome.

Thousands of studies have since harnessed CRISPR-Cas9 to probe gene function through targeted genetic knockouts. The DSBs induced by CRISPR-Cas9 are repaired by natural DNA repair pathways present within cells. The nature of DSB repair is partly dependent on both the state and type of the targeted cell, but in most contexts, DSBs are readily repaired by non-homologous end joining (NHEJ). NHEJ-mediated repair often results in the stochastic formation of small insertions or deletions (indels) in the sequence extending from the breakpoint [164]. When Cas9 is targeted to a coding region of a gene, the subsequent DSB repair frequently results in an out-of-frame indel, constituting a frameshift mutation that disrupts the encoded protein. This enables a powerful reverse genetics approach, whereby gene function can be understood through knockout and examination of the resulting phenotype.

This approach was first applied to cancer modeling *in vivo* in 2014, when work from our laboratory applied Cas9 and gRNAs targeted to tumor suppressor genes to facilitate tumorigenesis in adult mouse tissues. This was first demonstrated by Xue *et al.* (2014), who initiated liver tumors by delivering plasmids encoding Cas9 and sgRNAs targeting *Pten* and *Trp53* to otherwise wild-type mice [165]. Later that same year, Sánchez-Rivera, Papagiannakopoulos, and colleagues published a lentivirus delivery system based on an all-in-one vector capable of encoding Cas9, Cre recombinase, and a gRNA [166]. When applied to $Kras^{LSL\text{-}G12D/+}$ mice, this vector enables simultaneous activation of oncogenic *Kras* and disruption of a tumor suppressor gene targeted by Cas9 and the encoded gRNA. Harnessing this technology, our laboratory generated lung adenocarcinomas with modular knockout of different tumor suppressor genes. This enabled rapid interrogation of tumor suppressor genes such as *Apc*, hypothesized to impact lung tumorigenesis but not yet studied in GEMMs. Reminiscent of findings from pancreatic cancer GEMMs, researchers from our laboratory found that targeting different tumor suppressor genes yielded differences in histopathology, cell pathway signaling, and overall tumor burden in the lung. This study dramatically accelerated preclinical cancer modeling using CRISPR-Cas9, bypassing the need for generation of new alleles for targeted gene deletion.

Most cancer driver mutations represent single nucleotide variants (SNVs), not indels. Platt *et al.* (2014) were the first to demonstrate *in vivo* CRISPR modeling of a cancer SNV. To accomplish this, they installed germline alleles of Cas9 within *Rosa26*, a safe harbor locus where genes can be stably expressed without disrupting any apparent cell phenotypes [167]. Importantly, installation of foreign genes in the mouse genome subjects the resulting protein to central tolerance, even in the context of an LSL cassette, like *Rosa26*$^{LSL\text{-}Cas9}$ [168]. Thus, germline alleles can circumvent immune responses against Cas9, which is documented in both mice and humans [169,170]. Germline alleles also avoid certain delivery challenges. Cas9 is a relatively large gene, and common viral vectors like adeno-associated virus (AAV) and lentivirus have a limited DNA packaging capacity [171]. By installing Cas9 germline, more space is available for the delivery of other cargo. To this point, Platt *et al.* (2014) demonstrated delivery of multiple TSG-targeted gRNAs, enabling a multiplexed knockout of tumor suppressor genes within the context of the same tumor. In addition to TSG-targeted gRNAs, they also delivered a long genetic template for homology-directed repair (HDR). HDR is another major DSB repair pathway employed by mammalian cells. Harnessing HDR, researchers can design specific gene edits by encoding an HDR DNA template that bears homology to both regions flanking the targeted Cas9 cleavage position. The intervening sequence encodes genetic alterations of interest. At low efficiency, HDR will integrate the donor template sequence into the genome during the repair process, thereby permanently installing a designed edit. Platt *et al.* showed that this could be used to install *Kras*^{G12D} in otherwise wild-type *Rosa26*^{Cas9} mice. This approach has subsequently been adopted to model a subset of point mutations in a variety of oncogenes [172].

Though useful, the utility of HDR for precision genome editing is limited by low efficiency [173,174]. HDR activity is generally limited to dividing cells, especially during the S and G2 phases of the cell cycle [175], and its efficiency is variable across cell types. Even in cell types with efficient HDR, NHEJ is preferentially engaged for DSB repair. This has been demonstrated with DNA sequencing of targeted loci, which often reveal low-frequency point mutations in the presence of a higher frequency of indels [175].

This feature is prohibitive for some cancer applications, but not all. For example, activating point mutations in oncogenes confer a selective advantage that typically promotes increased cell division. Thus, tumors resulting from an HDR-mediated oncogene mutation should exhibit purely the point mutation, without evidence of indels [176]. However, point mutations are also common and functionally active in tumor suppressor genes, especially *TP53* [152]. Like point mutations, frameshifting indels in tumor suppressor genes are subject to positive selection. Thus, HDR-based edits in tumor suppressor genes are likely to be diluted by a higher frequency of indels, thus impeding the study of phenotypes linked to TSG point mutations.

The advent of base editing

With these and other limits of HDR in mind, second-generation CRISPR technologies were developed capable of precision genome editing without DSB intermediates. For example, CRISPR-Cas9 base editors are fusion enzymes capable of mediating precise single nucleotide substitutions without forming DSBs and without need for a donor DNA template. This technology employs a catalytically impaired Cas9 in which one or both nuclease domains are functionally inactivated. In the case of a Cas9 nickase, inactivation of only one domain enables a single-stranded DNA cut, while inactivation of both domains (dCas9 variants) avoids DNA cleavage altogether. Two primary classes of base editors include cytidine base editors (CBEs) and adenine base editors (ABEs). In these classes, the catalytically impaired Cas9 is fused to either an adenine or cytidine deaminase. CBEs further incorporate a uracil glycosylase inhibitor as a third enzyme in the complex. Both ABEs and CBEs work by deaminating the target nucleotide, thereby producing an inosine (ABEs) or uracil (CBEs) intermediate. The intermediate base pair represents a mismatch detectable by DNA mismatch repair pathways, which then resolve the mismatch and produce the desired single nucleotide substitution.

Base editors are extremely efficient, enabling targeted nucleotide conversions often at efficiencies of 60% or higher in mammalian cells [177]. High editing efficiency coupled to the benefits of avoiding DSBs led to rapid adoption of base editing by the cancer modeling field. Seminal work by Zafra and colleagues demonstrated somatic editing *in vivo* through exogenous delivery of base editors to the liver of mice [178]. In this study, the investigators delivered plasmids encoding a reengineered CBE, a sgRNA targeting the S45F mutation in the *Ctnnb1* oncogene, and a *Sleeping Beauty* transposon–based *Myc* cDNA. Together, these components generated liver tumors with the intended mutations in all recipient mice.

While base editing has expanded the precision of mouse cancer modeling, it is limited by the scope of achievable mutations. Traditional CBEs and ABEs are limited to transition mutations, and they can also suffer from bystander editing, which occurs when multiple cytidines or adenines co-exist within the base editing window. This can prohibit engineering the desired amino acid substitutions in certain contexts. While recent advances in transversion base editing will expand base editing scope, transversion base editors currently suffer from additional undesired on-target effects that are challenging to predict [179].

The advent of prime editing

With these considerations in mind, Anzalone and colleagues developed a third-generation technology called prime editing [180]. Prime editors are comprised of a Cas9 nickase fused to a reverse transcriptase (RT). Instead of a traditional 20-nucleotide gRNA, prime editors complex with longer prime editing guide RNAs (pegRNAs). pegRNAs encode a primer binding site (PBS) and a reverse transcriptase template (RTT), in addition to the targeting spacer and scaffold sequence that comprise traditional CRISPR gRNAs. This innovation enables a unique stepwise process in which 1) the prime editor enzyme is targeted to a specific genomic locus and induces a single-stranded DNA nick, 2) the 3' PBS hybridizes to one of the subsequent flapping DNA strands, 3) the transcriptase reverse transcribes the sequence encoded by the RTT, and 4) subsequent DNA repair pathways resolve the resulting mismatch, ultimately leaving an edited sequence of nucleotides encoded by the RTT. Prime editing thus enables a full suite of small somatic mutations, including all types of single nucleotide substitutions, multi-nucleotide substitutions, and small insertions and deletions. Beyond editing versatility, prime editing also benefits from high editing purity, with intended edits typically far outnumbering unintended indel byproducts.

While versatility is broad and purity is high, prime editing efficiency is typically low, much lower than single nucleotide conversion rates observed with base editors. To overcome this, several techniques can be employed. For example, varying the combination of PBS and RTT lengths can yield order of magnitude increases in editing efficiency. Furthermore, the selection of a protospacer near the targeted locus can influence efficiencies, as can the choice of the pegRNA scaffold sequence [181]. Engineered pegRNA motifs (epegRNAs) can also be appended to the 3’ extension of pegRNAs, which generally enhances efficiency by 3-4-fold through protecting the pegRNA from degradation [181].

As a more complex solution, pegRNAs can also be coupled with traditional CRISPR sgRNAs designed to induce a nick on the unedited DNA strand. This biases DNA repair toward incorporation of the prime edited nucleotides during mismatch repair, though this generates a higher rate of indels by increasing the likelihood of a DSB. Nicking gRNAs specific to the prime edited sequence circumvent this limitation, as the second nick will be induced only after the first has been resolved into the heteroduplex DNA intermediate, minimizing DSB formation [180]. Finally, pegRNAs can be further optimized by encoding benign or silent edits adjacent to the intended edit that can promote evasion of mismatch repair [182].

While these and other strategies frequently enable prime editing efficiencies of 20-50% in mammalian cells, they underscore an important caveat associated with this system. Extensive pegRNA optimization can be laborious and require testing dozens of designs to identify a highly efficient pegRNA. However, this caveat may be an acceptable tradeoff for researchers seeking to model mutations that are not amenable to base editing or CRISPR-Cas9-based HDR.

Preclinical modeling of cancer in mice with prime editing

Though prime editing offers many benefits for modeling cancer, only a handful of studies have pursued this, and many are limited to *in vitro* applications [183]. To my knowledge, the only *in vivo* studied published to date that employed prime editing for cancer used it to study liver cancer [184]. Similar to the base editing study described previously, these investigators initiated liver cancer by hydrodynamic co-injection of PE2 (split across two plasmids), *Ctnnb1*-targeted pegRNAs, and *Sleeping Beauty* transposon–based *Myc* cDNA. When applied using a dual-AAV delivery approach, the authors noted unacceptably low *in vivo* editing efficiencies and intolerable indel rates, which they posit could be linked to the reduced efficiency incurred by requiring the enzyme to be split across two AAV vectors.

The size of the prime editor enzyme may explain the lack of additional studies using prime editing to model cancer in mice. The gene encoding the traditional PE2 enzyme is approximately 6.3 kilobases in size [185]. This exceeds the packaging capacity of AAVs (~4.8 kilobases) and nearly exceeds that of lentivirus (~8 kilobases). Furthermore, the immunogenicity of Cas9 is well-documented, and the viral-derived RT may also harbor potent antigens [170]. The size and potential immunogenicity of the prime editor enzyme thus impose major obstacles to leveraging prime editing for modeling cancer in mice.

Prime editor GEMMs to model a broad spectrum of somatic mutations

With these considerations in mind, my colleagues and I developed a genetically engineered mouse model with the PE2 enzyme encoded within the *Rosa26* locus. As will be described in Chapter 3, this system enables high efficiency prime editing *in vivo* without exogenous delivery of the prime editor enzyme. This circumvents challenges associated with both delivery and potential immunogenicity, which has been documented with exogenous delivery approaches using the PE2 enzyme [170]. With this model, we demonstrated *ex vivo* and *in vivo* installation of a variety of genetic lesions, including precise insertions, deletions, and single and dinucleotide substitutions. In all cases, we observe negligible indel byproducts, including in the tumor suppressor gene *Trp53*. This enabled us to select pure populations of cells harboring *Trp53* mutations that are frequent in pancreatic cancer patients but understudied in preclinical models of the disease.

Furthermore, we demonstrated rapid generation of new pancreatic organoid models defined by distinct $Kras^{G12D}$ or $Kras^{G12C}$ mutations. Using these models, we found that pancreatic cancer organoids are uniformly more sensitive to KRAS^{G12D} or KRAS^{G12C} inhibitors when they are combined with the EGFR inhibitor, gefitinib. *in vivo*, we observed *Kras* allele-specific differences in tumor initiating potential in both an orthotopic transplant and autochthonous setting, nuancing prior observations made with LSL GEMMs. In the lung, we further find that prime editor GEMMs are capable of initiating lung adenocarcinoma even with lower efficiency pegRNAs, highlighting the broad utility of this system. We expect this model to accelerate preclinical functional studies of the vast landscape of untapped driver mutations described earlier in this chapter. By enabling rapid gene function studies *in vivo*, this system can both facilitate the discovery of new therapeutic vulnerabilities and the testing of new targeted therapies

Synopsis and outlook

Cancer is driven by genetic alterations that co-occur in thousands of combinations across different cancer patients. The precise effects of each mutation, as well as the interplay of these effects across multiple drivers, is a critical determinant of cancer progression and disease outcome. The technology presented in Chapter 3 will help fill critical gaps in our knowledge of how specific driver mutations and combinations thereof affect the progression, phenotype, and vulnerability of tumors. This will be accomplished in large part with prime editor GEMMs for the precise, simultaneous installation of multiple driver mutations (e.g., combined SNVs across an oncogene and tumor suppressor gene), as well as the combination of the *Rosa26*^{PE2} allele with preexisting technologies such as the *Kras*$^{LSL\text{-}G12D}$ allele.

Beyond classical driver mutations, I envision this system will be harnessed to examine historically unexplored patterns of mutations. For example, several studies highlight the contributions of noncoding mutations in driving cancer, such as historically overlooked silent mutations that co-occur with missense mutations [186]. Other studies pinpoint mutations in regulatory sequences that can affect gene expression programs in pancreatic cancer cells [187]. Through integration of long insertions, prime editing could also enable co-installation of a custom neoepitope sequence, enabling sophisticated cancer immunology studies. One potentially interesting application would involve screening cancer hotspot sites in murine driver genes for mutations that can both drive cancer and generate a neoepitope. This phenomenon is well-established in a subset of human cancers but understudied in animal models [80,85]. Combination of *Rosa26*^{PE2} with humanized *HLA* alleles could also enable *in vivo* testing of therapeutics targeted against such neoepitopes observed in human cancer.

The combination of multiple genetic alterations will become more challenging as the number of edits increase in each application. Low basal prime editing rates present an obstacle to introducing multiple intended edits. For secondary mutations that induce a neutral or negatively selected event (e.g., neoepitope installation), some tumors may grow out that bear the driver mutation but lack the second event. In some cases, this may be convenient, e.g., to study intratumoral heterogeneity. In other cases, this may not be an issue if both edits are necessary for tumor formation (e.g., combined oncogene activation and loss of a tumor suppressor gene).

The challenges associated with multiplexed editing experiments can be overcome with optimization of editing parameters. This includes various strategies for pegRNA optimization described and second-generation prime editor GEMMs based on more efficient PE enzymes, such as PEMax [182]. It is also conceivable to deliver custom pegRNAs in which the edit necessary for tumor formation (e.g., a *Kras*^{G12} mutation) are programmed by pegRNAs intentionally less efficient than those programming neutral edits. This would increase the probability of resulting tumors containing both intended edits. These and other techniques will expand the capabilities of prime editor GEMMs and ultimately reduce the need to generate and maintain GEMMs based on traditional alleles.

References

1. Pancreatic cancer statistics. *WCRF International* https://www.wcrf.org/cancer-trends/pancreatic-cancer-statistics/.
2. Prognosis. *Hirshberg Foundation for Pancreatic Cancer Research* https://pancreatic.org/pancreatic-cancer/about-the-pancreas/prognosis/.
3. Freelove, R. & Walling, A. D. Pancreatic Cancer: Diagnosis and Management. *Am. Fam. Physician* **73**, 485–492 (2006).
4. Rhim, A. D. *et al.* EMT and dissemination precede pancreatic tumor formation. *Cell* **148**, 349–361 (2012).
5. Kleeff, J. *et al.* Pancreatic cancer. *Nat. Rev. Dis. Primer* **2**, 1–22 (2016).
6. Lennon, A. M. *et al.* The Early Detection of Pancreatic Cancer: What Will it Take to Diagnose and Treat Curable Pancreatic Neoplasia? *Cancer Res.* **74**, 3381–3389 (2014).
7. Liu, M. C. *et al.* Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **31**, 745–759 (2020).
8. Walter, L. C., Lewis, C. L. & Barton, M. B. Screening for colorectal, breast, and cervical cancer in the elderly: A review of the evidence. *Am. J. Med.* **118**, 1078–1086 (2005).
9. Notta, F. *et al.* A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* **538**, 378–382 (2016).
10. Yachida, S. *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114–1117 (2010).
11. Hruban, R. H., Goggins, M., Parsons, J. & Kern, S. E. Progression Model for Pancreatic Cancer1. *Clin. Cancer Res.* **6**, 2969–2972 (2000).
12. Day, J. D. *et al.* Immunohistochemical evaluation of HER-2/neu expression in pancreatic adenocarcinoma and pancreatic intraepithelial neoplasms. *Hum. Pathol.* **27**, 119–124 (1996).
13. Lüttges, J. *et al.* The K-ras mutation pattern in pancreatic ductal adenocarcinoma usually is identical to that in associated normal, hyperplastic, and metaplastic ductal epithelium. *Cancer* **85**, 1703–1710 (1999).
14. Burris, H. A. *et al.* Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **15**, 2403–2413 (1997).
15. Conroy, T. *et al.* FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer. *N. Engl. J. Med.* **364**, 1817–1825 (2011).
16. Moore, M. J. *et al.* Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **25**, 1960–1966 (2007).
17. Raphael, B. J. *et al.* Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **32**, 185-203.e13 (2017).
18. Atkinson, M. A., Campbell-Thompson, M., Kusmartseva, I. & Kaestner, K. H. Organisation of the human pancreas in health and in diabetes. *Diabetologia* **63**, 1966–1973 (2020).

19. McKenna, L. R. & Edil, B. H. Update on pancreatic neuroendocrine tumors. *Gland Surg.* **3**, 258–275 (2014).
20. Flowers, B. M. *et al.* Cell of Origin Influences Pancreatic Cancer Subtype. *Cancer Discov.* **11**, 660–677 (2021).
21. Habbe, N. *et al.* Spontaneous induction of murine pancreatic intraepithelial neoplasia (mPanIN) by acinar cell targeting of oncogenic Kras in adult mice. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 18913–18918 (2008).
22. De La O, J.-P. *et al.* Notch and Kras reprogram pancreatic acinar cells to ductal intraepithelial neoplasia. *Proc. Natl. Acad. Sci.* **105**, 18907–18912 (2008).
23. Shi, C. *et al.* KRAS2 Mutations in Human Pancreatic Acinar-Ductal Metaplastic Lesions Are Limited to Those with PanIN: Implications for the Human Pancreatic Cancer Cell of Origin. *Mol. Cancer Res.* **7**, 230–236 (2009).
24. Lee, A. Y. L. *et al.* Cell of origin affects tumour development and phenotype in pancreatic ductal adenocarcinoma. *Gut* **68**, 487–498 (2019).
25. Boj, S. F. *et al.* Organoid Models of Human and Mouse Ductal Pancreatic Cancer. *Cell* **160**, 324–338 (2015).
26. Prior, I. A., Hood, F. E. & Hartley, J. L. The Frequency of Ras Mutations in Cancer. *Cancer Res.* **80**, 2969–2974 (2020).
27. Aguirre, A. J. *et al.* Real-time Genomic Characterization of Advanced Pancreatic Cancer to Enable Precision Medicine. *Cancer Discov.* **8**, 1096–1111 (2018).
28. Tada, M. *et al.* Analysis of K-ras gene mutation in hyperplastic duct cells of the pancreas without pancreatic disease. *Gastroenterology* **110**, 227–231 (1996).
29. Basturk, O. *et al.* A REVISED CLASSIFICATION SYSTEM AND RECOMMENDATIONS FROM THE BALTIMORE CONSENSUS MEETING FOR NEOPLASTIC PRECURSOR LESIONS IN THE PANCREAS. *Am. J. Surg. Pathol.* **39**, 1730–1741 (2015).
30. Matsuda, Y. *et al.* The Prevalence and Clinicopathological Characteristics of High-Grade Pancreatic Intraepithelial Neoplasia: Autopsy Study Evaluating the Entire Pancreatic Parenchyma. *Pancreas* **46**, 658–664 (2017).
31. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras Provokes Premature Cell Senescence Associated with Accumulation of p53 and p16INK4a. *Cell* **88**, 593–602 (1997).
32. Morton, J. P. *et al.* Mutant p53 drives metastasis and overcomes growth arrest/senescence in pancreatic cancer. *Proc. Natl. Acad. Sci.* **107**, 246–251 (2010).
33. Kamisawa, T., Wood, L. D., Itoi, T. & Takaori, K. Pancreatic cancer. *The Lancet* **388**, 73–85 (2016).
34. Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501 (2015).
35. Weinberg, R. A. & Weinberg, R. A. *The Biology of Cancer*. (W.W. Norton & Company, 2013). doi:10.1201/9780203852569.
36. di Magliano, M. P. & Logsdon, C. D. Roles for KRAS in Pancreatic Tumor Development and Progression. *Gastroenterology* **144**, 1220–1229 (2013).
37. Lauth, M. *et al.* DYRK1B-dependent autocrine-to-paracrine shift of Hedgehog signaling by mutant RAS. *Nat. Struct. Mol. Biol.* **17**, 718–725 (2010).
38. Thayer, S. P. *et al.* Hedgehog is an early and late mediator of pancreatic cancer tumorigenesis. *Nature* **425**, 851–856 (2003).

39. Bailey, J. M., Mohr, A. M. & Hollingsworth, M. A. Sonic hedgehog paracrine signaling regulates metastasis and lymphangiogenesis in pancreatic cancer. *Oncogene* **28**, 3513–3525 (2009).
40. Bailey, J. M. *et al.* Sonic Hedgehog Promotes Desmoplasia in Pancreatic Cancer. *Clin. Cancer Res.* **14**, 5995–6004 (2008).
41. Olive, K. P. *et al.* Inhibition of Hedgehog Signaling Enhances Delivery of Chemotherapy in a Mouse Model of Pancreatic Cancer. *Science* **324**, 1457–1461 (2009).
42. Uzunparmak, B. & Sahin, I. H. Pancreatic cancer microenvironment: a current dilemma. *Clin. Transl. Med.* **8**, 2 (2019).
43. Feig, C. *et al.* The Pancreas Cancer Microenvironment. *Clin. Cancer Res.* **18**, 4266–4276 (2012).
44. Hegde, S. *et al.* Dendritic Cell Paucity Leads to Dysfunctional Immune Surveillance in Pancreatic Cancer. *Cancer Cell* **37**, 289-307.e9 (2020).
45. Freed-Pastor, W. A. *et al.* The CD155/TIGIT axis promotes and maintains immune evasion in neoantigen-expressing pancreatic cancer. *Cancer Cell* **39**, 1342-1360.e14 (2021).
46. Vonderheide, R. H. CD40 Agonist Antibodies in Cancer Immunotherapy. *Annu. Rev. Med.* **71**, 47–58 (2020).
47. Murphy, K. & Weaver, C. *Janeway's Immunobiology*. (Garland Science, 2016).
48. Lythe, G., Callard, R. E., Hoare, R. L. & Molina-París, C. How many TCR clonotypes does a body maintain? *J. Theor. Biol.* **389**, 214–224 (2016).
49. Vallender, E. J. & Lahn, B. T. Positive selection on the human genome. *Hum. Mol. Genet.* **13**, R245–R254 (2004).
50. Cabeza-Cabrerizo, M., Cardoso, A., Minutti, C. M., Pereira da Costa, M. & Reis e Sousa, C. Dendritic Cells Revisited. *Annu. Rev. Immunol.* **39**, 131–166 (2021).
51. Banchereau, J. *et al.* Immunobiology of Dendritic Cells. *Annu. Rev. Immunol.* **18**, 767–811 (2000).
52. Prlic, M., Williams, M. A. & Bevan, M. J. Requirements for CD8 T-cell priming, memory generation and maintenance. *Curr. Opin. Immunol.* **19**, 315–319 (2007).
53. Starnes, C. O. Coley's toxins in perspective. *Nature* **357**, 11–12 (1992).
54. Redelman-Sidi, G., Glickman, M. S. & Bochner, B. H. The mechanism of action of BCG therapy for bladder cancer—a current perspective. *Nat. Rev. Urol.* **11**, 153–162 (2014).
55. Rosenberg, S. A. IL-2: The First Effective Immunotherapy for Human Cancer. *J. Immunol. Baltim. Md 1950* **192**, 5451–5458 (2014).
56. Allison, J. P., McIntyre, B. W. & Bloch, D. Tumor-specific antigen of murine T-lymphoma defined with monoclonal antibody. *J. Immunol.* **129**, 2293–2300 (1982).
57. Attia, P. #177 - Steven Rosenberg, M.D., Ph.D.: The development of cancer immunotherapy and its promise for treating advanced cancers. *Peter Attia* https://peterattiamd.com/stevenrosenberg/ (2021).
58. Rosenberg, S. A., Fox, E. & Churchill, W. H. Spontaneous regression of hepatic metastases from gastric carcinoma. *Cancer* **29**, 472–474 (1972).
59. Rosenberg, S. A., Mulé, J. J., Spiess, P. J., Reichert, C. M. & Schwarz, S. L. Regression of established pulmonary metastases and subcutaneous tumor mediated by the systemic administration of high-dose recombinant interleukin 2. *J. Exp. Med.* **161**, 1169–1188 (1985).
60. Francisco, L. M., Sage, P. T. & Sharpe, A. H. The PD-1 pathway in tolerance and autoimmunity. *Immunol. Rev.* **236**, 219–242 (2010).

61. Jubel, J. M., Barbati, Z. R., Burger, C., Wirtz, D. C. & Schildberg, F. A. The Role of PD-1 in Acute and Chronic Infection. *Front. Immunol.* **11**, (2020).
62. Marasco, M. *et al.* Molecular mechanism of SHP2 activation by PD-1 stimulation. *Sci. Adv.* **6**, eaay4458 (2020).
63. Sharpe, A. H. & Pauken, K. E. The diverse functions of the PD1 inhibitory pathway. *Nat. Rev. Immunol.* **18**, 153–167 (2018).
64. Cui, W. & Kaech, S. M. Generation of effector CD8+ T cells and their conversion to memory T cells. *Immunol. Rev.* **236**, 151–166 (2010).
65. Seidel, J. A., Otsuka, A. & Kabashima, K. Anti-PD-1 and Anti-CTLA-4 Therapies in Cancer: Mechanisms of Action, Efficacy, and Limitations. *Front. Oncol.* **8**, (2018).
66. Wing, K. *et al.* CTLA-4 control over Foxp3+ regulatory T cell function. *Science* **322**, 271–275 (2008).
67. Brunet, J.-F. *et al.* A new member of the immunoglobulin superfamily—CTLA-4. *Nature* **328**, 267–270 (1987).
68. Harper, K. *et al.* CTLA-4 and CD28 activated lymphocyte molecules are closely related in both mouse and human as to sequence, message expression, gene structure, and chromosomal location. *J. Immunol. Baltim. Md 1950* **147**, 1037–1044 (1991).
69. Leach, D. R., Krummel, M. F. & Allison, J. P. Enhancement of antitumor immunity by CTLA-4 blockade. *Science* **271**, 1734–1736 (1996).
70. Carlino, M. S. & Long, G. V. Ipilimumab Combined with Nivolumab: A Standard of Care for the Treatment of Advanced Melanoma? *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **22**, 3992–3998 (2016).
71. Treatment of Melanoma by Stage. https://www.cancer.org/cancer/melanoma-skin-cancer/treating/by-stage.html.
72. Hodi, F. S. *et al.* Safety and efficacy of combination nivolumab plus ipilimumab in patients with advanced melanoma: results from a North American expanded access program (CheckMate 218). *Melanoma Res.* **31**, 67–75 (2021).
73. Brahmer, J. R. *et al.* Safety and Activity of Anti–PD-L1 Antibody in Patients with Advanced Cancer. *N. Engl. J. Med.* **366**, 2455–2465 (2012).
74. Royal, R. E. *et al.* Phase 2 Trial of Single Agent Ipilimumab (Anti-CTLA-4) for Locally Advanced or Metastatic Pancreatic Adenocarcinoma. *J. Immunother. Hagerstown Md 1997* **33**, 828–833 (2010).
75. Fesnak, A. D., June, C. H. & Levine, B. L. Engineered T cells: the promise and challenges of cancer immunotherapy. *Nat. Rev. Cancer* **16**, 566–581 (2016).
76. Adoptive Cell Therapy: CAR T, TCR, TIL, NK. *Cancer Research Institute* https://www.cancerresearch.org/en-us/immunotherapy/treatment-types/adoptive-cell-therapy.
77. Akce, M., Zaidi, M. Y., Waller, E. K., El-Rayes, B. F. & Lesinski, G. B. The Potential of CAR T Cell Therapy in Pancreatic Cancer. *Front. Immunol.* **9**, 2166 (2018).
78. Leidner, R. *et al.* Neoantigen T-Cell Receptor Gene Therapy in Pancreatic Cancer. *N. Engl. J. Med.* **386**, 2112–2119 (2022).
79. Rollins, M. R., Spartz, E. J. & Stromnes, I. M. T Cell Receptor Engineered Lymphocytes for Cancer Therapy. *Curr. Protoc. Immunol.* **129**, e97 (2020).
80. Tran, E. *et al.* T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *N. Engl. J. Med.* **375**, 2255–2262 (2016).

81. Balachandran, V. P. *et al.* Phase I trial of adjuvant autogene cevumeran, an individualized mRNA neoantigen vaccine, for pancreatic ductal adenocarcinoma. *J. Clin. Oncol.* **40**, 2516–2516 (2022).
82. Cheever, M. A. & Higano, C. S. PROVENGE (Sipuleucel-T) in Prostate Cancer: The First FDA-Approved Therapeutic Cancer Vaccine. *Clin. Cancer Res.* **17**, 3520–3526 (2011).
83. Wells, D. K. *et al.* Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. *Cell* **183**, 818-834.e13 (2020).
84. Yarmarkovich, M. *et al.* Cross-HLA targeting of intracellular oncoproteins with peptide-centric CARs. *Nature* **599**, 477–484 (2021).
85. Hsiue, E. H.-C. *et al.* Targeting a neoantigen derived from a common TP53 mutation. *Science* **371**, eabc8697 (2021).
86. Gubin, M. M. *et al.* Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577–581 (2014).
87. Nevala-Plagemann, C., Hidalgo, M. & Garrido-Laguna, I. From state-of-the-art treatments to novel therapies for advanced-stage pancreatic cancer. *Nat. Rev. Clin. Oncol.* **17**, 108–123 (2020).
88. Macherla, S. *et al.* Emerging Role of Immune Checkpoint Blockade in Pancreatic Cancer. *Int. J. Mol. Sci.* **19**, 3505 (2018).
89. Hakim, N., Patel, R., Devoe, C. & Saif, M. W. Why HALO 301 Failed and Implications for Treatment of Pancreatic Cancer. *Pancreas Fairfax Va* **3**, e1–e4 (2019).
90. Torphy, R. J., Schulick, R. D. & Zhu, Y. Understanding the immune landscape and tumor microenvironment of pancreatic cancer to improve immunotherapy. *Mol. Carcinog.* **59**, 775–782 (2020).
91. Bailey, P. *et al.* Exploiting the neoantigen landscape for immunotherapy of pancreatic ductal adenocarcinoma. *Sci. Rep.* **6**, 35848 (2016).
92. Balachandran, V. P. *et al.* Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* **551**, 512–516 (2017).
93. Łuksza, M. *et al.* Neoantigen quality predicts immunoediting in survivors of pancreatic cancer. *Nature* **606**, 389–395 (2022).
94. Ehx, G. *et al.* Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes. *Immunity* **54**, 737-752.e10 (2021).
95. Ouspenskaia, T. *et al.* Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat. Biotechnol.* **40**, 209–217 (2022).
96. Roudko, V. *et al.* Shared Immunogenic Poly-Epitope Frameshift Mutations in Microsatellite Unstable Tumors. *Cell* **183**, 1634-1649.e17 (2020).
97. Smart, A. C. *et al.* Intron retention is a source of neoepitopes in cancer. *Nat. Biotechnol.* **36**, 1056–1058 (2018).
98. Kahles, A. *et al.* Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell* **34**, 211-224.e6 (2018).
99. Rajasagi, M. *et al.* Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* **124**, 453–462 (2014).
100. Huang, J. *et al.* T Cells Associated with Tumor Regression Recognize Frameshifted Products of the CDKN2A Tumor Suppressor Gene Locus and a Mutated HLA Class I Gene Product. *J. Immunol.* **172**, 6057–6064 (2004).
101. Turajlic, S. *et al.* Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol.* **18**, 1009–1021 (2017).

102. Parkhurst, M. R. *et al.* Unique Neoantigens Arise from Somatic Mutations in Patients with Gastrointestinal Cancers. *Cancer Discov.* **9**, 1022–1035 (2019).
103. Litchfield, K. *et al.* Escape from nonsense-mediated decay associates with anti-tumor immunogenicity. *Nat. Commun.* **11**, 3800 (2020).
104. Bigot, J. *et al.* Splicing Patterns in SF3B1-Mutated Uveal Melanoma Generate Shared Immunogenic Tumor-Specific Neoepitopes. *Cancer Discov.* **11**, 1938–1951 (2021).
105. Demircioğlu, D. *et al.* A Pan-cancer Transcriptome Analysis Reveals Pervasive Regulation through Alternative Promoters. *Cell* **178**, 1465-1477.e17 (2019).
106. Frankiw, L., Baltimore, D. & Li, G. Alternative mRNA splicing in cancer immunotherapy. *Nat. Rev. Immunol.* **19**, 675–687 (2019).
107. Intron-Retention Neoantigen Load Predicts Favorable Prognosis in Pancreatic Cancer | JCO Clinical Cancer Informatics. https://ascopubs.org/doi/full/10.1200/CCI.21.00124.
108. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
109. Ecker, J. R. *et al.* ENCODE explained. *Nature* **489**, 52–54 (2012).
110. Statello, L., Guo, C.-J., Chen, L.-L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).
111. Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* **4**, e08890 (2015).
112. Lu, S. X. *et al.* Pharmacologic modulation of RNA splicing enhances anti-tumor immunity. *Cell* **184**, 4032-4047.e31 (2021).
113. Keskin, D. B. *et al.* Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature* **565**, 234–239 (2019).
114. Ott, P. A. *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217–221 (2017).
115. Ott, P. A. *et al.* A Phase Ib Trial of Personalized Neoantigen Therapy Plus Anti-PD-1 in Patients with Advanced Melanoma, Non-small Cell Lung Cancer, or Bladder Cancer. *Cell* **183**, 347-362.e24 (2020).
116. Bauer, D. C., Zadoorian, A., Wilson, L. O. W., Melbourne Genomics Health Alliance & Thorne, N. P. Evaluation of computational programs to predict HLA genotypes from genomic sequencing data. *Brief. Bioinform.* **19**, 179–187 (2018).
117. Meléndez, B. *et al.* Methods of measurement for tumor mutational burden in tumor tissue. *Transl. Lung Cancer Res.* **7**, 661–667 (2018).
118. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
119. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020).
120. Sarkizova, S. *et al.* A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat. Biotechnol.* **38**, 199–209 (2020).
121. Jurtz, V. *et al.* NetMHCpan 4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol. Baltim. Md 1950* **199**, 3360–3368 (2017).
122. Sahin, U. *et al.* Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547**, 222–226 (2017).

123. Westcott, P. M. K. *et al.* Low neoantigen expression and poor T-cell priming underlie early immune escape in colorectal cancer. *Nat. Cancer* **2**, 1071–1085 (2021).
124. Chong, C. *et al.* Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat. Commun.* **11**, 1293 (2020).
125. Strønen, E. *et al.* Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science* **352**, 1337–1341 (2016).
126. Bradley, S. D. *et al.* Vestigial-like 1 is a shared targetable cancer-placenta antigen expressed by pancreatic and basal-like breast cancers. *Nat. Commun.* **11**, 5332 (2020).
127. Ma, Y. *et al.* Combination of PD-1 Inhibitor and OX40 Agonist Induces Tumor Rejection and Immune Memory in Mouse Models of Pancreatic Cancer. *Gastroenterology* **159**, 306-319.e12 (2020).
128. Klaeger, S. *et al.* Optimized Liquid and Gas Phase Fractionation Increases HLA-Peptidome Coverage for Primary Cell and Tissue Samples. *Mol. Cell. Proteomics* **20**, (2021).
129. Campbell, P. J. *et al.* Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
130. Knoechel, B. *et al.* An epigenetic mechanism of resistance to targeted therapy in T cell acute lymphoblastic leukemia. *Nat. Genet.* **46**, 364–370 (2014).
131. Awad, M. M. *et al.* Acquired Resistance to KRASG12C Inhibition in Cancer. *N. Engl. J. Med.* **384**, 2382–2393 (2021).
132. Jabbour, E. *et al.* Characteristics and outcomes of patients with chronic myeloid leukemia and T315I mutation following failure of imatinib mesylate therapy. *Blood* **112**, 53–55 (2008).
133. Cortes, J. e. *et al.* A Phase 2 Trial of Ponatinib in Philadelphia Chromosome–Positive Leukemias. *N. Engl. J. Med.* **369**, 1783–1796 (2013).
134. Daver, N. *et al.* Secondary mutations as mediators of resistance to targeted therapy in leukemia. *Blood* **125**, 3236–3245 (2015).
135. Canon, J. *et al.* The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* **575**, 217–223 (2019).
136. Reck, M., Carbone, D. P., Garassino, M. & Barlesi, F. Targeting KRAS in non-small-cell lung cancer: recent progress and new approaches. *Ann. Oncol.* **32**, 1101–1110 (2021).
137. Westcott, P. M. K. *et al.* The mutational landscapes of genetic and chemical models of Kras-driven lung cancer. *Nature* **517**, 489–492 (2015).
138. Donehower, L. A. *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221 (1992).
139. Lampreht Tratar, U., Horvat, S. & Cemazar, M. Transgenic Mouse Models in Cancer Research. *Front. Oncol.* **8**, 268 (2018).
140. Jonkers, J. & Berns, A. Conditional mouse models of sporadic cancer. *Nat. Rev. Cancer* **2**, 251–265 (2002).
141. Łobocka, M. B. *et al.* Genome of Bacteriophage P1. *J. Bacteriol.* **186**, 7032–7068 (2004).
142. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* **15**, 3243–3248 (2001).
143. Hingorani, S. R. *et al.* Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4**, 437–450 (2003).
144. Hingorani, S. R. *et al.* Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).

145. Lang, G. A. *et al.* Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. *Cell* **119**, 861–872 (2004).
146. Fontana, R., Ranieri, M., La Mantia, G. & Vivo, M. Dual Role of the Alternative Reading Frame ARF Protein in Cancer. *Biomolecules* **9**, 87 (2019).
147. Bardeesy, N. *et al.* Both p16Ink4a and the p19Arf-p53 pathway constrain progression of pancreatic adenocarcinoma in the mouse. *Proc. Natl. Acad. Sci.* **103**, 5947–5952 (2006).
148. Westphalen, C. B. & Olive, K. P. Genetically Engineered Mouse Models of Pancreatic Cancer. *Cancer J. Sudbury Mass* **18**, 502–510 (2012).
149. Kojima, K. *et al.* Inactivation of Smad4 Accelerates KrasG12D-Mediated Pancreatic Neoplasia. *Cancer Res.* **67**, 8121–8130 (2007).
150. Zafra, M. P. *et al.* An In Vivo Kras Allelic Series Reveals Distinct Phenotypes of Common Oncogenic Variants. *Cancer Discov.* **10**, 1654–1671 (2020).
151. Hunter, J. C. *et al.* Biochemical and Structural Analysis of Common Cancer-Associated KRAS Mutations. *Mol. Cancer Res.* **13**, 1325–1335 (2015).
152. Freed-Pastor, W. A. & Prives, C. Mutant p53: one name, many proteins. *Genes Dev.* **26**, 1268–1286 (2012).
153. Schulz-Heddergott, R. *et al.* Therapeutic Ablation of Gain-of-Function Mutant p53 in Colorectal Cancer Inhibits Stat3-Mediated Tumor Growth and Invasion. *Cancer Cell* **34**, 298-314.e7 (2018).
154. Barrangou, R. & Horvath, P. A decade of discovery: CRISPR functions and applications. *Nat. Microbiol.* **2**, 1–9 (2017).
155. Lander, E. S. The Heroes of CRISPR. *Cell* **164**, 18–28 (2016).
156. Nakata, A., Amemura, M. & Makino, K. Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome. *J. Bacteriol.* **171**, 3553–3556 (1989).
157. Mojica, F. J., Juez, G. & Rodríguez-Valera, F. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol. Microbiol.* **9**, 613–621 (1993).
158. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182 (2005).
159. Jansen, R., Embden, J. D. A. van, Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002).
160. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315**, 1709–1712 (2007).
161. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).
162. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (2013).
163. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–826 (2013).
164. Lodish, U. H. *et al. Molecular Cell Biology*. (Macmillan, 2008).
165. Xue, W. *et al.* CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature* **514**, 380–384 (2014).
166. Sánchez-Rivera, F. J. *et al.* Rapid modelling of cooperating genetic events in cancer through somatic genome editing. *Nature* **516**, 428–431 (2014).

167. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
168. Cheung, A. F., DuPage, M. J. P., Dong, H. K., Chen, J. & Jacks, T. Regulated Expression of a Tumor-Associated Antigen Reveals Multiple Levels of T-Cell Tolerance in a Mouse Model of Lung Cancer. *Cancer Res.* **68**, 9459–9468 (2008).
169. Mehta, A. & Merkel, O. M. Immunogenicity of Cas9 Protein. *J. Pharm. Sci.* **109**, 62–67 (2020).
170. Böck, D. *et al.* In vivo prime editing of a metabolic liver disease in mice. *Sci. Transl. Med.* **14**, eabl9238 (2022).
171. Choi, J.-H. *et al.* Optimization of AAV expression cassettes to improve packaging capacity and transgene expression in neurons. *Mol. Brain* **7**, 17 (2014).
172. Winters, I. P. *et al.* Multiplexed in vivo homology-directed repair and tumor barcoding enables parallel quantification of Kras variant oncogenicity. *Nat. Commun.* **8**, 2053 (2017).
173. Wang, B. *et al.* Highly efficient CRISPR/HDR-mediated knock-in for mouse embryonic stem cells and zygotes. *BioTechniques* **59**, 201–202, 204, 206–208 (2015).
174. Newby, G. A. & Liu, D. R. In vivo somatic cell base editing and prime editing. *Mol. Ther.* **29**, 3107–3124 (2021).
175. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
176. Platt, R. J. *et al.* CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. *Cell* **159**, 440–455 (2014).
177. Sánchez-Rivera, F. J. *et al.* Base editing sensor libraries for high-throughput engineering and functional analysis of cancer-associated single nucleotide variants. *Nat. Biotechnol.* **40**, 862–873 (2022).
178. Zafra, M. P. *et al.* Optimized base editors enable efficient editing in cells, organoids and mice. *Nat. Biotechnol.* **36**, 888–893 (2018).
179. Koblan, L. W. *et al.* Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nat. Biotechnol.* **39**, 1414–1425 (2021).
180. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
181. Nelson, J. W. *et al.* Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**, 402–410 (2022).
182. Chen, P. J. *et al.* Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635-5652.e29 (2021).
183. Petri, K. *et al.* CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nat. Biotechnol.* **40**, 189–193 (2022).
184. Liu, P. *et al.* Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. *Nat. Commun.* **12**, 2121 (2021).
185. Liu, B. *et al.* A split prime editor with untethered reverse transcriptase and circular RNA template. *Nat. Biotechnol.* 1–6 (2022) doi:10.1038/s41587-022-01255-9.
186. Kobayashi, Y. *et al.* Silent mutations reveal therapeutic vulnerability in RAS Q61 cancers. *Nature* **603**, 335–342 (2022).
187. Feigin, M. E. *et al.* Recurrent noncoding regulatory mutations in pancreatic ductal adenocarcinoma. *Nat. Genet.* **49**, 825–833 (2017).

CHAPTER 2: A BROADENED HLA LIGANDOME UNCOVERS NEW IMMUNOTHERAPY TARGETS FOR PANCREATIC CANCER

Authors: Zackery A. Ely[1,2], William Freed-Pastor[1,3], Eva Verzani[4], Zachary Kulstad[1,3], Jenn Abelin[4], Karl Clauser[4], Susan Klaeger[4], Miles Agus[1], Jennifer Su[1], Kevin Kapner[3], Andy Aguirre[3], Steven Carr[4], Tyler Jacks[1,2]

This chapter is a modified version of a manuscript currently in preparation for submission.

Affiliations:
[1] David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA
[4] Broad Institute of MIT and Harvard, Cambridge, MA, USA

Contributions: Z.A.E, W.F.P., and T.J., conceived of and designed the study. Z.A.E. conducted all bioinformatic analyses. Z.A.E., W.F.P, and Z.K. conducted T cell expansion assays and related analyses. A.A. and K.K. provided patient-derived organoid specimens and related sequencing data. M.A. and Z.K. contributed to analysis of nuORF and canonical gene expression in normal tissues. J.S. conducted irradiation. E.V., K.C., S.K., and J.A. conducted all immunopeptidomics and related peptide identifications, with guidance from S.C. Z.A.E. wrote the manuscript with input from W.F.P. and T.J.

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is a lethal disease largely recalcitrant to traditional immunotherapy. A large subset of PDAC tumors is computationally predicted to harbor potentially immunogenic peptides for MHC class I (MHC-I) presentation, but the nature, expression, and cancer-specificity of these peptides has yet to be determined. The only prior study of the PDAC immunopeptidome focused on profiling MHC class I-associated peptides (MAPs) from canonical proteins in bulk tumor samples; however, non-malignant cell populations comprise most of the tumor mass in PDAC, obscuring the identity of MAPs that derive specifically from cancer cells. We overcame this challenge by expanding patient-derived PDAC organoids prior to profiling with whole genome sequencing, RNA sequencing, and immunopeptidomics. We find that organoids enable a higher resolution view of the PDAC transcriptome and immunopeptidome than bulk tumor populations, and they improve detection of protein-coding variants. We harnessed these data to construct patient-specific proteome databases, enabling the identification of MAPs derived from somatic and germline mutations, retained introns, and PDAC-enriched novel unannotated open reading frames (nuORFs) associated with noncoding RNAs, untranslated regions, and pseudogenes. We identified a large cohort of nuORFs and canonical genes that encode MAPs shared by multiple patients. These include nuORFs and genes that exhibit PDAC-restricted expression, validated by examination of RNA sequencing and immunopeptidome data derived from hundreds of healthy normal tissue specimens. We investigated the immunogenicity of several of these PDAC-restricted MAPs and discovered that $CD8^+$ T cells specific to nuORF-derived MAPs can be readily generated from healthy donor peripheral blood mononuclear cells (PBMCs). These results shed new light into the immunopeptidome of pancreatic cancer cells and furnish a large set of novel targets for immunotherapies in PDAC.

INTRODUCTION

Immunotherapy—the therapeutic manipulation of the immune system to target cancer—has revolutionized cancer treatment in the last decade, providing extended life and occasionally cures for a subset of patients. This achievement has invigorated the field of oncology with new hope, yet immunotherapies still fail to benefit patients suffering from most malignancies, including pancreatic cancer, an extremely lethal disease with a 5-year survival rate of approximately 10%. Many studies have explained this by invoking the concept of pancreatic cancer as an ‘immunologically cold’ malignancy, in which a low mutational burden helps conceal the cancer from recognition by immune cells [1,2]. In principle, a low mutational burden should give rise to fewer neoepitopes, mutant peptides that are empirically validated targets of anti-tumor T cells and many immunotherapies [3]. However, recent computational studies have challenged this notion for pancreatic cancer, reporting dozens of predicted neoepitopes in the typical patient [4–6] To date, the field has primarily analyzed pancreatic cancer neoepitopes derived from missense mutations that are typically unique to individual patients, limiting therapeutic applicability. However, neoepitopes can derive from other classes of somatic mutations. For example, frameshift mutations, in principle, should generate long, ‘foreign’ peptide sequences that could be recurrent among patients, as different frameshifts in the same gene could yield similar alternative reading frames. Frameshift mutations are often excluded from neoepitope predictions in PDAC [4,5].

Beyond mutations, a burst of studies over recent years has elucidated numerous sources of cancer-restricted antigens. For example, mRNA splicing often becomes severely dysregulated in numerous cancer types [7], leading to the production of novel RNA isoforms encoding cancer-specific protein sequences that can be presented by MHC class I [8,9]. Although numerous mis-splicing events have been identified as potential sources of cancer-specific epitopes, several papers have highlighted intron retention as the most frequent splicing source of cancer-restricted epitopes [9]. Yet, intron retention has not been explored as a source of neoepitopes in pancreatic cancer cells beyond a predictive study in bulk tumor samples [10].

Other studies have explored novel or unannotated open reading frames (nuORFs) as new sources of cancer-restricted epitopes. Canonically thought not to encode protein, several classes of nuORFs spanning long noncoding RNAs (lncRNAs), 5’ and 3’ untranslated regions (UTRs), pseudogenes, and other regions have been shown to be bound by ribosomes and actively translated into peptides [11,12]. Most recently, Ouspenskaia and colleagues employed high-resolution immunopeptidomics to validate MHC presentation of thousands of nuORF-derived peptides on the surface of melanoma, chronic lymphocytic leukemia, and glioblastoma cells [12]. Notably, many of the transcripts encoding these peptides were shown to be transcribed or translated in a cancer-specific fashion by RNA-Sequencing (RNA-Seq) and ribosome profiling, highlighting the potential of nuORF-derived peptides as a class of novel immunotherapy targets.

Despite advances in our understanding of sources of cancer-restricted epitopes, efforts to profile these in pancreatic cancer have lagged far behind. To our knowledge, only one prior study has employed immunopeptidomics—the state-of-the-art technique for empirical detection of MHC-associated peptides (MAPs)—to analyze bulk pancreatic cancer specimens [13]. Although this study revealed a promising epitope derived from the placentally expressed gene, *VGLL1*, analysis was limited to the reference proteome, thereby excluding mutations, intron retention, and nuORFs as epitope sources.

Both the single empirical study and the numerous computational predictions of the PDAC immunopeptidome are typically performed using bulk tumor samples, which contain cells from the surrounding stroma, like immune cells or fibroblasts, in additional to cancer cells. Pancreatic cancer is notorious for its high stromal content [14]. With most of the sequencing data derived from these non-mutated stromal cells, it is challenging to estimate the fraction of tumor cells containing a particular mutation, a metric known as variant allele frequency, and some mutations may even go undetected. Furthermore, transcriptomic and proteomic datasets profiling bulk PDAC samples are often contaminated with confounding signals from stromal populations [15,16]. We sought to overcome these limitations by using patient-derived organoids (PDOs) to expand a pure malignant population for whole-genome sequencing, RNA-Sequencing, and immunopeptidomics. Importantly, we also employed a diverse set of bioinformatic techniques to construct a broad proteogenomic search space for MAPs, accounting for germline and somatic mutations, nuORFs, and retained introns as potential sources of MAPs. We applied this approach to seven patient biospecimens and detected 17,000-20,500 unique MAPs per sample, a dramatic increase in both depth and resolution over prior efforts. By analyzing the transcriptome and immunopeptidome of thousands of normal tissue specimens, we identified numerous MAPs spanning nuORFs, retained introns, and canonical proteins with detectable expression restricted solely to pancreatic cancer cells or restricted to both pancreatic cancer cells and select normal tissues. We also identified mutation-derived MAPs in the majority of PDOs, including a frameshift mutation derived from a tumor suppressor gene. In total, our findings elucidate the landscape of MAPs presentable by pancreatic cancer, broadening the repertoire of potential immunotherapy targets for this devastating disease.

RESULTS

Patient-derived pancreatic cancer organoids possess a greater somatic mutational burden than matched bulk tumor samples. Cancer cells typically represent <20% of the tumor volume in PDAC [14]. Thus, the contribution of sequencing reads from stromal cells could significantly deflate DNA variant allele frequency (VAF), a critical factor influencing the sensitivity of somatic variant detection [17]. With these considerations in mind, we hypothesized that prior studies estimating tumor mutational burden (TMB) and predicted neoepitopes in pancreatic cancer may have underestimated PDAC TMB, as these studies are traditionally limited to bulk tumor samples [14,18]. To assess this hypothesis, we computed TMB (i.e., the number of mutations per coding Mb) across three cohorts of patient-derived samples, including 148 primary bulk samples from The Cancer Genome Atlas (TCGA; WES), 57 metastatic bulk samples from the Dana-Farber Institute's PancSeq (WES), and 41 patient-derived organoids (WGS) **(Fig. 1a**). The expansion of patient-derived cancer cells as organoids enables enrichment for cancer cells prior to analysis, thereby reducing or eliminating stromal contamination.

We detected a significant difference in average TMB among all three cohorts ($P < 0.0001$, Kruskal-Wallis), with the highest TMB load observed in patient-derived organoids, followed by the metastatic DFCI cohort and lastly the primary TCGA cohort (**Fig. 1b**). The significant increase in TMB in patient0derived organoids is not attributable to sequencing depth, as this cohort was sequenced using WGS at a significantly lower depth than the WES used to profile bulk tumors [19]. Furthermore, these organoids were sequenced after only an average of eight passages, and thus little time was available to accumulate additional mutations in culture. However, we cannot rule out the possibility that organoid generation selected for sub-clones of cancer cells, which may have impacted TMB assessment.

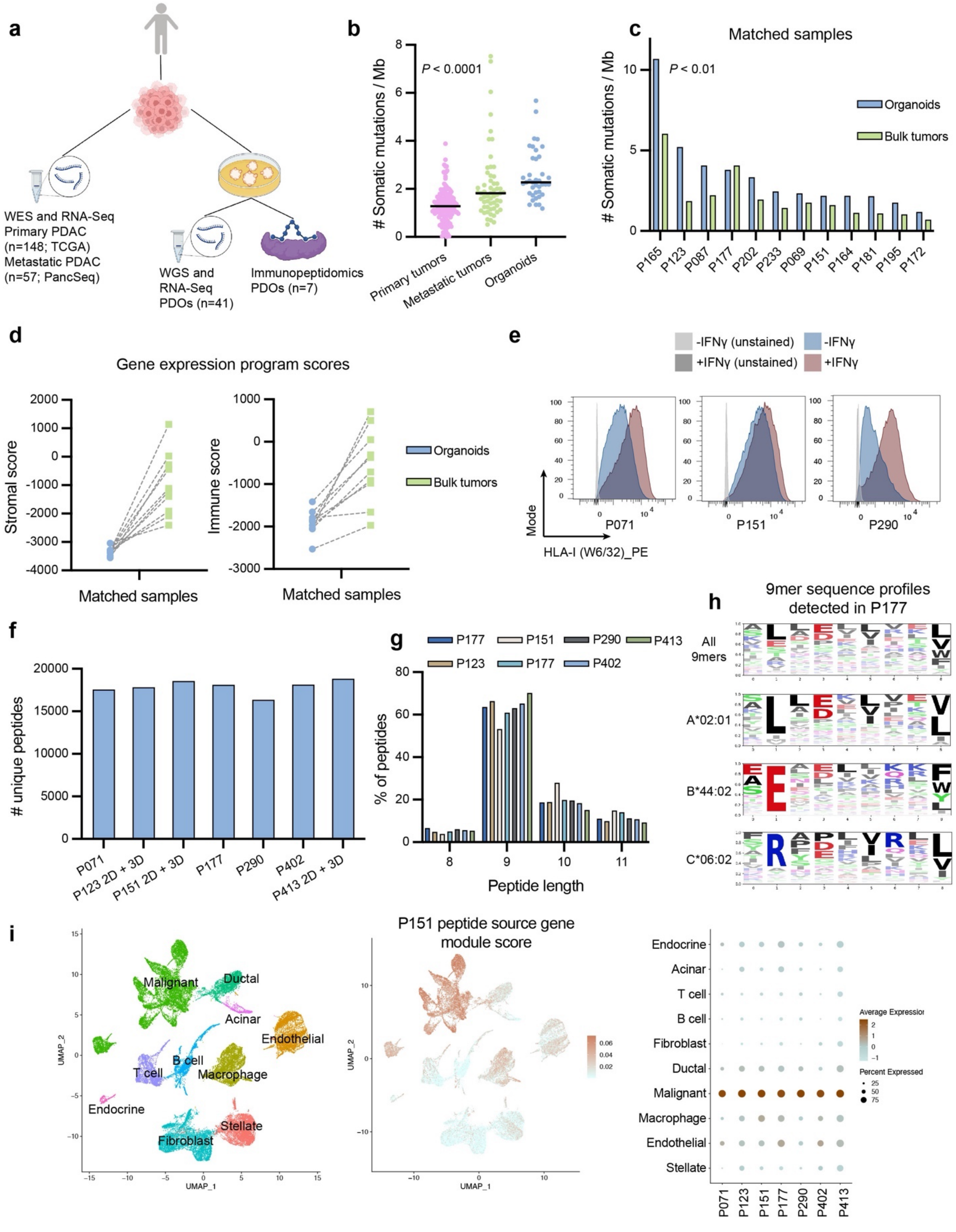


Figure 1: Pancreatic cancer organoids expand the neoepitope landscape and enable a high-resolution characterization of the immunopeptidome.

a. Schematic depicting the cohorts of pancreatic cancer specimens analyzed and corresponding sequencing datasets.
b. Tumor mutational burden (TMB) comparisons across TCGA primary bulk samples, PancSeq metastatic bulk tumor samples, and patient-derived organoids. Horizontal bars indicate median TMB. Two outlier data points (TMB = 10.7 and 34.6) from two organoid samples are excluded from the graph. *P*-value evaluated with the Kruskal-Wallis test.
c. TMB comparison between pancreatic cancer organoids and matched bulk tumor specimens that were derived from the same patient. *P*-value determined by the Wilcoxon ranked sign test.
d. Quantification of stromal and immune gene expression program scores calculated by the ESTIMATE algorithm for organoids and matched bulk tumor samples.
e. Flow cytometry quantifying the surface expression of MHC class I proteins in pancreatic cancer organoids with (red) or without (blue) interferon gamma stimulation, compared to unstained controls (gray). Plots for four other samples are shown in Supplementary Figure 3.
f. Total number of unique peptide sequences identified in the immunopeptidome of seven different patient-derived organoid samples. +2D indicates that data also include peptides identified from a matched cell line derived from the organoid.
g. Distribution of peptide lengths for all mapped spectra, categorized for each organoid sample.
h. Peptide sequence profile plots representing the amino acid profile of all 9mer peptides (top) or specific subsets of 9mers predicted to bind A*02:01 (second row), B*44:02 (third row), or C*06:02 (bottom), for peptides derived from sample P177. HLA allele binding partner predictions were made with HLAthena [20].
i. Uniform Manifold Approximation and Projection (UMAP) depicting annotated cell types derived from >20 individual patient tumor samples (left). Expression of an aggregate gene module score derived from all source genes encoding peptides detected in the immunopeptidome for sample PANFR0151 is plotted in the middle. Average expression of seven gene modules derived from the immunopeptidomes of seven organoid samples shows specific enrichment in the malignant population (right).

We also compared TMB among twelve matched samples for which both bulk tumors and organoids were derived and sequenced. We found that organoids almost uniformly presented a higher TMB than their bulk tumor counterparts (**Fig. 1c**; $P < 0.01$, Wilcoxon signed-rank test). Though a modest increase in overall TMB, these data support the notion that profiling cancer-cell enriched organoids may increase our ability to discover neoepitopes in pancreatic cancer. Yoshihara and colleagues previously showed that tumor purity affects the number of mutations detected for a small fraction of cancer types, though PDAC was not examined. To test whether tumor purity might impact TMB within bulk tumor cohorts, we categorized TCGA patients into high- and low-purity classes (top 25% vs bottom 25%) and compared TMB between groups. TMB was significantly higher in TCGA samples marked as high purity, regardless of the method for quantifying purity (**Supplementary Fig. 1**). While TMB also trended slightly higher in metastatic patients, it was not statistically significant. We confirmed that these predictions also hold when comparing predicted neoepitope burden between matched organoid and bulk tumor samples and between high- and low-purity samples (**Supplementary Fig. 1**). In total, these findings indicate that tumor purity has a pronounced effect on the estimate of mutational burden in PDAC, and they highlight the utility of patient-derived organoids for uncovering a broader landscape of mutations.

Variant allele frequency (VAF) is a commonly used threshold for filtering candidate neoepitopes predicted by computational pipelines, including in studies examining PDAC [5,21,22]. In general, VAF cutoffs below 10% increase the risk of false-positive calls that can arise from sequencing errors [17], but many *bona fide* PDAC mutations may be well below these cutoffs for the typical tumor with neoplastic cellularity below 20%. To quantify the impact of VAF cutoffs, we compared the number of predicted neoepitopes between matched bulk and organoid samples, considering only neoepitopes derived from somatic mutations called in both samples for each patient. We found that a standard VAF cutoff of 0.1 in bulk tumors resulted in a mean exclusion of 4.5 variants (0-17) encoding predicted neoepitopes (IC_{50} < 500 nM) that were detected in the corresponding organoid with VAF > 0.1 (**Supplementary Fig. 1**). Using a conservative VAF cutoff of 0.07 employed in prior studies of neoepitopes in PDAC [22], we found that 2.7 variants (0-11) were excluded on average. Notably, this effect was most dramatic for lower purity tumors (**Supplementary Fig. 1**). Collectively, these results further show that conventional methods of predicting neoepitopes in PDAC can overlook somatic variants that encode potential neoepitopes.

Patient-derived organoids provide a higher resolution profiling of the PDAC transcriptome and immunopeptidome. Beyond impacting variant detection, stromal populations in bulk PDAC samples also contribute noise to transcriptomic and proteomic measurements [15,16]. To compare the presence of stromal gene expression programs between bulk pancreatic tumors (PancSeq) and pancreatic cancer organoids, we employed the Estimation of STromal and Immune cells in MAlignant Tumours using Expression data (ESTIMATE) algorithm and quantified the degree of stromal and immune cell contributions in these transcriptomes. As expected, bulk tumor populations contained a dramatically higher level of stromal and immune gene signatures than organoids ($P < 0.0001$ in both cases; Welch's *t*-test), including a uniform increase in these signatures between ten bulk tumors and organoids derived from the same patients (**Fig. 1d**; **Supplementary Fig. 2**). Of note, 54% of the genes comprising the stromal and immune signatures were not expressed at all (mean raw RSEM counts < three) in the organoid cohort, while 98% were expressed in the bulk tumor cohort (**Supplementary Fig. 2**). The genes in each of these categories that were detected in PDOs were expressed at extremely low levels, reflected in the low stromal and immune scores for these samples.

As expected, gene modules derived from these missing genes are absent in malignant cell populations derived from human PDAC single-cell expression data [23], instead enriching strongly in other cells from the tumor microenvironment (**Supplementary Fig. 2**). The aggregate ESTIMATE score and tumor purity calculated by ABSOLUTE showed a mild association (correlation = -0.56) comparable to that noted for other bulk tumor RNA-Seq datasets (**Supplementary Fig. 2**) [24]. Collectively, these data confirm that organoids provide a purer view of the pancreatic cancer transcriptome, devoid of the stromal gene expression signatures present in bulk samples that could confound interpretation of pancreatic cancer expression programs. Prior work suggests that autophagy-related pathways may degrade MHC class I in pancreatic cancer cells, thereby enabling evasion of an anti-PDAC immune response [25]. To assess whether MHC class I is expressed in our PDOs, we performed flow cytometry to profile baseline and interferon-gamma inducible surface expression of MHC-I. In all eight examined patient samples, we identified significant cell surface baseline expression of MHC-I that was further elevated by exposure to interferon gamma (**Fig. 1e**; **Supplementary Figure 3**). These data suggest that most pancreatic tumors still present sufficient levels of MHC class I for presentation of potential cancer-restricted epitopes, motivating further characterization.

To enable a higher resolution of the immunopeptidome in human PDAC, we expanded patient-derived organoid lines (n = 7) to 50-100 million cells. These samples harbored driver mutation patterns resembling those observed in sequencing studies of larger PDAC cohorts (**Supplementary Fig. 3**) [14]. We also adapted three of these organoid samples to monolayer cell lines (2D) for concomitant expansion and analysis. After expansion, we stimulated each line with interferon gamma to increase MHC class I expression and peptide detection sensitivity [20]. After harvesting each sample and eluting peptides bound by MHC class I, we identified over 17,000 MHC-associated peptides (MAPs) in each sample (**Fig. 1f**), a dramatic increase over the only prior study of the PDAC immunopeptidome [13].

The characteristics of detected peptides matched expectations, including appropriate peptide length distribution (8-11 amino acids) and sequence motifs resembling established peptide-binding profiles of patient-specific HLA alleles (**Fig. 1g, h**). These trends were consistent between matched organoid and 2D cell lines (**Supplementary Fig. 3**). Importantly, gene module scores based on peptides' source genes were specifically enriched in malignant pancreatic cancer cells (*P* ~ 0; XL-minimum hypergeometric test; see Methods), relative to all stromal cell types in a single cell expression dataset derived from more than 20 patient tumor samples (**Fig. 1i**) [26]. These results further reflect the malignant cell purity of organoids, and they suggest that our cancer cell enrichment enables unambiguous identification of peptides derived from genes expressed by pancreatic cancer cells *in vivo*.

A computational proteogenomics approach broadens the PDAC immunopeptidome search space. Identification of peptides in proteomics data critically depends upon the protein sequence search database [27]. To enable identification of peptides derived from sources beyond the reference proteome, we constructed patient-specific databases informed by WGS and RNA-Seq of each sample (**Fig. 2a**). This included recoding the reference proteome with both germline and somatic variants. To account for cases in which nearby variants co-occur on the same haplotype, we also performed variant phasing and integrated this into the recoded search space. Beyond genetic variants, we applied a modified version of a previous pipeline for detection of retained introns (RIs) [8]. We then further augmented our search spaces with protein sequences translated from the open frames of retained introns detected in each sample, as well as the two alternative reading frames to account for possible indels. In addition to mutations and retained introns, we also appended a search space encompassing the entire catalog of novel unannotated reading frames (nuORFs) identified by Ouspenskaia *et al* (2021).

Using this augmented search space, we identified hundreds of peptides derived from nuORFs, dozens from germline variants and RIs, and five from somatic mutations (**Fig. 2b-d**). Notably, we identified 1-2 neoepitopes per sample in 4/7 (57%) PDOs, similar to detection rates previously reported in other cancer types like melanoma [28]. All of these peptides were predicted by our neoepitope pipeline to bind with <300 nM affinity to their corresponding MHC allele (**Fig 2b**). Most notably, one of these peptides derives from a frameshift mutation in the tumor suppressor gene (TSG) *SMAD4* (**Supplementary Fig. 4**). The corresponding indel is detected at 100% DNA VAF in the organoid and conserved in the matched bulk tumor, suggesting the mutation is clonal, as would be predicted for a TSG mutation in PDAC [29]. Prior studies have suggested the possibility of TSG-derived frameshift neoepitopes [30], but this, to our knowledge, is the first empirical identification of one presented in the immunopeptidome. Given that frameshift mutations in *SMAD4*, *TP53*, and *CDKN2A* are frequent drivers of PDAC, we analyzed the TCGA and PancSeq cohorts to estimate both the frequency and recurrence of TSG-FS neoepitopes in PDAC. We find that several recurrent TSG-FS neoepitopes are predicted in cohorts of up to 3% of patients (collectively >10% of all patients), and we find the corresponding mutations are typically expressed at the RNA-Seq level, suggesting TSG frameshift mutations could represent a common source of shared, driver gene-derived neoepitopes (**Supplementary Fig. 4**).

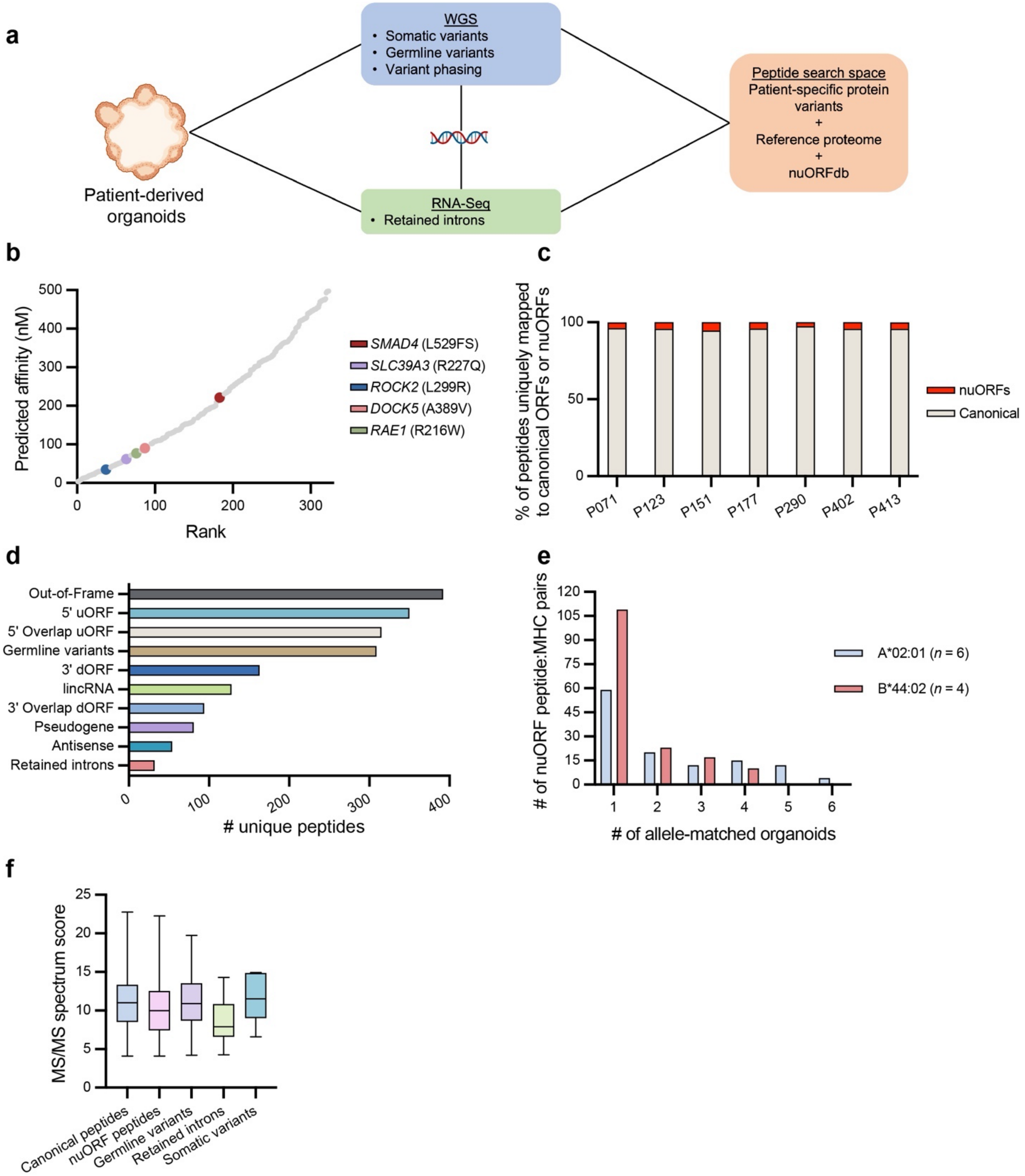


Figure 2: **The pancreatic cancer immunopeptidome harbors hundreds of noncanonical peptides.**

a. Schematic depicting the workflow for establishing patient-specific protein search spaces for mass spectrometry (WGS = whole genome sequencing).
b. Rank-ordered predicted MHC affinity for candidate neoepitopes predicted by the neoepitope prediction pipeline (see Methods). Only neoepitopes that derive from the

same sample and are predicted to bind the superset of the same alleles (B*38:01, A*02:01, B*44:02, A*01:01, C*05:01, B*07:02, and B*35:01) are included. Empirically detected neoepitopes are highlighted by colored, enlarged points. *SMAD4* (L529FS) = frameshift mutation.

c. Proportion of identified peptides that map to canonical ORFS and not nuORFs vs peptides that map to nuORFs but not canonical ORFs (mean nuORF % = 4.05%).
d. The number of identified peptides that uniquely derive from a single noncanonical protein sequence entry across different classes of noncanonical protein sequences. Values are summed from all seven samples.
e. Histogram depicting the frequency of nuORF-derived peptides predicted to specifically bind A*02:01 or B*44:02 and uniquely map to one or more sequences in nuORFdb and not the canonical proteome. Data incorporate all B*44:02 organoids and all A*02:01 organoids.
f. LC–MS/MS Spectrum Mill identification score for peptides derived from different protein sources. Only peptides that map to a single protein sequence entry are considered.

Of the ~78,000 unique peptides detected across all samples, 2,203 derived from nuORFs (**Fig. 2c**). More than 50% of these peptides derive from out-of-frame ORFs or ORFs present in 5' "untranslated" regions (UTR) (**Fig. 2d**). We observed 128 peptides derived from lncRNAs and 81 from pseudogenes. We also found that many nuORF peptides predicted to bind the same MHC alleles were detected in multiple allele-matched samples, including ten detected in all B*44:02 samples (*n* = 4) and four detected in all A*02:01 samples (*n* = 6) (**Fig. 2e**). Like prior studies of melanoma and glioblastoma, these results indicate that noncanonical open reading frames frequently encode peptides presented by MHC-I in pancreatic cancer cells.

At the transcriptional level, we detect thousands of retained introns (RIs) in each organoid sample (**Supplementary Fig. 5)**. In the immunopeptidome, we detected two to nine potential RI-derived peptides in every organoid, and two were shared by more than one sample (**Supplementary Fig. 5**). The difference in number of retained introns detected transcriptionally versus the number detected in the immunopeptidome is similar to observations by Smart *et al.* (2018). Most classes of peptides showed similar Spectrum Mill MS/MS identification scores (**Fig. 2f**), reflecting similarly high quality in peptide-spectrum mapping. However, Spectrum

Mill scores were notably lower for a subset of retained intron peptides, suggesting these may warrant further investigation and reconsideration of the assigned peptide sequence. We manually discarded low-scoring candidate retained intron peptides from subsequent analysis (see Methods).

In total, our results show that retained introns, nuORFs, and somatic mutations can contribute to the immunopeptidome of pancreatic cancer cells, and a subset may thus represent previously unexplored immunotherapy targets for this disease.

A subset of peptide-generating nuORFs and canonical genes exhibit PDAC-restricted transcriptional expression. We next sought to determine whether nuORFs, retained introns, or canonical genes could encode PDAC-specific or PDAC-associated MAPs via cancer restricted expression patterns (**Fig. 3a**). For nuORFs, we examined parental transcript expression across 26 normal tissue categories, encompassing 46 total tissue subtypes, from the Genotype-Tissue Expression (GTEx) project (*n* = 9-10 samples within each tissue subtype; *n* = 255 for Cortex) (**Supplementary Table 1**) [12,31]. We excluded reproductive tissues from consideration, and we examined only the parental transcript that had been assigned to each nuORF by Ouspenskaia *et al*. (2022) (see Discussion). We classified PDAC-specific transcripts as the subset of transcripts whose 90th percentile Transcripts per Million (TPM) value was less than one in all normal tissue categories (see Methods). PDAC-associated transcripts were classified as the subset that exhibited limited expression in certain normal tissues for which tissue-related toxicities could be clinically managed (**Supplementary Fig. 6**) [13]. All final transcripts were also required to be expressed at a TPM > 1 in at least one of the seven organoids profiled by immunopeptidomics. We took a similar approach to curate a list of seven PDAC-specific and PDAC-associated canonical genes encoding peptides in our mass spectrometry data, requiring that the candidates show increased expression in profiled organoids.

With this approach, we identified dozens of annotated and unannotated ORFs that exhibit PDAC-specific or PDAC-associated expression and encode MAPs detected in the immunopeptidome of pancreatic cancer cells (**Fig. 3b-d, g**). In total, we identified 47 MAPs from candidate PDAC-specific nuORFs or genes (35 nuORF transcripts and three canonical genes) and 92 MAPs from candidate PDAC-associated nuORFs or genes (57 nuORF transcripts and four canonical genes). Many of these genes and nuORF-associated transcripts were significantly overexpressed in the majority of the broader PDAC organoid cohort ($n = 48$) relative to normal tissues (**Fig. 3e**).

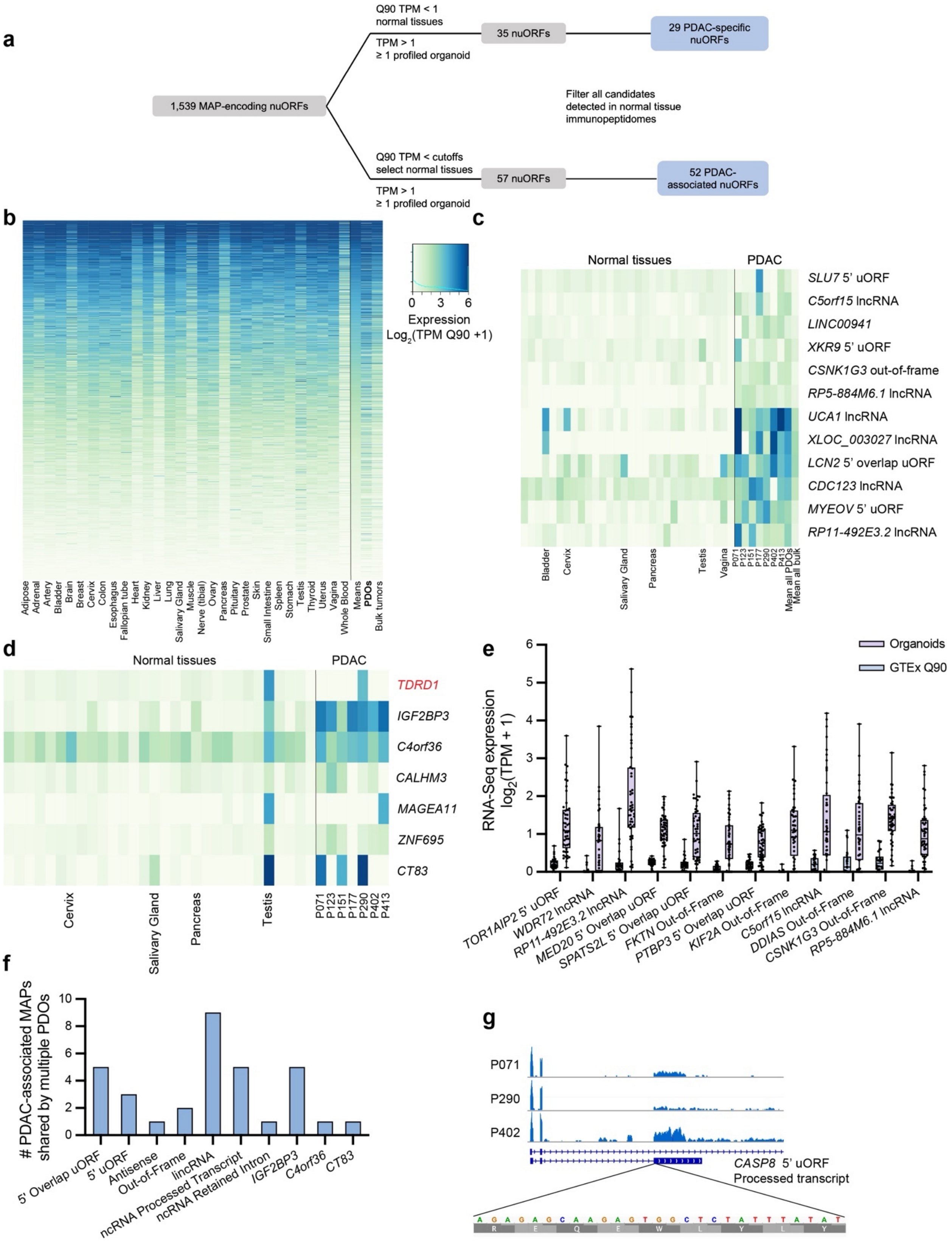

a
1,539 MAP-encoding nuORFs
Q90 TPM < 1
normal tissues
TPM > 1
≥ 1 profiled organoid
35 nuORFs
29 PDAC-specific nuORFs
Filter all candidates
detected in normal tissue
immunopeptidomes
Q90 TPM < cutoffs
select normal tissues
TPM > 1
≥ 1 profiled organoid
57 nuORFs
52 PDAC-associated nuORFs
b
0 3 6
Expression
Log2(TPM Q90 +1)
Adipose
Adrenal
Artery
Bladder
Brain
Breast
Cervix
Colon
Esophagus
Fallopian tube
Heart
Kidney
Liver
Lung
Salivary Gland
Muscle
Nerve (tibial)
Ovary
Pancreas
Pituitary
Prostate
Skin
Small Intestine
Spleen
Stomach
Testis
Thyroid
Uterus
Vagina
Whole Blood
Means
PDOs
Bulk tumors
c
Normal tissues
PDAC
SLU7 5' uORF
C5orf15 lncRNA
LINC00941
XKR9 5' uORF
CSNK1G3 out-of-frame
RP5-884M6.1 lncRNA
UCA1 lncRNA
XLOC_003027 lncRNA
LCN2 5' overlap uORF
CDC123 lncRNA
MYEOV 5' uORF
RP11-492E3.2 lncRNA
Bladder
Cervix
Salivary Gland
Pancreas
Testis
Vagina
P071
P123
P151
P177
P290
P402
P413
Mean all PDOs
Mean all bulk
d
Normal tissues
PDAC
TDRD1
IGF2BP3
C4orf36
CALHM3
MAGEA11
ZNF695
CT83
Cervix
Salivary Gland
Pancreas
Testis
P071
P123
P151
P177
P290
P402
P413
e
Organoids
GTEx Q90
RNA-Seq expression log2(TPM + 1)
0
1
2
3
4
5
6
TOR1AIP2 5' uORF
WDR72 lncRNA
RP11-492E3.2 lncRNA
MED20 5' Overlap uORF
SPATS2L 5' Overlap uORF
FKTN Out-of-Frame
PTBP3 5' Overlap uORF
KIF2A Out-of-Frame
C5orf15 lncRNA
DDIAS Out-of-Frame
CSNK1G3 Out-of-Frame
RP5-884M6.1 lncRNA
f
# PDAC-associated MAPs shared by multiple PDOs
0
2
4
6
8
10
5' Overlap uORF
5' uORF
Antisense
Out-of-Frame
lincRNA
ncRNA Processed Transcript
ncRNA Retained Intron
IGF2BP3
C4orf36
CT83
g
P071
P290
P402
CASP8 5' uORF
Processed transcript
A G A G A G C A A G A G T G G C T C T A T T T A T A T
R E Q E W L Y L Y

Figure 3: **A subset of canonical and noncanonical transcripts exhibit PDAC-restricted expression patterns and encode MAPs presented by multiple patient samples.**

a. Schematic depicting the filtering process used to nominate PDAC-specific and PDAC-associated nuORFs.
b. Heatmap depicting the expression of all transcripts associated with 1,539 nuORFs with unambiguously mapped peptides from the immunopeptidome of PDAC organoids. Expression values were calculated using $\log_2$(TPM+1) of the 90th percentile (Q90) value of each nuORF in each of 30 normal tissues. NuORFs are ordered according to the mean Q90 expression value calculated across normal tissues (third column from right). The $\log_2$(TPM+1) value based on Q90 TPM across all PDAC organoids (*n*=48) and bulk tumors (*n*=49) are depicted in the two rightmost columns. Expression values were capped at $\log_2$(TPM+1) = 6 for plotting. A black bar separates data for normal tissues (left) vs mean values and PDAC samples (right).
c. Heatmap depicting the expression of the six PDAC-specific (top) and six PDAC-associated (bottom) nuORF transcripts with the highest average expression in seven organoids profiled with immunopeptidomics. Expression is based on the Q90 TPM value in normal tissues (30 leftmost columns), the calculated TPM value in each of the seven organoids (next seven columns), and the mean TPM value across all organoids and bulk tumors (two rightmost columns). All values are expressed as $\log_2$(TPM+1). Select normal tissues are labeled along the bottom of the heatmap.
d. Heatmap depicting expression of three PDAC-specific (*TDRD1*, *CALHM3*, *MAGEA11*) and four PDAC-associated (*IGF2BP3*, *C4orf36*, *ZNF695*, and *CT83*) canonical genes. One gene (*TDRD1*) with one or more peptides detected in normal tissue immunopeptidomes is labeled in red. All values are expressed as $\log_2$(TPM+1) as described in (C).
e. Box plots depicting the median (middle line), 25th and 75th percentiles (box limits) and minimum and maximum (whiskers) expression values of the top 12 nuORF-associated transcripts whose 25th percentile expression across PDAC organoids is two times the maximum Q90 TPM across normal tissues. Dots reflect expression values of individual tissues (*n* = 29, excluding testis) and PDAC organoids (*n* = 48).
f. Bar plots depicting the total number of MAPs detected in two or more organoid samples that derive from nuORFs or canonical genes classified as PDAC-associated or PDAC-specific.
g. RNA-Seq coverage across an example PDAC-associated nuORF transcript (bottom annotation track) derived from *CASP8*. Plots depict coverage in three different B*44:02 PDAC samples in which the same peptide was detected (bottom). This nuORF also gives rise to a second peptide detected in two A*03:01 samples (not shown).

To further validate the PDAC-restricted expression of these MAPs, we also analyzed data derived from the immunopeptidomes of 31 different normal tissues across a total of 245 autopsy specimens (excluding reproductive tissues), and we removed peptides that were derived from parental genes or nuORF-associated transcripts encoding the same or different 8-11mer peptides in normal tissue immunopeptidomes [32]. After this final filtration, we retained 37 MAPs from PDAC-specific candidate genes or nuORFs (29 nuORF transcripts and two genes) and 71 MAPs from PDAC-associated candidate genes or nuORFs (52 nuORF transcripts and four genes).

A subset of PDAC-specific and -associated MAPs were detected in multiple PDO samples and may thus represent shared immunotherapy targets (**Fig. 3f**). Four nuORF-derived peptides were particularly recurrent, including one 5' overlap uORF peptide detected in five A*02:01 samples (transcript assigned by Ouspenskaia *et al.*: ENST00000512813.5_1), two ncRNA peptides each detected in four A*02:01 samples (transcripts assigned by Ouspenskaia *et al.*: ENST00000509913.1_1 and ENST00000478653.6_1), and one 5' uORF peptide detected in three B*44:02 samples (transcript assigned by Ouspenskaia *et al.*: ENST00000490682.5_1) (**Fig. 3d**). Notably, multiple organoids shared canonical peptides derived from *IGF2BP3*, encoding Insulin-like growth factor 2 mRNA-binding protein 3, which has been shown to promote metastasis in pancreatic cancer (**Fig. 3d, f**) [33]. We also identified further evidence of PDAC-specific IGF2BP3 protein expression by examining the Human Protein Atlas (**Supplementary Fig. 7**) [34]. Though *IGF2BP3* qualifies as a PDAC-associated gene according to our transcriptomic analysis, we initially filtered it due to a single peptide detected in a normal cerebellum immunopeptidome specimen. Since this peptide was also mappable to a nuORF transcript, *IGF2PB3* may still represent a promising target for immunotherapy, especially given that it encodes four different peptides that each are detected in the immunopeptidome of three or more PDAC organoid samples.

We analyzed another cohort of normal GTEx tissues for retained introns by employing the same intron retention pipeline applied to our PDO samples. We found that the number of retained introns in different normal tissues generally tracked with relative estimates established by

Braunschweig *et al*. (2014), with, as examples, heart, muscle, and colon tissues exhibiting lower levels of intron retention [35]. To designate retained introns that define each normal tissue category, we took the intersection of introns detected for all samples within each tissue (see Methods; **Supplementary Fig. 5**). We then took the union of all retained introns called for each tissue and used this cohort to filter introns in organoids down to a set of PDAC-specific retained introns (**Supplementary Fig. 5**). After filtering, we identified hundreds of PDAC-specific retained introns shared by multiple organoids (**Supplementary Fig. 5**). Despite an apparently high load of cancer-specific retained introns, only 2/48 intron-derived peptides detected in our immunopeptidome dataset derive from cancer-specific retained introns (**Supplementary Fig. 5**). The Spectrum Mill scores for these two peptides were >10, supporting a correct peptide sequence assignment.

Generation of T cells specific to PDAC-associated MAPs using healthy donor PBMCs. PDAC-specific and -associated MAPs could represent valuable immunotherapy targets for cancer vaccination strategies or adoptive cell therapy (ACT) with MAP-specific T cells. Prior preclinical and clinical studies suggest that the immune system can fail to recognize tumor-specific antigens expressed in PDAC [6,36]. This suggests that the TCR repertoire of patient TILs may not be an ideal source for gauging the immunogenicity of PDAC-associated MAPs, nor an ideal source for identifying T cells suitable for expansion and ACT. Thus, we opted to employ partially HLA-matched healthy donor PBMCs for the expansion and identification of T cells specific to PDAC-associated or PDAC-specific MAPs [37]. We adapted a procedure described by Rollins and colleagues to prime and repeatedly stimulate CD8$^+$ T cells with autologous monocyte-derived dendritic cells (Mo-DCs) pulsed with PDAC-associated or PDAC-specific MAPs (**Fig. 4a**; **Supplementary Fig. 8**) [38]. For predicted A*02:01 binders, we stained CD8$^+$ T cells with peptide-loaded tetramers conjugated to two distinct fluorophores to enumerate antigen-specific T cells. After expansion and repeated stimulation, we used this method to compare the frequency of dual-tetramer$^+$ cells in the stimulated cohorts to the frequency among autologous CD8$^+$ T cells not previously stimulated with the peptide of interest. For peptides predicted to bind other alleles (i.e., not A*02:01), we examined the frequency of CD8$^+$ T cells expressing interferon gamma and tumor necrosis factor alpha after co-culture with peptide-pulsed DCs. Notably, of the six nuORF-derived peptides tested, three generated sizable populations of antigen-specific CD8$^+$ T cells from undetectable precursor populations (**Fig. 4b-c; Supplementary Fig. 8**). The immunogenicity of these peptides was validated in additional experiments with PBMCs derived from a separate healthy donor, or with orthogonal validation using peptide-mediated stimulation of effector cytokine production (**Fig. 4d**). Of note, one peptide, ALSPRNWTL, derives from a PDAC-associated lncRNA. The corresponding nuORF overlaps both the transcript assigned by Ouspenskaia *et al*. (2022) and two others derived from the same lncRNA. The assigned transcript is expressed at Q90 TPM < 0.9 in all normal GTEx tissues, except whole blood (Q90 TPM = 1.3), and is expressed at TPM = 1.6 and 3.2 in the two different organoids in which the peptide was detected. The other two overlapping transcripts are expressed more highly in organoids and more broadly in normal tissues, challenging whether the transcript assigned by Ouspenskaia *et al*. (2022) is the correct source of the peptide (see Discussion). Regardless of the transcript source, this lncRNA does not generate a peptide detected in any normal tissue immunopeptidome, suggesting at least a pattern of cancer-restricted translation. In total, these results indicate that T cells specific to PDAC-associated, nuORF-derived MAPs can be readily generated from healthy donor PBMCs, which may thus represent a potent source for T cell-based immunotherapies for PDAC.

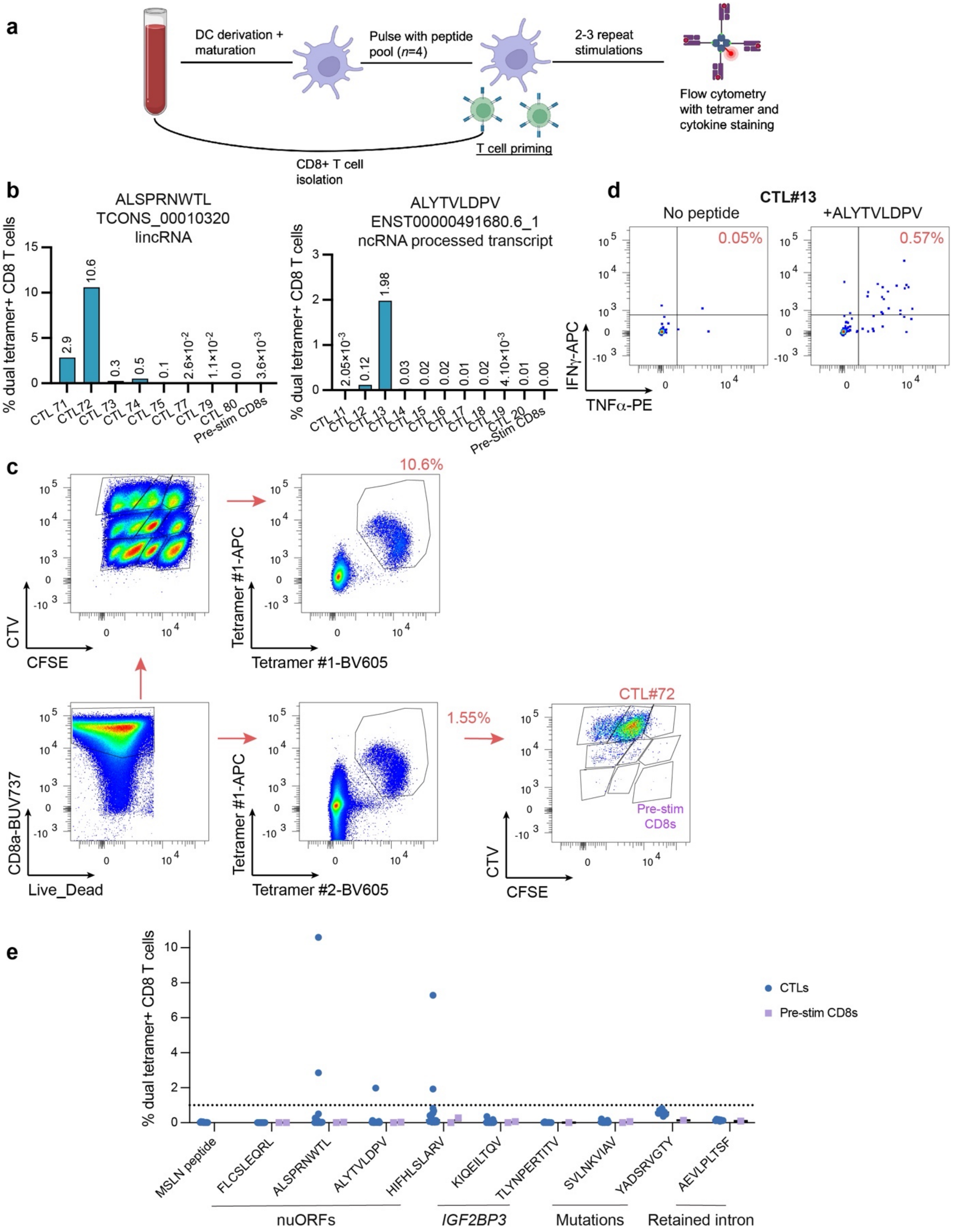


Figure 4: **T cells specific to PDAC-associated nuORF MAPs are readily generated from healthy donor PBMCs.**

a. Schematic depicting the procedure for initial T cell priming and subsequent restimulations using partially HLA-matched healthy donor PBMCs. 20 CD8$^+$ cytotoxic lymphocyte lines (CTL) were derived from each donor. Each set of 10 CTLs was primed and restimulated two to three times with a pool of four specific peptides.
b. The percentage of dual-tetramer-positive cells among live CD8$^+$ cells using A*02:01 tetramers loaded with a peptide derived from a PDAC-associated lncRNA (left) or a peptide derived from a ncRNA transcript whose expression is TPM < 1 in all normal tissue samples (right). Bars indicate the percentage (value labeled above bars) across multiple CTLs derived from the same donor and expanded with the same peptide pool. The percentage among autologous CD8$^+$ cells not previously stimulated with the peptide is depicted in the rightmost column. Only CTL lines that had a sufficient number of cells after expansion were included in the analysis.
c. Flow cytometry plots depicting a large population of expanded CD8$^+$ T cells stained with dual tetramers loaded with lincRNA-derived peptide, ALSPRNWTL. 10 independent CTL lines were pooled for analysis and demultiplexed using fluorescent staining with CellTrace™ Violet (CTV) and CarboxyFluoroscein Succinimidyl Ester (CFSE). 1.55% of all live CD8$^+$ cells were dual tetramer positive (bottom middle), and most of these were cells derived from CTL #72 (bottom right). Dual-tetramer-positive cells represented 10.6% of all CTL #72 live CD8+ cells (upper right) and a negligible fraction of control cells not previously stimulated with the peptide ("Pre-stim CD8s" labeled in purple in bottom right panel).
d. Flow cytometry plot depicting the percentage of CD8$^+$ cells expressing or not expressing IFNγ and/or TNFα after treatment with ncRNA-derived ALYTVLDPV, or no peptide as a control. These data are derived the same donor depicted in the right panel of (b).
e. The percentage of dual-tetramer-positive cells among live CD8$^+$ cells using A*02:01 tetramers loaded with peptides derived from nuORFs, *IGF2BP3*, missense mutations, or a retained intron. Peptide source categories are indicated below the X-axis. CTLs previously stimulated with the indicated peptide are labeled in blue, and autologous CD8$^+$ CTLs not previously stimulated are labeled in purple. All technical replicates (different CTLs; same donor) and biological replicates (different CTLs; different donor) are pooled together in each column. The percentage of dual-tetramer-positive cells among live CD8$^+$ cells using A*02:01 tetramers loaded with an irrelevant mesothelin (MSLN)-derived peptide are included on the left as an additional negative control.

In contrast to nuORFs, we did not observe a T cell response against any of the PDAC missense-derived neoepitopes tested (**Fig. 4e; Supplementary Fig. 8**), though we did observe a strong and robust response against two melanoma-derived neoepitopes reported in a previous study using a similar healthy donor PBMC-based assay (**Supplementary Fig. 8**) [37]. We also did not observe a response against any of three tested *IGF2BP3*-derived peptides, nor against a retained intron-derived peptide (**Fig. 4e; Supplementary Fig. 8**). Of note, we detected *IGF2BP3*-derived peptides in the immunopeptidome of three thymus donor tissues, so it is likely that central tolerance excludes *IGF2BP3*-specific peptides from the naïve T cell repertoire. However, *IGF2BP3* could still represent a valuable target for adoptive cell therapy if *IGF2BP3*-peptide-specific TCRs could be identified or otherwise engineered. In total, these results suggest that nuORFs may represent a more immunogenic class of epitopes than canonical genes or somatic mutations, though more expansive testing with numerous different PBMC donors and additional epitopes is warranted.

Discussion

PDAC is notoriously recalcitrant to immunotherapy [39,40]. This has been partly attributed to its mutational burden [1,2]; however, recent studies have elucidated numerous non-mutational sources of cancer-restricted antigens that have remained largely unexplored in PDAC [8,12,41,42]. In this study, we performed the deepest characterization of the PDAC immunopeptidome to-date, uncovering dozens of MAPs from noncanonical sources with transcription or translation limited to PDAC cells or PDAC cells and select normal tissues. Importantly, we were able to unambiguously confirm the malignant origin of these peptides by employing patient-derived PDAC organoids, thereby avoiding signals from stromal cell populations that confound bulk pancreatic tumor datasets.

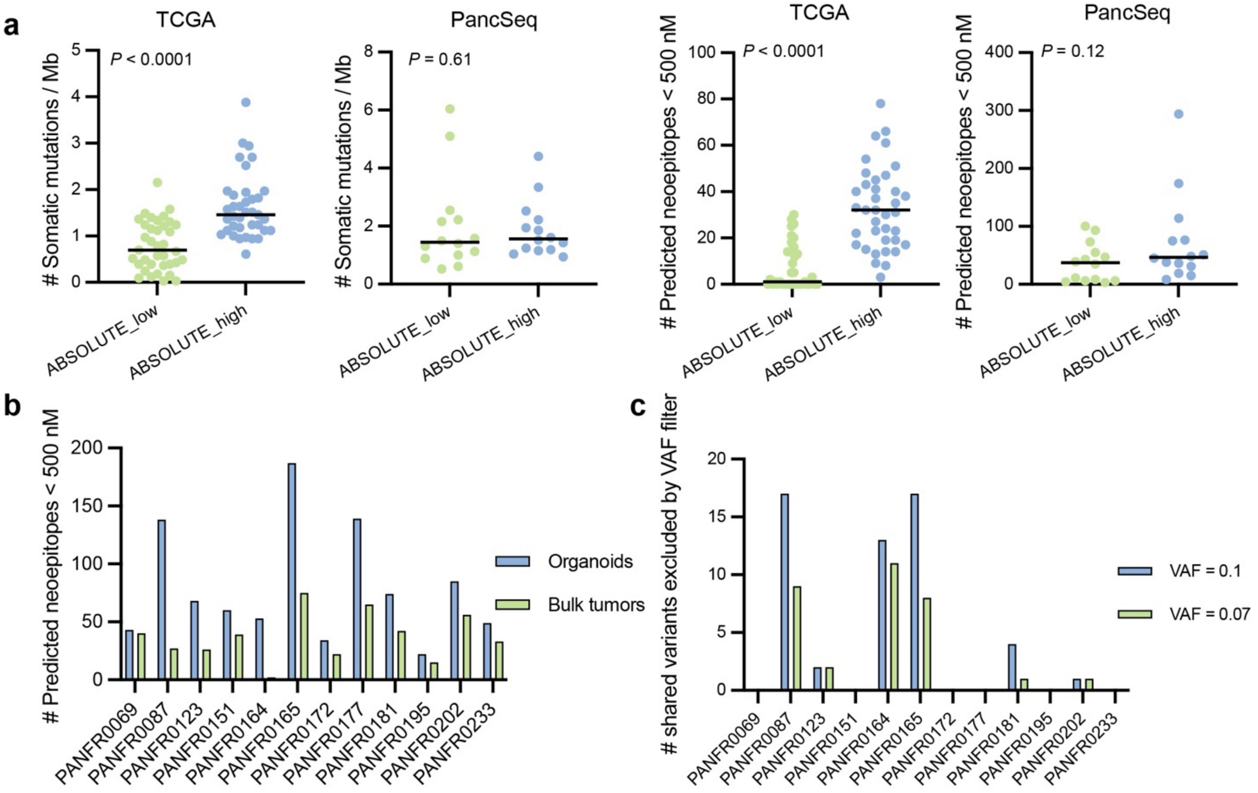
In additional to noncanonical sources, we also frequently identified mutation-derived neoepitopes in most organoid samples. This suggests that pancreatic tumors do in fact present neoepitopes despite a low mutational burden. By comparing predicted neoepitope burden across primary and metastatic bulk tumors and tumor-derived organoids, we found that the overall neoepitope burden in PDAC may be moderately higher than prior estimates based on primary tumors [4]. We further found that even for metastatic samples, tumor-derived organoids typically harbor several neoepitope-encoding mutations that would be excluded by variant allele frequency filters used in most bioinformatic pipelines [22]. In light of this, current clinical efforts aimed at neoepitope-based vaccinations may benefit by expanding and sequencing tumor-derived organoids prior to neoepitope prediction and prioritization.

In the case of retained introns, we found that while pancreatic cancer cells often present retained intron-derived epitopes, most of those detected are not cancer-specific. Prior studies have often limited their analysis of retained introns in normal tissues to the tissue-of-origin [8,9]; however, it has been shown that intron retention profiles can be lineage-specific and thus considerably discordant among different tissue types [35,43]. This emphasizes the importance of examining a broad cohort of normal tissues when designating cancer-specific retained introns. Our results warrant caution for clinical efforts that may employ retained intron-derived epitopes as therapeutic targets. However, our identification of a PDAC-restricted retained intron epitope suggests that a subset of patients may indeed harbor therapeutically actionable epitopes derived from retained introns.

We found that pancreatic cancer can harbor hundreds of MAPs derived from lncRNAs, 5' UTRs, pseudogenes, out-of-frame ORFs, and canonical protein-coding ORFs. We systematically interrogated source gene expression patterns in the transcriptomes and immunopeptidomes of hundreds of normal tissue specimens. Through this process, we were able to designate a rigorous cohort of previously unreported candidate PDAC-specific or PDAC-associated MAPs. These included a subset encoded by canonical genes like *IGF2BP3*. We were unable to generate a T cell response against *IGF2BP3*-derived epitopes. This is consistent with our identification of *IGF2BP3*-derived peptides in thymic immunopeptidomes, direct evidence that canonical genes like *IGF2BP3* are more likely subject to AIRE (autoimmune regulator)-induced thymic expression and central tolerance than nuORFs. In contrast, nuORF-derived peptides in the cancer immunopeptidome may be due to cancer-specific transcription and/or translation that does not occur in the thymus. Indeed, most nuORF-associated transcripts nominated by RNA-Seq analyses were not detected in the immunopeptidomes of any normal tissue, including the thymus. One limitation of our study involves the construction of the nuORF database published by Ouspenskaia *et al*. (2022). In this study, nuORFs were inferred from ribosome profiling data. The subsequent nuORF ID was assigned to a specific transcript overlapping the nuORF coordinates. Upon manual review, we noted that many nuORFs overlap several transcripts without an intervening splice site that could be used to distinguish the proper isoform assignment (i.e., the transcript isoform that most likely generated the peptide). In the absence of sample-matched RNA-Seq, Ouspenskaia *et al*. (2022) arbitrarily assigned a transcript to the nuORF's ID. Consequently, the normal tissue expression analysis in both their study and our study is confounded in cases where this assignment may have been incorrect. To address these ambiguous assignments, we are restructuring the nuORF database by re-assigning each ambiguous nuORF to the candidate transcript exhibiting the highest expression in organoids with a nuORF-encoded peptide. This will enable more precise curation of nuORF-derived MAPs with PDAC-restricted expression patterns.

Importantly, we found that T cells specific to nuORF-derived peptides were easily generated from healthy donor immune cell populations. In all cases, we could not identify a preexisting population of nuORF-specific T cells, suggesting that these cells began as naïve precursor populations. In total, our results suggest that nuORFs are an especially immunogenic class of PDAC-restricted epitopes, and naïve T cell repertoires represent a convenient source for developing nuORF-specific T cells. This finding has implications for both cancer vaccine strategies and adoptive cell therapy, both of which may benefit from targeting peptides derived from PDAC-specific or PDAC-associated nuORFs, or nuORFs that may be translated in a PDAC-restricted fashion. In general, the MAPs identified in our study represent a promising class of targets for a broad array of traditional and next-generation immunotherapies.

Supplementary Figures and Legends:



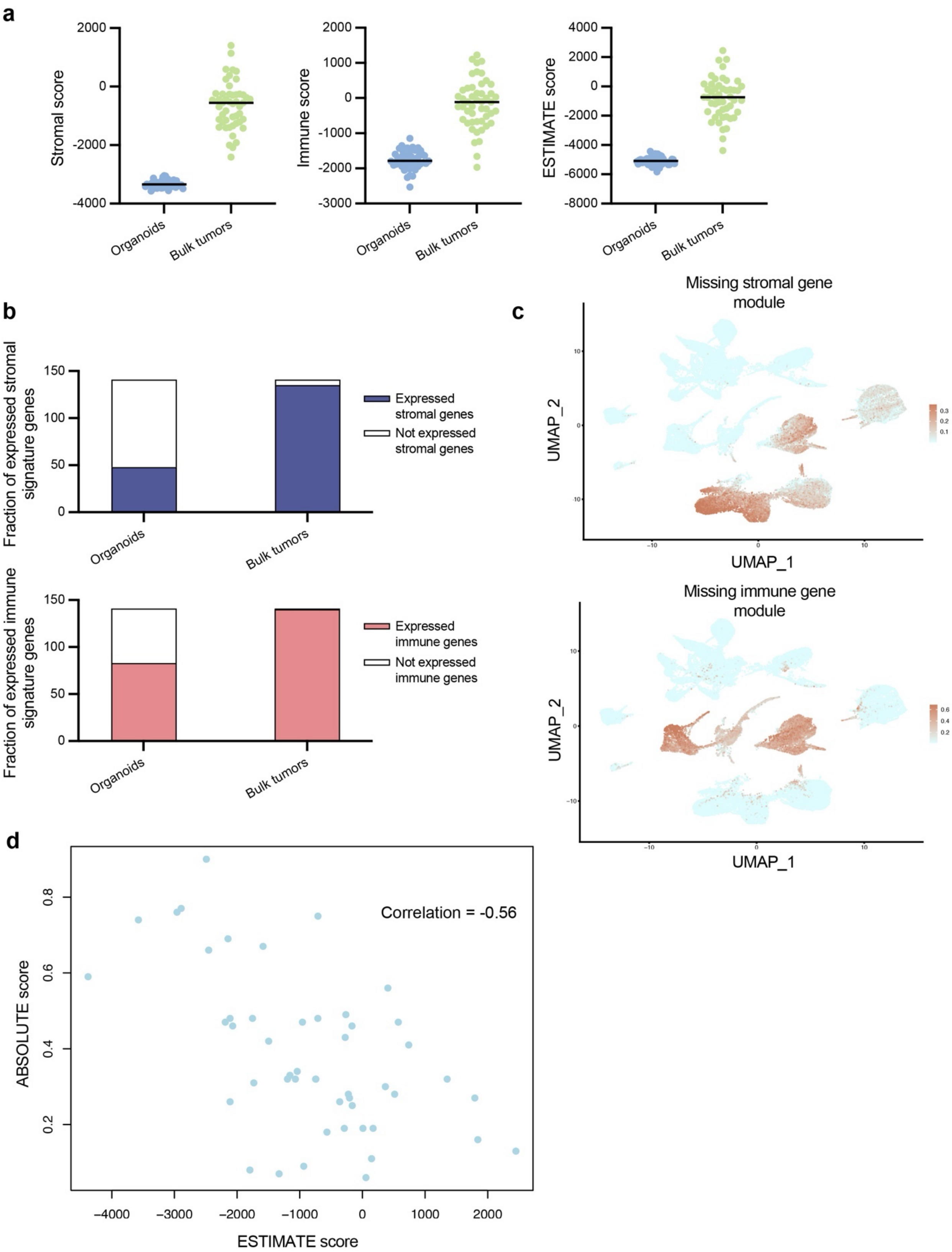
d

Sample	ABSOLUTE purity
PANFR0069	0.32
PANFR0087	0.19
PANFR0123	0.32
PANFR0151	0.53
PANFR0164	0.06
PANFR0165	0.16
PANFR0172	0.48
PANFR0177	0.47
PANFR0181	0.24
PANFR0195	0.66
PANFR0202	0.74
PANFR0233	0.59

Supplementary Figure 1: Low tumor purity diminishes estimated TMB and predicted neoepitope burden in PDAC.

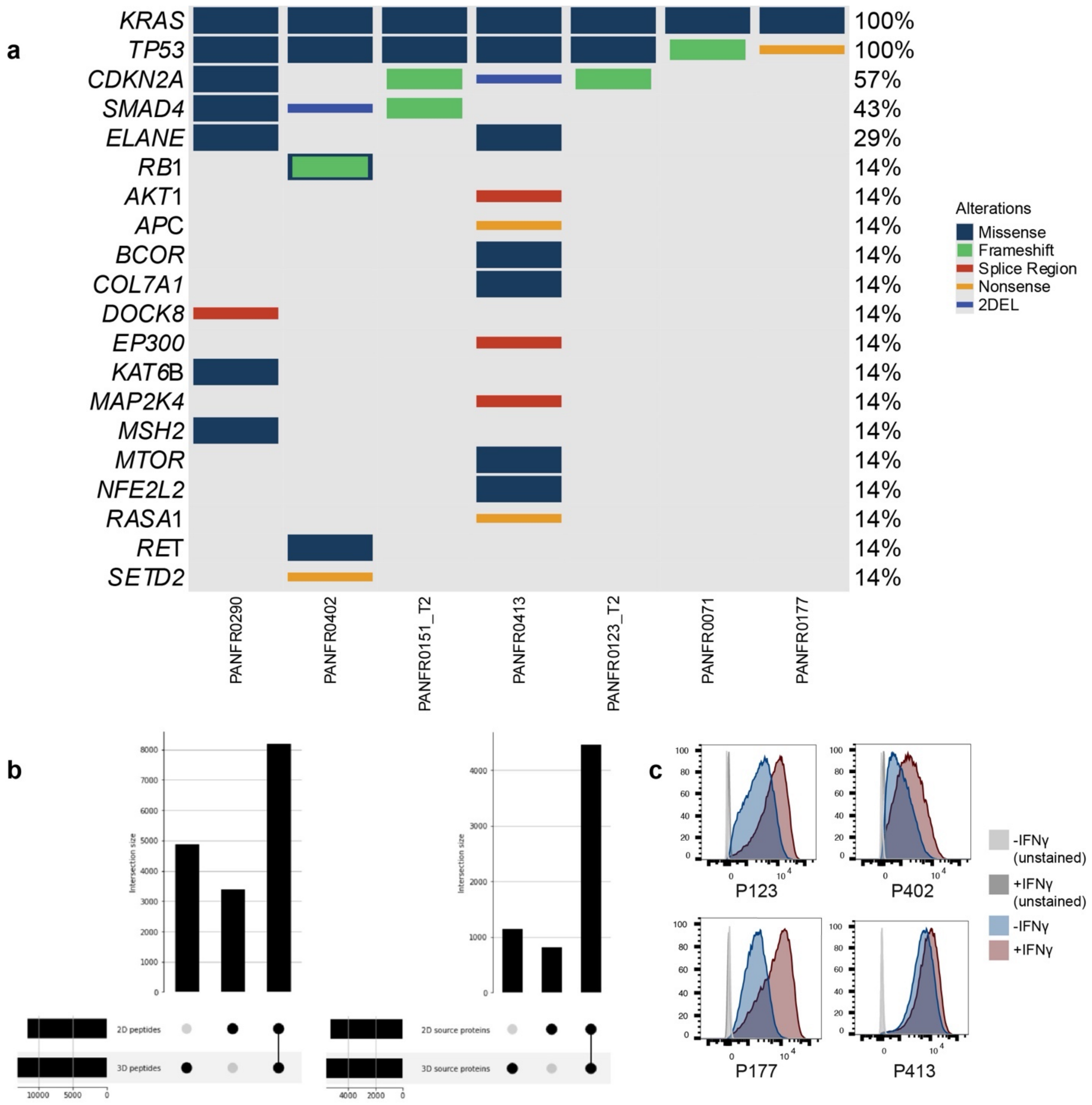
a. Comparison of TMB (left two plots) and predicted neoepitope burden (two right plots) in low- vs high-purity bulk tumors derived from TCGA and PancSeq cohorts.
b. The number of predicted neoepitopes in organoids vs matched bulk tumor samples.

c. The number of variants encoding candidate neoepitopes that are excluded by a traditional VAF cutoff of 0.1 or a more conservative cutoff of 0.07 across samples. Only variants called for both the organoid and bulk tumor are considered.
d. ABSOLUTE-estimated bulk tumor purity for all 12 samples with matched organoids.



Supplementary Figure 2: Stromal gene expression signatures significantly contaminate the transcriptomes of bulk PDAC tumor samples.

a. ESTIMATE-calculated stromal, immune, and ESTIMATE scores for all PDAC organoids (n=48) and all PancSeq bulk tumors (n=49).
b. The fraction of genes comprising ESTIMATE's stromal and immune modules that are detectably expressed (mean RSEM raw counts > 3) across all organoids or bulk tumor samples.
c. Modules of stromal and immune signature genes detected in bulk tumors but not organoids derive from non-malignant cell types.
d. ESTIMATE scores show a modest negative correlation with ABSOLUTE-calculated tumor purity in bulk PDAC samples.

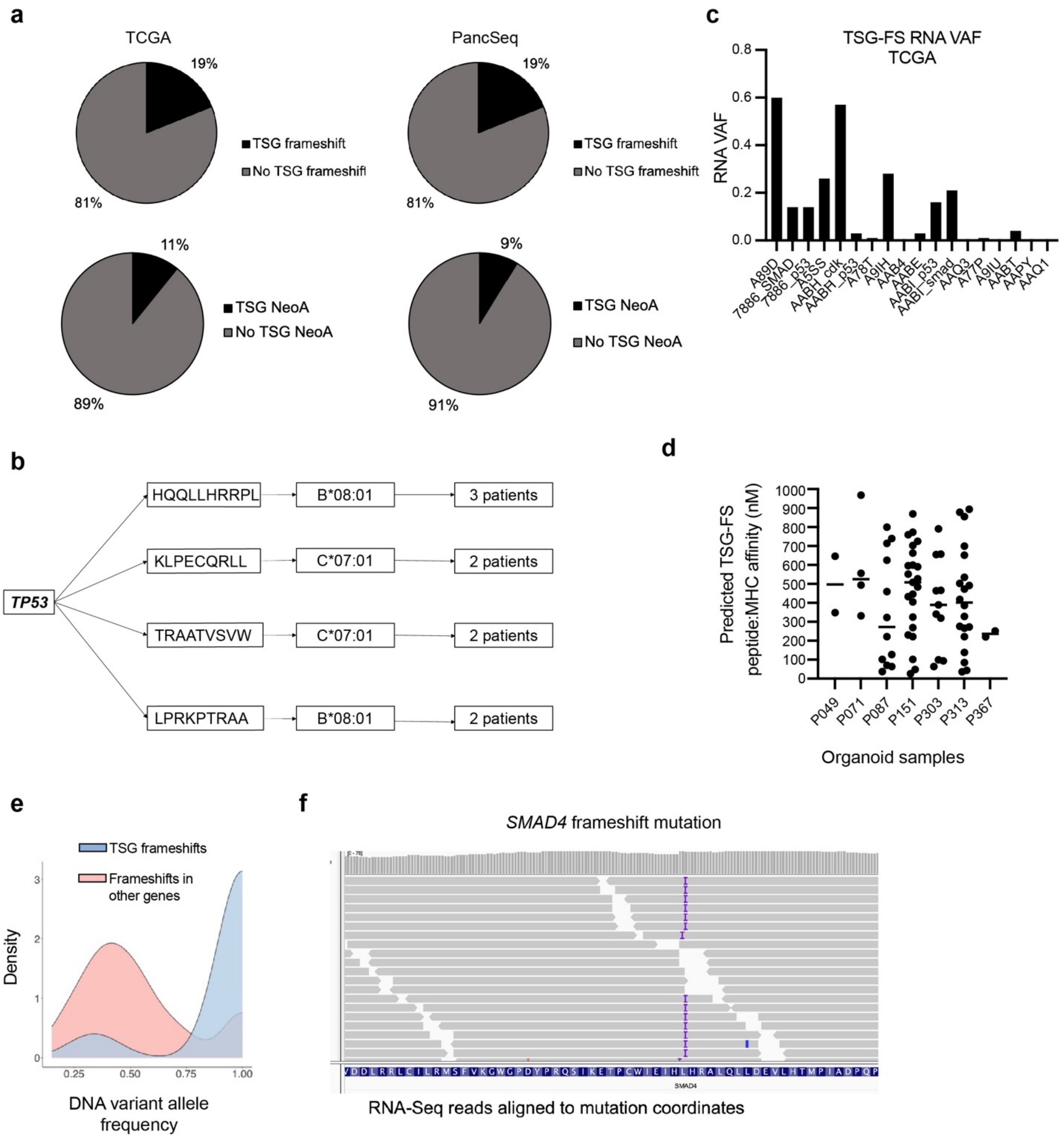


d

Organoid line	HLA-A		HLA-B		HLA-C	
PANFR0123	A*02:01	A*25:01	B*13:02	B*57:01	C*06:02	C*06:02
PANFR0413	A*02:01	A*24:02	B*07:02	B*35:01	C*04:01	C*07:02
PANFR0402	A*02:01	A*03:01	B*44:02	B*35:03	C*04:01	C*05:01
PANFR0290	A*02:01	A*01:01	B*44:02	B*41:01	C*05:01	C*17:01
PANFR0071	A*02:01	A*31:01	B*44:02	B*51:01	C*05:01	C*15:02
PANFR0151	A*25:01	A*03:01	B*27:05	B*38:01	C*01:02	C*12:03
PANFR0177	A*02:01	A*02:01	B*44:02	B*57:01	C*06:02	C*05:01

Supplementary Figure 3: Mutational profile and HLA types of organoids and derived 2D cell lines expanded for immunopeptidomics.

a. Oncoprint depicting the most commonly mutated genes among seven organoids profiled with immunopeptidomics. Legend on the right indicates mutation class.
b. UpSet plots depicting the number of MHC class I-bound peptides (left) or peptide source proteins (right) shared between or unique to 2D cells and organoids derived from PANFR0413.
c. Flow cytometry quantifying the surface expression of MHC class I proteins in pancreatic cancer organoids with (red) or without (blue) interferon gamma stimulation, compared to unstained controls (gray). These plots depict the results for the four organoid samples not depicted in Figure 1e.
d. HLA alleles for each of the seven PDAC patient samples tested via immunopeptidomics.



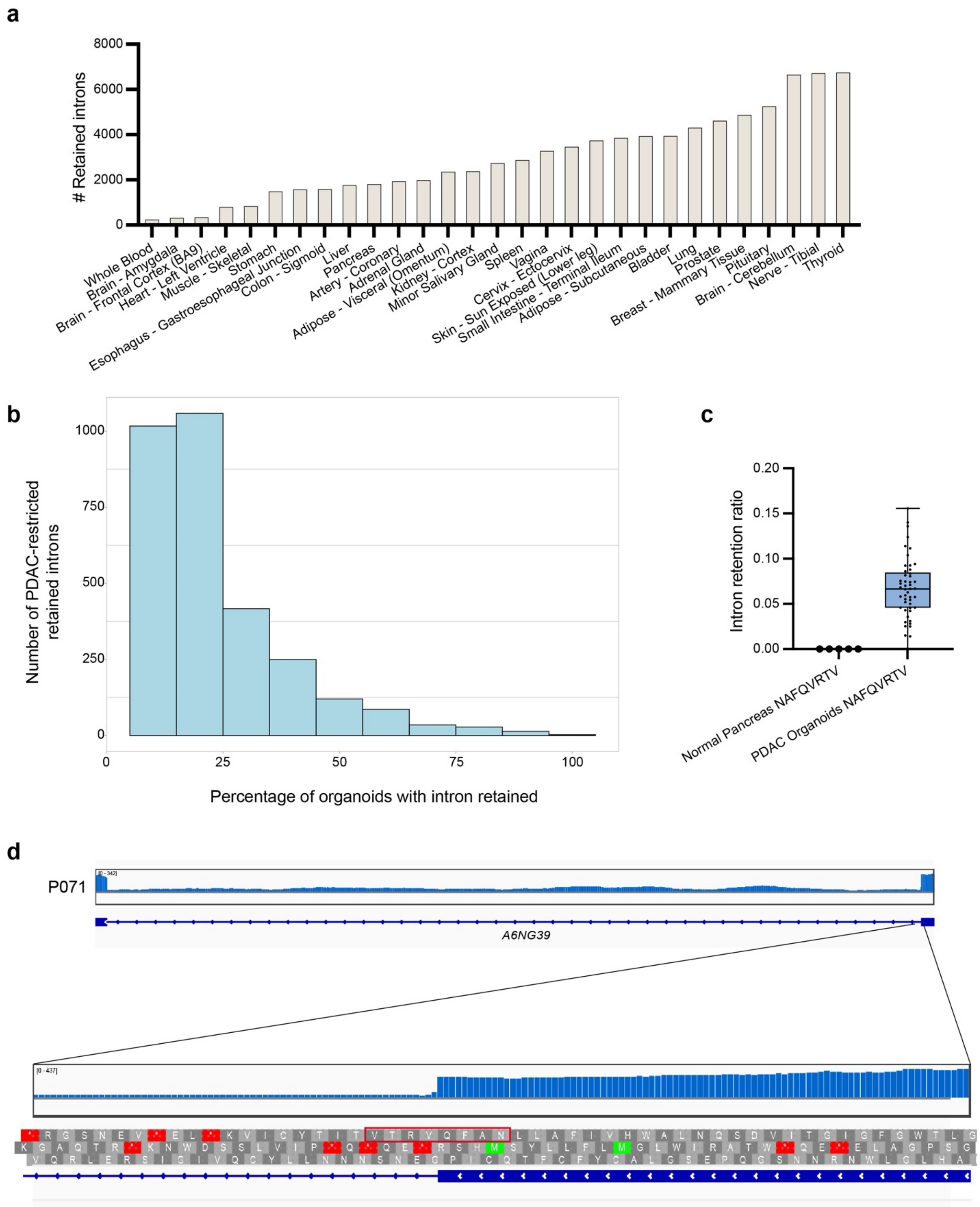
g

```
Score             Expect    Identities
22.3 bits(45)     35        7/12(58%)

Query  1     YHADC-----RP  7
             YHADC     RP
Sbjct  2142  YHADCLNLTKRP  2153
```

Supplementary Figure 4: Frameshift mutations in tumor suppressor genes are predicted to generate shared neoepitopes in multiple cohorts of PDAC patients.

a. The fraction of patients with a predicted TSG frameshift (top) or predicted TSG neoantigen (bottom) in TCGA and PancSeq.
b. Example of several *TP53*-derived neoepitopes predicted to bind to the same MHC allele in multiple samples from TCGA.
c. RNA variant allele frequency of indels in tumor suppressor genes for all TCGA samples with frameshift mutations in either *TP53*, *CDKN2A*, or *SMAD4*.
d. Predicted TSG frameshift-derived peptide affinity for patient-specific MHC alleles for all organoid samples with a predicted TSG frameshift neoepitope.
e. 90% of frameshift mutations in TSGs (blue curve) are present at 100% DNA variant allele frequency in organoid samples, a significant increase compared to frameshift mutations in other genes (pink curve).
f. RNA-Sequencing reads aligned to the frameshift mutation that gives rise to the *SMAD4* frameshift neoepitope (YHADCRPTTF). Purple bars indicate the presence of an indel.
g. BLAST results for the best wild-type peptide match comparing YHADCRPTTF against the reference proteome, confirming the peptide is unique to the frameshifted protein.

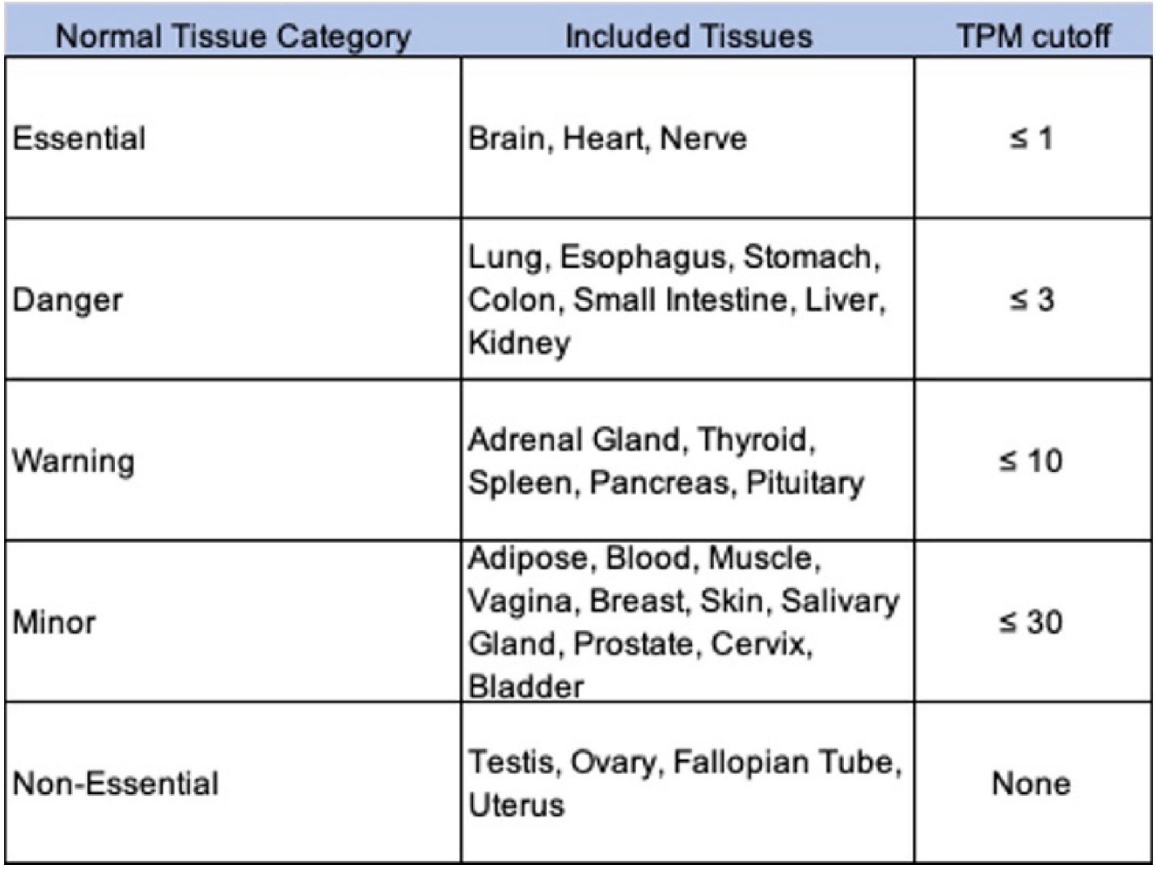


Supplementary Figure 5: **PDAC organoids harbor hundreds of shared, cancer-restricted retained introns.**

a. Number of retained introns called for each GTEx normal tissue subtype

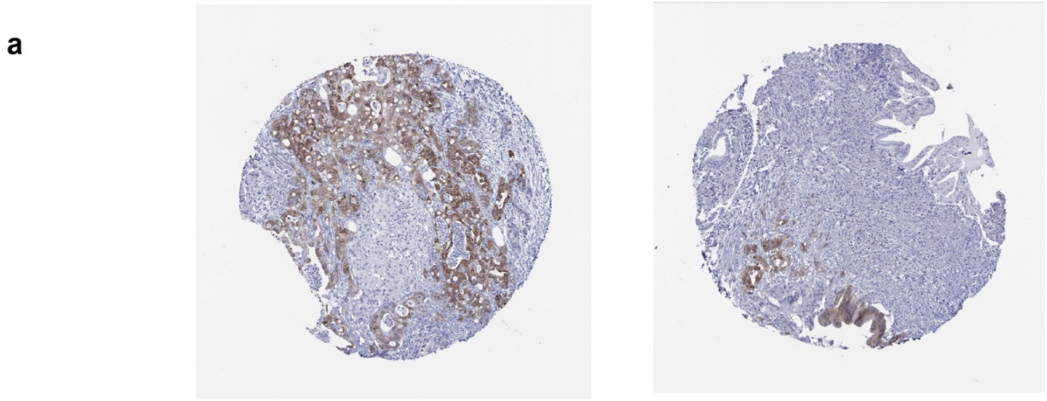
b. Frequency distribution of shared, PDAC-restricted retained introns in organoids (*n*=45). Plot begins with introns retained in 10% of organoids.
c. Intron retention ratios for a PDAC-restricted, *A6NG39*-derived retained intron in five normal pancreatic tissue specimens (GTEx; left) and all PDAC organoids (right). Peptide sequence reflects the empirically detected MHC-bound peptide from sample P071.
d. RNA-Seq coverage of the retained intron encoding peptide NAFQVRTV in the sample P071. Highlight shows a zoomed-in version including the peptide sequence highlighted in red (bottom).

a

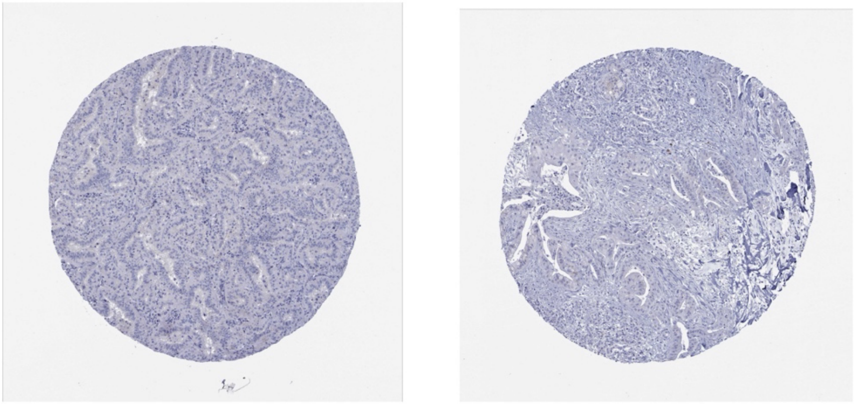


Normal Tissue Category	Included Tissues	TPM cutoff
Essential	Brain, Heart, Nerve	≤ 1
Danger	Lung, Esophagus, Stomach, Colon, Small Intestine, Liver, Kidney	≤ 3
Warning	Adrenal Gland, Thyroid, Spleen, Pancreas, Pituitary	≤ 10
Minor	Adipose, Blood, Muscle, Vagina, Breast, Skin, Salivary Gland, Prostate, Cervix, Bladder	≤ 30
Non-Essential	Testis, Ovary, Fallopian Tube, Uterus	None

Supplementary Figure 6: **Thresholds used to designate PDAC-associated genes and nuORFs from Bradley *et al*. 2020.**
a. Table describing the different tissues tiers and corresponding TPM cutoffs.



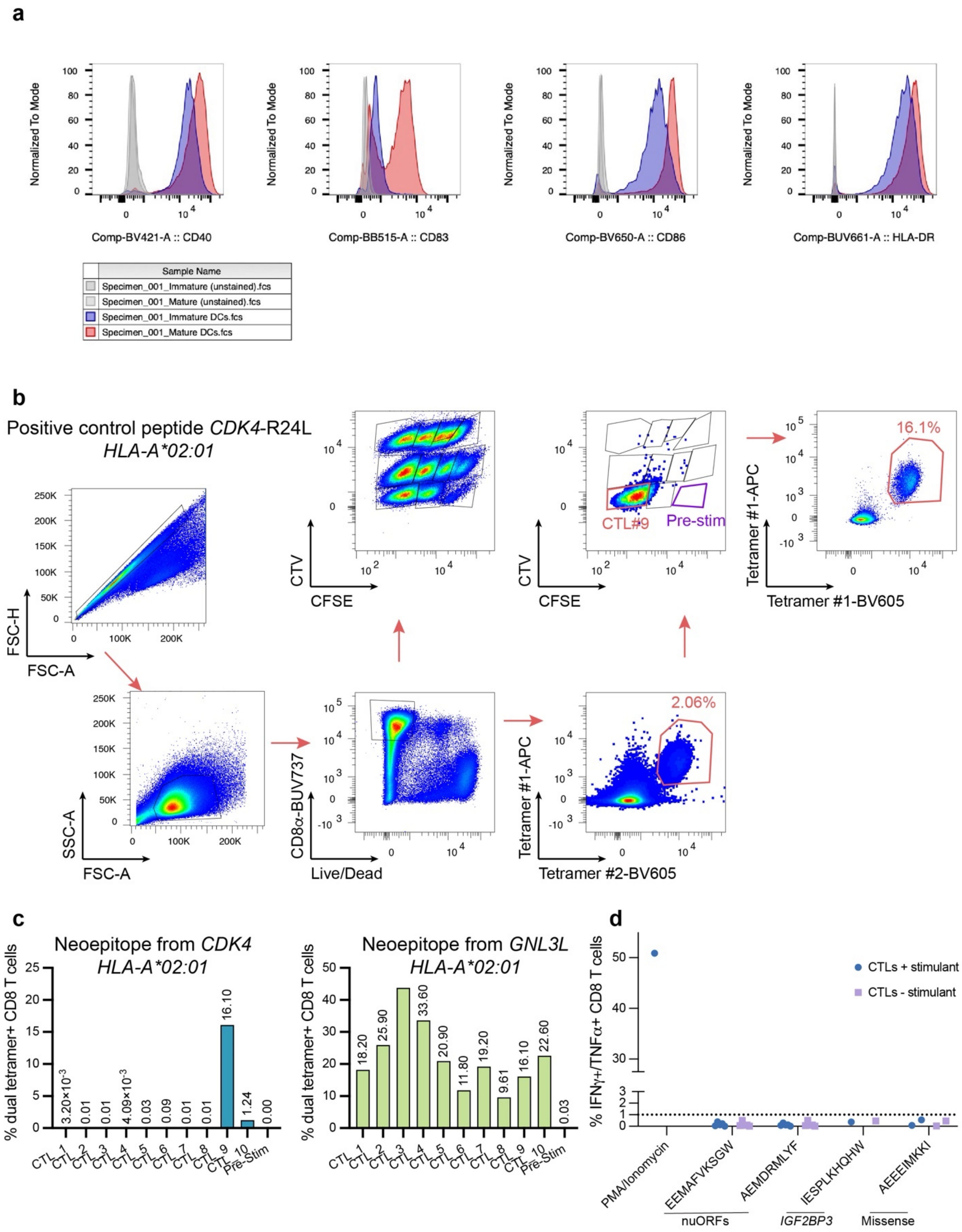
IGF2BP3 IHC in two positive human PDAC
samples (Human Protein Atlas)



IGF2BP3 IHC in two negative human PDAC
samples (Human Protein Atlas)

Supplementary Figure 7: **Immunohistochemistry staining supports the expression of IGF2BP3 in human PDAC.**

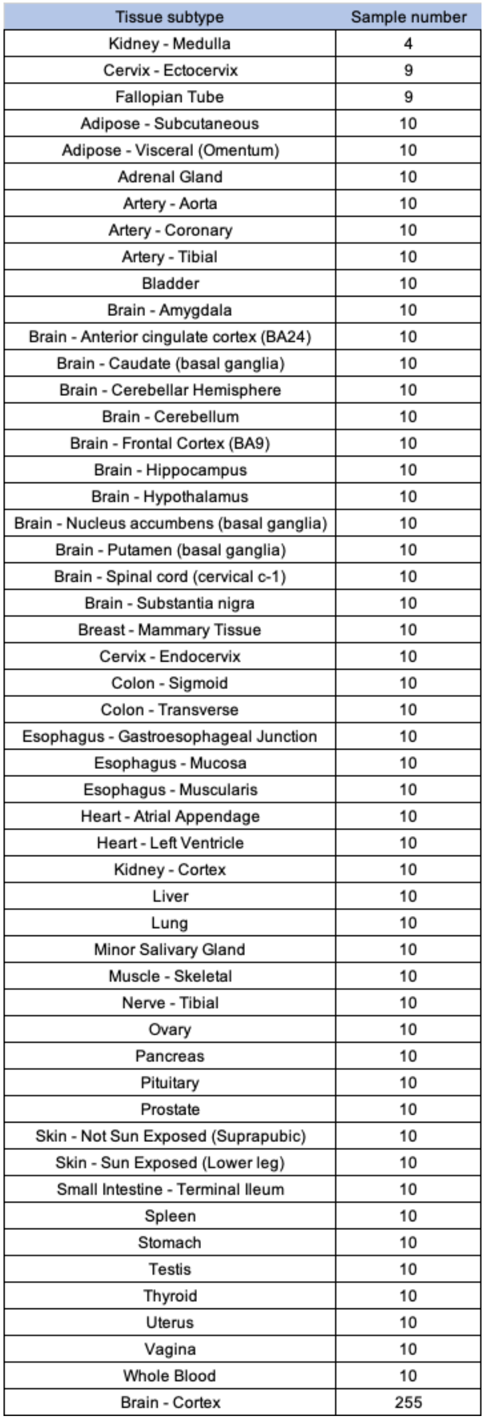
a. Immunohistochemistry (IHC) staining for two human PDAC samples positive for IGF2BP3 protein expression (top) and two that were negative (bottom). These data were retrieved from the Human Protein Atlas [34].



Supplementary Figure 8: **Successful dendritic cell derivation and maturation and expansion of CD8+ T cells specific to *bona fide* melanoma neoepitopes.**

a. Dendritic cells show increased expressed of maturation and activation markers after treatment with cytokines and adjuvants.
b. Successful expansion of T cells recognizing a *CDK4*-derived melanoma neoepitope from a previously undetectable population. Controls cells not previously stimulated with the peptide do not stain positive for peptide loaded dual-tetramers (upper middle plot; purple label).
c. Frequency of CD8+ T cells recognizing a *CDK4*-derived neoepitope (left) or *GNL3L*-derived neoepitope (right) across ten CTL lines derived from the same PBMC donor.
d. The percentage of interferon gamma+ and TNF-alpha+ cells among live $CD8^+$ cells stimulated or unstimulated with peptides derived from nuORFs, *IGF2BP3*, or a missense mutation. Peptide source categories are indicated below the X-axis. CTLs stimulated with the peptide they were expanded with are labeled in blue, and the same CTLs untreated with the peptide are labeled in purple. A pool of the CTL lines stimulated with PMA and ionomycin are included as a positive control. All technical replicates (different CTLs; same donor) (*n*=1=5) are pooled together in each column.

Supplementary Table 1: Number of samples in each tissue subtype included in normal tissue expression analysis.



Tissue subtype	Sample number
Kidney - Medulla	4
Cervix - Ectocervix	9
Fallopian Tube	9
Adipose - Subcutaneous	10
Adipose - Visceral (Omentum)	10
Adrenal Gland	10
Artery - Aorta	10
Artery - Coronary	10
Artery - Tibial	10
Bladder	10
Brain - Amygdala	10
Brain - Anterior cingulate cortex (BA24)	10
Brain - Caudate (basal ganglia)	10
Brain - Cerebellar Hemisphere	10
Brain - Cerebellum	10
Brain - Frontal Cortex (BA9)	10
Brain - Hippocampus	10
Brain - Hypothalamus	10
Brain - Nucleus accumbens (basal ganglia)	10
Brain - Putamen (basal ganglia)	10
Brain - Spinal cord (cervical c-1)	10
Brain - Substantia nigra	10
Breast - Mammary Tissue	10
Cervix - Endocervix	10
Colon - Sigmoid	10
Colon - Transverse	10
Esophagus - Gastroesophageal Junction	10
Esophagus - Mucosa	10
Esophagus - Muscularis	10
Heart - Atrial Appendage	10
Heart - Left Ventricle	10
Kidney - Cortex	10
Liver	10
Lung	10
Minor Salivary Gland	10
Muscle - Skeletal	10
Nerve - Tibial	10
Ovary	10
Pancreas	10
Pituitary	10
Prostate	10
Skin - Not Sun Exposed (Suprapubic)	10
Skin - Sun Exposed (Lower leg)	10
Small Intestine - Terminal Ileum	10
Spleen	10
Stomach	10
Testis	10
Thyroid	10
Uterus	10
Vagina	10
Whole Blood	10
Brain - Cortex	255

Methods

Patient tumor organoid derivation and sequencing.

The following two sections are taken from Raghavan *et al*. (2021), who first described the derivation and sequencing of the organoids described in Chapter 2 [19]. These sections will be modified before a final submission of this manuscript to reflect an upcoming publication describing this resource in more detail.

Human specimens

Eligible participants were recruited from outpatient clinics and inpatient units at the Dana-Farber Cancer Institute and the Brigham and Women's Hospital. Investigators obtained written, informed consent from patients at least 18 years old with pancreatic cancer for Dana-Farber/Harvard Cancer Center Institutional Review Board (IRB)-approved protocols 11-104, 17-000, 03-189, and/or 14-408 for tissue collection, molecular analysis, and organoid generation. Organoid samples for bulk genomic and transcriptomic analyses were collected between May 2015 and January 2018. Core needle biopsy specimens were collected, and the first core was sent for pathologic analysis. One or more additional cores were then allocated for organoid generation.

Bulk RNA- and DNA-sequencing of organoids

RNA was obtained for bulk RNA-sequencing from established organoids using one of two approaches. Dissociated organoids were resuspended into cold Matrigel, added as droplets to tissue culture plates (Greiner BioOne), and allowed to polymerize for 30 min before addition of media. Organoids were grown for 14-21 days (until confluent) under these conditions with regular media changes. At the time of harvest, cells were washed with cold phosphate buffered saline (PBS) at 4C, then lysed with Trizol (Invitrogen) before snap-freezing. To isolate RNA, we performed chloroform extraction with isolation of the aqueous phase before processing RNA as per protocols outlined in the QIAGEN AllPrep DNA/RNA/miRNA Universal kit. In the second approach, used to obtain both RNA and DNA, dissociated organoids were resuspended in a solution of 10% Matrigel in complete organoid media (volume/volume) and cultured in ultralow-attachment culture flasks (Corning). Organoids were grown for 14-21 days (until confluent) before pelleting, washing with cold PBS at 4C until most Matrigel was dissipated, and then snap frozen. Cell pellets were homogenized using buffer RLT Plus (QIAGEN) and a Precellys homogenizer. Samples were then processed for both DNA extraction and RNA isolation as per the QIAGEN AllPrep DNA/RNA/miRNA Universal kit. Purified RNA and DNA were then submitted for sequencing by the Broad Institute Genomics Platform. For bulk RNA-sequencing, total RNA was quantified using the Quant-iT RiboGreen RNA Assay Kit (Thermo Fisher) and normalized to 5 ng/mL. Following plating, 2 mL of a 1:1000 dilution of ERCC RNA controls (Thermo Fisher) were spiked into each sample. An aliquot of 200 ng for each sample was transferred into library preparation which uses an automated variant of the Illumina TruSeq Stranded mRNA Sample Preparation Kit. This method preserves strand orientation of the RNA transcript and uses oligo dT beads to select mRNA from the total RNA sample followed by heat fragmentation and cDNA synthesis from the RNA template. The resultant 400 bp cDNA then goes through dual-indexed library preparation: 'A' base addition, adaptor ligation using P7 adapters, and PCR enrichment using P5 adapters. After enrichment, the libraries were quantified using Quant-iT PicoGreen (1:200 dilution; Thermo Fisher). After normalizing samples to 5 ng/mL, the set was pooled and quantified using the KAPA Library Quantification Kit for Illumina Sequencing Platforms. The entire process was performed in 96-well format and all pipetting was done by either Agilent Bravo or Hamilton Starlet. Pooled libraries were normalized to 2 nM and denatured using 0.1 N NaOH prior to sequencing. Flowcell cluster amplification and sequencing were performed according to the manufacturer's protocols using the NovaSeq 6000. Each run was a 101 bp paired-end with an eight-base index barcode read. Data were analyzed using the Broad Picard pipeline which includes de-multiplexing and data aggregation (https://broadinstitute.github.io/picard/).

For whole genome sequencing, 350 ng of genomic DNA was fragmented using a Covaris focused-ultrasonicator targeting 385bp fragments followed by size selection using SPRI cleanup. Library preparation was performed using a KAPA HyperPrep without amplification kit (KAPA Biosystems) with palindromic forked adapters with unique 8-base index sequences embedded within the adaptor (Roche). Libraries were then quantified using quantitative PCR (kit purchased from KAPA Biosystems) with probes specific to the adaptor ends on an Agilent Bravo liquid handling platform. Libraries were normalized to 2.2 nM, pooled into 24-plexes, combined with NovaSeq Cluster Amp Reagents DPX1, DPX2, and DPX3, and loaded into single lanes of a NovaSeq 6000 S4 flowcell using a Hamilton Starlet Liquid Handling system. Cluster amplification and sequencing occurred utilizing sequencing-by-synthesis kits to produce 151bp paired-end reads. Output from Illumina software was processed by the Broad Picard pipeline (https:// broadinstitute.github.io/picard/) to yield BAM files containing demultiplexed, aggregated aligned reads.

Single nucleotide variant calls from whole-genome sequencing of organoids.

For targeted DNA-sequencing of clinical samples, next-generation sequencing using a custom-designed hybrid capture library preparation was performed on an Illumina HiSeq 2500 with 2x100 paired-end reads. Sequence reads were aligned to reference sequence b37 edition from the Human Genome Reference Consortium using bwa, and further processed using Picard (version 1.90, http://broadinstitute.github.io/picard/) to remove duplicates and Genome Analysis Toolkit (GATK, version 1.6-5-g557da77) to perform localized realignment around indel sites. Single nucleotide variants were called using MuTect v1.1.45 [44].

Organoid propagation and preparation of organoids for immunopeptidomics

Pancreatic organoid isolation and propagation has been previously described [45]. Human patient derived organoids were cultured in three-dimensional domes of Matrigel (80-90% Matrigel) and passaged every 3-4 days. Prior to HLA-I immunopeptidomics, organoids were treated with hIFNg (PeproTech; 100 ng/mL) for 36-40 hours to upregulate surface MHC class I expression [20]. Organoids were mechanically dissociated and then subjected to a gentle enzymatic digestion (TrypLE for less than 10 minutes at 37°C) to digest Matrigel. Cells were washed in 1x PBS and cell pellets were snap frozen and stored at -80°C for batch processing.

Media for pancreatic organoids was formulated based on L-WRN cell conditioned media (L-WRN CM) [46]. Briefly, L-WRN CM was generated by collecting 8 days of supernatant from L-WRN cells, grown in Advanced DMEM/F12 (Gibco) supplemented with 20% fetal bovine serum (Hyclone), 2 mM GlutaMAX, 100 U/mL of penicillin, 100 μg/mL of streptomycin, and 0.25 μg/mL amphotericin. L-WRN CM was diluted 1:1 in Advanced DMEM/F12 (Gibco) and supplemented with additional RSPO-1 conditioned media (10% v/v), generated using Cultrex HA-R-Spondin1-Fc 293T Cells. The following molecules were also added to the growth media: B27 (Gibco), 1 μM N-acetylcysteine (Sigma-Aldrich), 10 μM nicotinamide (Sigma-Aldrich), 50 ng/mL EGF (Novus Biologicals), 500 nM A83-01 (Cayman Chemical), 10 μM SB202190 (Cayman Chemical), and 500 nM PGE2 (Cayman Chemical).

Flow cytometry

Pancreatic organoids

Pancreatic organoids were grown as described above. Where indicated, organoids were treated with interferon-gamma (100 ng/mL; PeproTech) for 36-40 hours prior to analysis. Organoids were dissociated using TrypLE (10 minutes to minimize cleavage of surface proteins) washed with PBS. Single cell suspensions were pelleted at 2000 rpm and transferred to 96-well round-bottom plates for flow cytometric staining. Prior to surface staining, cell pellets were resuspended in Live/Dead dye (Ghost Dye Red 780, Tonbo Biosciences) diluted 1:1000 in PBS on ice for 20 minutes in the dark. Surface staining was performed using HLA-A/B/C (clone W6/32; PE; 1:40) on cells in PBS with 1% heat-inactivated FBS on ice for 30 min in the dark.

Monocyte-derived dendritic cells

MoDCs were prepared as previously described [38]. Briefly, monocytes were isolated from autologous leukapheresis samples using the adherence method. To generate immature dendritic cells, cells were cultured for 16-24 hours in recombinant human GM-CSF (800 U/ml) and recombinant human IL-4 (35 ng/ml). MoDCs were differentiated using a combination of recombinant human TNF-α (10 ng/ml), recombinant human IL-1β (10 ng/ml), recombinant human IL-6 (10 ng/ml), and recombinant human PGE2 (1 μg/ml).

MoDCs were harvested for flow cytometric staining, washed with PBS, and transferred to 96-well U-bottom plates. Prior to surface staining, cell pellets were resuspended in Live/Dead dye (Ghost Dye Red 780, Tonbo Biosciences) diluted 1:1000 in PBS on ice for 20 minutes in the dark. Surface staining was performed on ice for 30 min in the dark using the following antibodies: anti-CD14 (clone M5E2; BUV737), anti-CD11c (clone Bu15; PE), anti-CD83 (clone HB15e; FITC), anti-CD86 (clone 2331; BV650); anti-HLA-DR (clone G46-6; BUV661), anti-CD40 (clone 5C3; BV421). All antibodies were used at recommended (1:20) dilution unless otherwise indicated.

Antigen-specific T cells

To analyze antigen-specificity of expanded cytotoxic T lymphocytes lines (CTLs), we employed both fluorescently conjugated multimer analysis and peptide-mediated stimulation of cytokine production.

Multimer analysis

Briefly, each CTL line was harvested independently, washed in PBS, counted with a manual hemocytometer, and 3×10^5 cells (per multimer) were transferred to 96-well U-bottom plates. Individual CTL lines were fluorescently barcoded using a 3x4 grid of serially-diluted CellTrace CFSE (CFSE) and CellTrace Violet (CTV) (ref-Shumacher 2019). Cells were barcoded on ice for 5 minutes in the dark and excess dye was quenched with HI-FBS. Following fluorescent-barcoding, CTL lines from a given pool (typically 10 CTL lines plus one negative control CTL line) were combined into one reaction and then stained for flow cytometric analysis. Prior to surface staining, cell pellets were resuspended in Live/Dead dye (Ghost Dye Red 780, Tonbo Biosciences) diluted 1:1000 in PBS on ice for 20 minutes in the dark. Fluorescently-conjugated HLA-A*02:01 tetramers (APC and BV605) were loaded with peptide immediately prior to staining and cells were first incubated with tetramers at room temperature for 15 mins, prior to surface staining using anti-CD8 (clone SK1; BUV737) for 30 minutes on ice.

Cytokine analysis

Briefly, each CTL line was harvested independently, washed in PBS, counted with a manual hemocytometer, and 3×10^5 CTLs were plated per well in a 96-well U-bottom plate (typically 5 wells per CTL line, one well for each experimental peptide and one well for a no peptide control). In each well with CTLs, autologous CD8^{neg} PBMCs (1:1 ratio with CTLs), 5 ug/mL peptide, and GolgiPLUG/GolgiSTOP were added. Cells were allowed to incubate at 37^{o}C for 5-6 hours prior to flow cytometric staining. Cells were washed and prior to surface staining, cell pellets were resuspended in Live/Dead dye (Ghost Dye Red 780, Tonbo Biosciences) diluted 1:1000 in PBS on ice for 20 minutes in the dark. Cells then underwent surface staining using anti-CD8 (clone SK1; SK1) at 1:40 dilution for 30 minutes on ice. CTLs were fixed and permeabilized using FoxP3 fix (BD Biosciences) and followed by intracellular cytokine staining with anti-IFNg (clone B27; APC) and anti-TNFa (clone MAb11; PE), both used at 1:80 dilution. For all flow cytometry experiments, samples were acquired on BD LSR II or LSR Fortessa machines.

T cell culture and stimulation

$CD8^+$ T cell isolation, culture, and peptide stimulation was carried out exactly as described in the protocol published by Rollins *et al*. (2020).

Neoepitope Prediction for bulk tumor datasets

In the TCGA cohort, 148 PDAC patients were analyzed (of 150). One patient lacking a normal BAM file was excluded, and another patient was also excluded due to hypermutation [47]. In the DFCI-PancSeq cohort, 57 patients with (1) annotated mutations, (2) both WES and RNA-Seq data, and (3) sufficient tumor purity [48] were analyzed. Binary Alignment Map (BAM) files were obtained for PancSeq (aligned to GRCh37) and for TCGA (aligned to GRCh38). Thus, GRCh37 was used as the reference genome for the PancSeq cohort in all downstream analyses, and GRCh38 was used for the TCGA cohort.

HLA typing was performed using two programs and with both RNA-Seq and WES data to assess robustness of HLA allele calls. HLA alleles for classical genes (HLA-A, -B, and C) were called using the HLA genotyping algorithm, OptiType, version 1.3.1 [49], as well as seq2HLA, version 2.3 [50], which was also used to identify alleles for HLA-E. Tumor and normal WES BAM files were used to create inputs to OptiType, which outperforms peer programs in WES-based HLA-typing [51], and RNA-Sequencing BAMs were used to create inputs to seq2HLA. WES BAMs were filtered to retain only reads mapping to the HLA region (6:28477897-33448354 in GRCh37; chr6:28510120-33480577 in GRCh38) with the genomics software suite, Samtools, version 1.10 [52]. The BAMs were then converted to FASTQ format, and then filtered with the genome mapping tool, RazerS 3, version 3.5.8 [53], as recommended in the OptiType documentation. RNA-Seq BAMs were sorted, converted to FASTQ format, and compressed before being used as inputs to seq2HLA. Both programs were run with default parameters.

A custom python script was then employed to evaluate concordance between (1) normal and tumor HLA allele calls from WES and (2) seq2HLA and OptiType calls. Only 4/342 alleles (0.58%) in the PancSeq cohort and only 2/888 (0.23%) alleles in the TCGA cohort were called differently between tumor and normal WES-based calls. Given the consistency of OptiType calls for tumor and normal WES data, the OptiType allele was accepted as the final call to resolve discrepancies between OptiType and Seq2HLA.

Mutation Annotation Format (MAF) files were obtained for patients in both datasets and converted to Variant Call Format (VCF) files. VCF files were filtered to only retain single nucleotide variants (SNVs). Only PASS variants were available in the PancSeq MAF file and were thus not filtered further [48]. Mutations in the TCGA cohort included non-PASS variants, which were all filtered in this cohort with the exception of some non-PASS mutations in known PDAC-associated genes that had been marked as either panel_of_normals, clustered_events, or homologous_mapping_event in the TCGA MAF file. For these cases, genes that had variants marked as non-PASS more than twice by at least one of these filters were reconsidered. The following genes with a known association with PDAC based on a literature search were retained: *KRAS*, *TP53*, *GNAS*, *RNF43*, *PLEC*, *FLG*, *AHNAK*, *APOB*, *CSMD1*, *PLXNA1*, *MCM6, MKI67*, and *SIPA1*. This step was intended to reduce false negatives, and in the case of *KRAS*, this step retrieved 30 variants at residue position 12, a site known to confer oncogenic properties when mutated.

Indel variants were called using the variant callers, Strelka2, version 2.9.2 [54], and Scalpel, version 0.5.4 [55]. The structural variant and indel caller, Manta, version 1.6.0 [56], was run prior to Strelka2 and these results were incorporated into the indelCandidates parameter for Strelka2. Scalpel was run with default parameters, with a bed file derived from the CGHub bitbucket account (https://cghub.ucsc.edu; whole_exome_agilent_1.1_refseq_plus_3_boosters.targetIntervals.bed). For the PancSeq cohort, the unmodified first 3 columns of this file were used. For the TCGA cohort, the coordinates in this file were converted to GRCh38 coordinates using the LiftOver tool from the UCSC genome browser [57]. Scalpel failed to call variants for 13/148 TCGA patients due to excessive read buildup at some loci. To enable variant calling with Scalpel for these patients, the Picard tools' DownsampleSam function (http://broadinstitute.github.io/picard/) was employed to randomly downsample reads in the tumor BAM files of these patients by decrements of 10%, starting at 50%. This was done until Scalpel successfully called variants for each patient. Ten patients succeeded at 50%, two at 40%, and one at 30%.

To reduce the contribution of caller-specific biases and hence the indel false positive rate, only those indels that were called and marked as PASS by both Scalpel and Strelka2 were retained. Variant call format (VCF) files containing the union of PASSed variants from Strelka2 and Scalpel were generated with a custom batch script, and variant allele frequencies were calculated using statistics output by Strelka2. These indel VCF files were then merged with the corresponding SNV VCF files for each patient using the vcf-shuffle-cols and vcf-concat functions from VCFtools, version 0.1.13 [58].

Variant consequence was then annotated using the Ensembl Variant Effect Predictor (VEP), version 99 [59]. The corresponding VEP cache for both GRCh37 and GRCh38 was downloaded and used to run the software offline. VEP was run using the Wildtype and Downstream plugins to annotate the effects of indels. The following parameters were employed: --symbol, --terms=SO, --cache, --offline, --transcript_version, --pick. The --pick parameter was reordered from the default to report the transcript with the most extreme consequence for each variant: rank, canonical, appris, tsl, biotype, ccds, length, mane.

Neoepitopes were predicted with the HLA allele calls and variant effect predictions using the antigen prediction toolkit, pVACtools, version 1.5.7 [60]. For each mutation, mutant peptides were generated for lengths of 8-, 9-, 10-, and 11- amino acids, the spectrum of peptide lengths known to bind to MHC class I. MHC:peptide binding affinity was predicted for all peptide:MHC allele pairs with NetMHC-4.0, NetMHCpan-4.0, SMM (version 1.0), and SMMPMBEC (version 1.0), and the median value across all affinity predictions was taken as a final, composite measure of binding affinity.

After predictions were made by pVACtools, candidate neoepitopes from all patients were merged into a single matrix and filtered using a custom python script and the following criteria (based on parameters output by pVACtools): median peptide:MHC binding affinity < 500 nM, tumor DNA depth >= 5, tumor DNA variant allele frequency >= 0.07, cysteine_count <=1, and a median wildtype:mutant peptide binding affinity fold-change >= 1.

Neoepitope predictions for patient-derived pancreatic organoids

Somatic mutation calls and neoepitope predictions for organoids were made as described for bulk organoids, with the following exception. All tools were run with parameters prescribed for whole genome sequencing datasets. In the case of Scalpel, indel calls were restricted to the exome intervals using the bed file described above. For both neoepitope and somatic mutational burden comparisons between organoids and bulk tumors, only the variants overlapping these intervals in bulk samples and organoid samples were considered.

To analyze the impact of variant allele frequency (VAF) on neoepitope prediction, neoepitopes were predicted with all of the same parameters described above for both matched organoid and bulk tumor samples, with only VAF being altered. A VAF of 0.1 was used for the matched comparison bar plot of total predicted neoepitope load.

Quantification of tumor mutational burden

We defined tumor mutational burden as the number of all somatic mutations per megabase across coding regions of the genome. Coding regions were classified as the set of genomic regions in final_whole_exome_agilent_1.1_refseq_plus_3_boosters.targetIntervals.bed (totaling 32,950,014 bases or 32.950014 Mb). To make a fair comparison between patient-derived organoids and bulk tumor samples, we considered only the subset of mutations that overlapped the Agilent file for calculating TMB. TMB was then calculated as the number of variants in the VCF (grep -v "#" | wc -l) divided by 32.950014 Mb.

Construction of patient-specific proteome database for mass spectrometry analysis

Germline mutations and variant phasing

Germline SNVs were called using Strelka with default parameters and the reference genome file from human assembly 19. Only variants marked PASS were considered for downstream analysis. These were merged with somatic variants into single VCF files using the Genome Analysis

Toolkit (GATK), version 4.1.2.0 [61]. VCF files were then coordinate sorted, followed by read-backed variant phasing performed using the ReadBackedPhasing function from GATK, version 3.7. Final variant-recoded protein sequences were then retrieved using pVACtools' Generate Protein Fasta function, with flanking sequence length set arbitrarily to 30,000 and with phasing enabled. Phased and unphased protein sequences were then added to patient-specific protein search databases.

nuORF mutations

Coordinates corresponding to nuORFs were retrieved from the nuORFdb_v1.0.bed file published by Ouspenskaia *et al*. (2022), downloaded from the NCBI Gene Expression Omnibus (GSE143263). We then took somatic SNVs and indels intersecting these coordinates and generated mutation-encoded nuORF protein sequences using a custom python script modified from one published by Ouspenskaia *et al*. (2022). Full-length mutant protein sequences were then aggregated into a single FASTA file for each patient.

Retained intron analysis

Raw RNA-seq FASTQ files were pseudoaligned to a transcriptome augmented to contain both exonic and intronic sequences, with intronic sequences set to extend 25 base pairs into adjacent exons, as described by Smart *et al*. (2018). TPM values were calculated using Kallisto, version 0.45.0 [62].

Retained introns were called using the KeepMeAround algorithm, as implemented by Smart *et al*. (2018). Retained introns were called for all PDAC organoid samples and 239 GTEx samples, representing five randomly selected samples per each tissue subtype from the cohort of GTEx samples used by Ouspenskaia *et al*. (2022). Introns were considered retained if they passed the following thresholds: intron retention ratio > 0.05, intron TPM > 1, exonic TPM > 1, and ≥ 5 unique reads map to the intron. To define sets of tissue-specific retained introns, we required each intron to pass thresholds in all samples (*n* = 4-5) of a given tissue subtype. After taking the intersection of retained introns within each tissue, we then established the union of all of these sets as the full cohort of introns retained in normal tissues. We excluded reproductive tissues from consideration: testis, ovary, fallopian tube, and uterus. PDAC organoid-specific retained introns were then defined as those that pass the aforementioned thresholds in at least one organoid sample but were not contained in the set of normal tissue retained introns.

For the seven organoid samples profiled with immunopeptidomics, we modified python scripts published by Smart *et al*. (2018) to retrieve the open reading frames and coordinates associated with sample-specific retained introns. We then used bedtools, version 2.29.1, to retrieve all germline and somatic SNVs intersecting retained intron coordinates from VCF files established in the mutation calling procedures described in the previous section [63]. We then employed BCFtools in consensus mode to recode intronic sequences with all variants for both haplotypes separately and combined (to account for potential errors in variant phasing) [64]. The resulting sequences were then all translated across all three reading frames, accounting for potential indels, to generate the retained intron-encoded protein sequence search space.

HLA peptide immunoprecipitation and peptide sequencing by tandem mass spectrometry

MHC-I peptide identification with Spectrum Mill

Final, patient-specific protein sequences were amalgamated from retained introns, somatic and germline protein-coding variants, and nuORF-altering variants into a single FASTA file. MS/MS spectra for each organoid sample were searched against the patient-specific set of protein sequences appended to nuORFdb v.1.0 and a base reference proteome containing all UCSC Genome Browser genes with hg19 annotation of the genome and its nonredundant protein-coding transcripts (52,788 entries) as well as 264 common laboratory contaminants, including proteins present in cell culture media and immunoprecipitation reagents [12].

Soluble lysates from up to 0.1 to 0.2 g organoid cells were immunoprecipitated with W6/32 antibody (sc-32235, Santa Cruz). 10 mM iodoacetamide was added to the lysis buffer to alkylate cysteines during the lysis and incubation step (3 h, 4 °C). Peptides acid eluted either on StageTips or SepPak cartridges and analyzed in technical duplicates using high-resolution LC–MS/MS on a QExactive Plus, QExactive HF or Fusion Lumos mass spectrometer (Thermo Scientific).

Target-decoy FDR estimation was enabled by Spectrum Mill with on-the-fly generation of decoy sequences during searches. For each candidate sequence passing the precursor mass tolerance filter, the internal sequence was reversed, while holding fixed the second position and the peptide C terminus, to maintain not only equal size target and decoy search spaces, but also comparable HLA class I binding motifs among the sequence candidate population.

PSMs were consolidated to the peptide level to generate lists of confidently observed peptides for each allele using the Spectrum Mill protein/peptide summary module’s peptide-distinct mode with filtering distinct peptides set to case sensitive. A distinct peptide was the single highest scoring PSM of a peptide detected for each allele. MS/MS spectra for a particular peptide may have been recorded multiple times (for example, as different precursor charge states, from replicate immunopeptidomes, from replicate LC–MS/MS injections). Different modification states observed for a peptide were each reported when containing amino acids configured to allow variable modification; a lowercase letter indicates the variable modification (C-cysteinylated, c-carbamidomethylated) [12].

Bulk RNA-Sequencing analysis of nuORF transcripts and canonical genes

All transcript-specific TPM values for 777 GTEx normal tissue samples were retrieved from the Gene Expression Omnibus (GSE143263). To compare expression of these transcripts in PDAC bulk tumors and organoid samples, we employed identical tools and parameters described by Ouspenskaia *et al*. (2022) and detailed at https://github.com/broadinstitute/gtex-pipeline/blob/master/TOPMed_RNAseq_pipeline.md. Briefly, BAM files were sorted by query name using the collate function from Samtools, version 1.10. and then converted to FASTQ files. Sequencing reads were then aligned to reference human genome GRCh37 using STAR, version 2.6.1 [65], coupled with a transcriptome annotation comprised of GENCODE v26lift37, transcripts annotated as tstatus 'unannotated' from MiTranscriptome annotation [66], and a cohort of all lncRNAs reported by Cabili and colleagues [67]. Isoform and gene expression was then quantified with RSEM, version 1.3.1 [68]. Isoform TPM values were compared to GTEx using the dataset retrieved from GEO, and gene-level TPMs were compared to GTEx using TPM values for all samples from a matrix downloaded through the GTEx expression website (https://storage.googleapis.com/gtex_analysis_v8/rna_seq_data/GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_tpm.gct.gz).

Designation of PDAC-associated and PDAC-specific nuORF-associated transcripts and canonical genes

nuORF-associated transcripts and canonical genes were considered for analysis if they had at least one peptide sequence unambiguously mapped to them from the immunopeptidomics dataset (i.e., the corresponding peptide sequence is uniquely encoded by the associated transcript). Transcripts and genes were designated PDAC-specific if their 90th percentile TPM values was less than one across all 26 normal tissue categories, which excluded testis, fallopian tube, uterus, and ovary. Tissue subtypes were aggregated into a single category (n = 26) before calculating the

90th percentile. PDAC-associated transcripts and genes were designated according to the cutoffs described in Supplementary Figure 7 and by Bradley *et al*. (2020). All candidates were required to be expressed at a TPM > 1 in at least one of the seven organoid samples profiled with immunopeptidomics.

From these transcripts, we further filtered PDAC-specific and PDAC-associated nuORF candidates and canonical genes according to their detection in healthy tissue immunopeptidomes. For this analysis, we re-analyzed MHC-I peptide spectra derived from 31 different normal tissues previously reported [32], searching against both the canonical reference proteome and nuORFdb as described above for organoid samples. *CT83* had a 7mer detected in a single bladder entry, so this gene was retained. *IGF2BP3* was detected in three thymus samples and one cerebellum entry, though this latter entry was also mappable to a nuORF sequence. *TDRD15* was detected in the kidney, liver, and lung. *C4orf36*, *MAGEA11*, *ZNF695*, and *CALHM3* were not detected in any normal tissue immunopeptidome.

Quantification of stromal gene signature programs in bulk tumors and organoids

Expected counts calculated by RSEM were extracted from gene-level output files; only features annotated with "ENSG" Ensembl codes were retained for downstream analysis of stromal signatures. We further filtered genes down to a list of 17,165 canonical protein-coding genes by intersecting this list with all Ensembl gene IDs designated as protein-coding and "canonical" in the nuORFdb annotation file. Raw counts for 49 bulk tumor samples with more than 50 million RNA-Seq reads and 48 pancreatic cancer organoids were then aggregated within a single matrix for downstream analysis. We then took a subset of only the genes expressed in bulk tumor samples, defined as genes with a mean log2(count +1) > 2 across bulk tumor samples.

Raw counts were normalized using the estimateSizeFactors function from DESeq2 and then log2-transformed [69]. The resulting transformed expression matrix was used an input for the Estimation of STromal and Immune cells in MAlignant Tumours using Expression data (ESTIMATE) package, using default parameters outlined in the package vignette. The computed stromal, immune, and ESTIMATE scores were then extracted to compare differences between bulk and organoid samples. Different normalization methods were tested, including the variance stabilizing transformation from DESeq2, and expression data for the two cohorts was also normalized separately as well as together. All of these methods robustly and uniformly indicated higher stromal and immune signatures in bulk tumor samples compared to organoids. We proceeded with the results obtained from log2-transformed expression data (normalized for both cohorts together) for computing the results described in the Main Text. For this analysis and the isoform-based RNA-Seq analyses described above, we excluded bulk tumor samples with a library size less than 50 million reads, including two samples with matched bulk tumor datasets (PANFR0151 and PANFR0181). To compare bulk tumor ESTIMATE scores against tumor purity, we used tumor purity estimates based on the ABSOLUTE algorithm [70].

Human Protein Atlas

Immunohistochemistry staining of IGF2BP3 depicted in Supplementary Figure 8 for human PDAC samples was retrieved from the Human Protein Atlas [34].

Single cell RNA-Sequencing and Gene module analysis

Human PDAC scRNA-Seq data [26] was downloaded from the Genome Sequencing Archive (accession: CRA001160). A count matrix of 41,986 pre-processed cells was prepared from these data and used to create a Seurat object. Genes expressed in less than 10 cells were filtered out. Data normalization and scaling, variable feature selection, cell clustering, and differential gene expression analysis was performed using Seurat. Data were normalized by total expression per cell and scaled using a factor of 10,000 and log transformed (natural scale). The top 2,000 variable genes were selected using Seurat's default "vst" method. The expression of these genes was then scaled and centered, and these genes were then used for all downstream analysis. Principal component analysis (PCA) was then performed for dimensionality reduction. The first 15 principal components were used for the construction of the k-nearest neighbor graph and the UMAP plot (metric parameter = Euclidean). Clusters were then assigned using the Louvain method with a resolution of 1. Clusters were re-annotated by cell type according to gene expression patterns as detailed by Peng *et al*. 2019. UMAP feature plots were generated using order=TRUE and min.cutoff = "q01", max.cutoff = "q99" as parameters.

Seurat's AddModuleScore function (control parameter = 8) was used to calculate gene module scores for all cells. Peptide signature gene set scores depicted in Figure 1 were derived from immunopeptidomics data for each patient sample by taking all of the peptides that mapped to a single canonical Ensembl protein ID and then matching these with the corresponding gene symbol. Stromal and immune gene modules depicted in Supplementary Figure 2 were derived from all of the genes excluded as unexpressed in PDAC organoids as described in ESTIMATE analysis.

After cluster annotation and gene module scoring, the Seurat object metadata table was exported for analysis using the XL-mHG package in Python [71]. This package and the XL-mHG hypergeometric test was used to determine whether gene modules were significantly enriched within each cell type.

Statistical methods

Except for XL-mHG, all statistical tests mentioned in this manuscript were conducted in PRISM. For each dataset, we evaluated whether they followed a Gaussian distribution using the Anderson-Darling test. Parametric or nonparametric tests were then selected based on this result. Results were considered significant if the computed tested statistic was associated with a *P*-value < 0.05.

References

1. Hakim, N., Patel, R., Devoe, C. & Saif, M. W. Why HALO 301 Failed and Implications for Treatment of Pancreatic Cancer. *Pancreas Fairfax Va* **3**, e1–e4 (2019).
2. Macherla, S. *et al.* Emerging Role of Immune Checkpoint Blockade in Pancreatic Cancer. *Int. J. Mol. Sci.* **19**, 3505 (2018).
3. Gubin, M. M. *et al.* Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577–581 (2014).
4. Bailey, P. *et al.* Exploiting the neoantigen landscape for immunotherapy of pancreatic ductal adenocarcinoma. *Sci. Rep.* **6**, 35848 (2016).
5. Balachandran, V. P. *et al.* Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* **551**, 512–516 (2017).
6. Balachandran, V. P. *et al.* Phase I trial of adjuvant autogene cevumeran, an individualized mRNA neoantigen vaccine, for pancreatic ductal adenocarcinoma. *J. Clin. Oncol.* **40**, 2516–2516 (2022).
7. Frankiw, L., Baltimore, D. & Li, G. Alternative mRNA splicing in cancer immunotherapy. *Nat. Rev. Immunol.* **19**, 675–687 (2019).
8. Smart, A. C. *et al.* Intron retention is a source of neoepitopes in cancer. *Nat. Biotechnol.* **36**, 1056–1058 (2018).
9. Ehx, G. *et al.* Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes. *Immunity* **54**, 737-752.e10 (2021).
10. Intron-Retention Neoantigen Load Predicts Favorable Prognosis in Pancreatic Cancer | JCO Clinical Cancer Informatics. https://ascopubs.org/doi/full/10.1200/CCI.21.00124.
11. Aspden, J. L. *et al.* Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. *eLife* **3**, e03528 (2014).
12. Ouspenskaia, T. *et al.* Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat. Biotechnol.* **40**, 209–217 (2022).
13. Bradley, S. D. *et al.* Vestigial-like 1 is a shared targetable cancer-placenta antigen expressed by pancreatic and basal-like breast cancers. *Nat. Commun.* **11**, 5332 (2020).
14. Raphael, B. J. *et al.* Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **32**, 185-203.e13 (2017).
15. Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **47**, 1168–1178 (2015).

16. Cao, L. *et al.* Proteogenomic characterization of pancreatic ductal adenocarcinoma. *Cell* **184**, 5031-5052.e26 (2021).
17. Stenzinger, A. *et al.* Tumor mutational burden standardization initiatives: Recommendations for consistent tumor mutational burden assessment in clinical samples to guide immunotherapy treatment decisions. *Genes. Chromosomes Cancer* **58**, 578–588 (2019).
18. Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501 (2015).
19. Raghavan, S. *et al.* Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. *Cell* **184**, 6119-6137.e26 (2021).
20. Sarkizova, S. *et al.* A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat. Biotechnol.* **38**, 199–209 (2020).
21. Campbell, J. D. *et al.* Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* **48**, 607–616 (2016).
22. Parkhurst, M. R. *et al.* Unique Neoantigens Arise from Somatic Mutations in Patients with Gastrointestinal Cancers. *Cancer Discov.* **9**, 1022–1035 (2019).
23. Peng, J. *et al.* Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* **29**, 725–738 (2019).
24. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
25. Yamamoto, K. *et al.* Autophagy promotes immune evasion of pancreatic cancer by degrading MHC-I. *Nature* **581**, 100–105 (2020).
26. Peng, J. *et al.* Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.* **29**, 725–738 (2019).
27. Mani, D. R. *et al.* Cancer proteogenomics: current impact and future prospects. *Nat. Rev. Cancer* **22**, 298–313 (2022).
28. Bassani-Sternberg, M. *et al.* Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat. Commun.* **7**, 13404 (2016).
29. Westphalen, C. B. & Olive, K. P. Genetically Engineered Mouse Models of Pancreatic Cancer. *Cancer J. Sudbury Mass* **18**, 502–510 (2012).
30. Turajlic, S. *et al.* Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol.* **18**, 1009–1021 (2017).
31. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
32. Marcu, A. *et al.* HLA Ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. *J. Immunother. Cancer* **9**, e002071 (2021).
33. Taniuchi, K., Furihata, M., Hanazaki, K., Saito, M. & Saibara, T. IGF2BP3-mediated translation in cell protrusions promotes cell invasiveness and metastasis of pancreatic cancer. *Oncotarget* **5**, 6832–6845 (2014).
34. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010).
35. Braunschweig, U. *et al.* Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* **24**, 1774–1786 (2014).
36. Hegde, S. *et al.* Dendritic Cell Paucity Leads to Dysfunctional Immune Surveillance in Pancreatic Cancer. *Cancer Cell* **37**, 289-307.e9 (2020).

37. Strønen, E. *et al.* Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science* **352**, 1337–1341 (2016).
38. Rollins, M. R., Spartz, E. J. & Stromnes, I. M. T Cell Receptor Engineered Lymphocytes for Cancer Therapy. *Curr. Protoc. Immunol.* **129**, e97 (2020).
39. Brahmer, J. R. *et al.* Safety and Activity of Anti–PD-L1 Antibody in Patients with Advanced Cancer. *N. Engl. J. Med.* **366**, 2455–2465 (2012).
40. Royal, R. E. *et al.* Phase 2 Trial of Single Agent Ipilimumab (Anti-CTLA-4) for Locally Advanced or Metastatic Pancreatic Adenocarcinoma. *J. Immunother. Hagerstown Md 1997* **33**, 828–833 (2010).
41. Lu, S. X. *et al.* Pharmacologic modulation of RNA splicing enhances anti-tumor immunity. *Cell* **184**, 4032-4047.e31 (2021).
42. Roudko, V. *et al.* Shared Immunogenic Poly-Epitope Frameshift Mutations in Microsatellite Unstable Tumors. *Cell* **183**, 1634-1649.e17 (2020).
43. Dvinge, H. & Bradley, R. K. Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.* **7**, 45 (2015).
44. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. 861054 Preprint at https://doi.org/10.1101/861054 (2019).
45. Boj, S. F. *et al.* Organoid models of human and mouse ductal pancreatic cancer. *Cell* **160**, 324–338 (2015).
46. VanDussen, K. L., Sonnek, N. M. & Stappenbeck, T. S. L-WRN conditioned medium for gastrointestinal epithelial stem cell culture shows replicable batch-to-batch activity levels across multiple research teams. *Stem Cell Res.* **37**, 101430 (2019).
47. Cancer Genome Atlas Research Network. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **32**, 185-203.e13 (2017).
48. Aguirre, A. J. *et al.* Real-time Genomic Characterization of Advanced Pancreatic Cancer to Enable Precision Medicine. *Cancer Discov.* **8**, 1096–1111 (2018).
49. Szolek, A. *et al.* OptiType: Precision HLA typing from next-generation sequencing data. *Bioinformatics* **30**, 3310–3316 (2014).
50. Boegel, S. *et al.* HLA typing from RNA-Seq sequence reads. *Genome Med.* **4**, 102 (2012).
51. Bauer, D. C., Zadoorian, A., Wilson, L. O. W. & Thorne, N. P. Evaluation of computational programs to predict HLA genotypes from genomic sequencing data. *Brief. Bioinform.* **19**, 179–187 (2018).
52. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
53. Weese, D., Holtgrewe, M. & Reinert, K. RazerS 3: Faster, fully sensitive read mapping. *Bioinformatics* **28**, 2592–2599 (2012).
54. Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* **15**, 591–594 (2018).
55. Narzisi, G. *et al.* Accurate de novo and transmitted indel detection in exome-capture data using microassembly. *Nat. Methods* **11**, 1033–1036 (2014).
56. Chen, X. *et al.* Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
57. Haeussler, M. *et al.* The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* **47**, D853–D858 (2019).
58. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

59. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
60. Hundal, J. *et al.* PVACtools: A computational toolkit to identify and visualize cancer neoantigens. *Cancer Immunol. Res.* **8**, 409–420 (2020).
61. Auwera, G. A. V. der & O'Connor, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. (O'Reilly Media, Inc., 2020).
62. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
63. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
64. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
65. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).
66. Iyer, M. K. *et al.* The Landscape of Long Noncoding RNAs in the Human Transcriptome. *Nat. Genet.* **47**, 199–208 (2015).
67. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
68. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
69. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
70. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
71. Wagner, F. The XL-mHG Test For Enrichment: A Technical Report. (2015) doi:10.48550/arXiv.1507.07905.

CHAPTER 3: A PRIME EDITOR MOUSE TO MODEL A BROAD SPECTRUM OF SOMATIC MUTATIONS *IN VIVO*

Authors: Zackery A. Ely[1,2], Nicolas Mathey-Andrews[1,2,3], Santiago Naranjo[1,2], Samuel I. Gould[1,2], Christina M. Cabana[1,2], Kim L. Mercer[1], William M. Rideout III[1], Gregory A. Newby[4,5,6], Grissel Cervantes Jaramillo[1,7], Katie Holland[1,8], Peyton B. Randolph[4,5,6], William A. Freed-Pastor[1,9], Jessie R. Davis[4,5,6], Peter M.K. Westcott[1], Andrew V. Anzalone[4,5,6], Nimisha B. Pattada[1], Francisco J. Sánchez-Rivera[1,2], David R. Liu[4,5,6], Tyler Jacks[1,2]

This chapter is a modified version of a manuscript currently in review for publication.

Affiliations:
[1] David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3] Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA
[4] Merkin Institute of Transformative Technologies in Healthcare, Broad Institute of Harvard and MIT, Cambridge, MA, USA
[5] Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA
[6] Howard Hughes Medical Institute, Harvard University, Cambridge, MA, USA
[7] Harvard-MIT Division of Health Sciences and Technology, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[8] Department of Biology, Angelo State University, San Angelo, TX, USA
[9] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

Contributions: Z.A.E. and T.J. conceived of the study. Z.A.E., N.M.A, S.N., and T.J., designed the experiments. S.N. and Z.A.E. designed and constructed all DNA vectors described, including the transgene cassette. Z.A.E. and N.M.A. designed and performed all types of experiments reported in the study. Z.A.E., N.M.A., P.R.B., J.R.D., K.H., and G.A.N. collectively designed all pegRNAs. S.I.G. and F.J.S.R. designed the computational pipeline to analyze patient mutation data, and S.I.G. conducted related bioinformatics analyses. G.C.J. conducted pancreatic orthotopic transplant experiments. C.M.C. and S.N. conducted experiments involving lung organoids. N.M.A., Z.A.E., N.B.P., and W.A.F.-P. conducted *in vitro* experiments using pancreatic organoids. W.M.R. conducted mESC targeting and chimera generation. K.L.M. provided animal husbandry expertise and conceptual advice. S.N., G.A.N., P.R.B., A.V.A., W.A.F.-P., P.M.K.W., K.L.M., F.J.S.R., and D.R.L. provided conceptual advice. Z.A.E., N.M.A., S.I.G., F.J.S.R., and T.J. wrote the manuscript with input from all authors.

Abstract

Genetically engineered mouse models (GEMMs) are used extensively to study the effects of a range of cancer-associated mutations *in vivo*, but current models only capture a small fraction of genetic lesions observed in human cancer. GEMMs based on CRISPR-Cas9 have enabled rapid assessment of additional mutations, primarily loss-of-function alleles that arise from insertions or deletions. However, the utility of these models for engineering precise mutations is limited due to their reliance on error-prone DNA repair mechanisms. Here, we describe the development of a facile system for performing *in vivo* prime editing in murine tissues by encoding a Cre-inducible prime editor enzyme in the mouse germline. We show that this model allows rapid and precise engineering of a wide range of mutations in cell lines and organoids derived from multiple primary tissues of these mice, including clinically relevant *Kras* mutations associated with resistance to targeted therapies, as well as *Trp53* hotspot mutations commonly observed in pancreatic cancer patients. Lastly, we illustrate the utility of these mice for modeling lung and pancreatic cancer *in vivo* through lentiviral delivery of prime editing guide RNAs or syngeneic orthotopic transplantation of prime edited organoids. We anticipate that prime editing GEMMs will accelerate preclinical functional studies of cancer-associated alleles and complex genetic combinations that are challenging to model by traditional approaches.

Introduction

Cancer is driven by somatic mutations that accumulate throughout progression and display extensive intertumoral heterogeneity, occurring in thousands of different combinations across human cancer[1,2]. The precise nature of driver mutations can profoundly influence how cancers initiate, progress, and respond to therapy, thereby establishing tumor genotype as a critical determinant of disease outcome[3,4]. Emerging precision oncology treatment paradigms aim to match specific therapies with tumor genotypes, and this strategy has shown promise for a handful of genetic lesions[5,6]. To expand the promise of precision oncology to more patients, it is critical for the field to develop tools to systematically interrogate the effects of distinct genetic lesions and combinations thereof on the overall tumor phenotype, particularly *in vivo*.

GEMMs have been proven invaluable for elucidating the mechanisms by which cancer drivers promote tumor development and progression *in vivo*[7,8]. However, generating new GEMMs using traditional approaches is an expensive, laborious, and time-consuming process. Furthermore, established GEMMs can take months for investigators to acquire and often require laborious breeding programs to combine multiple alleles of interest and to establish a colony of sufficient size for experimental cohorts. These factors impede studies aimed at developing precision oncology treatments for tumors driven by specific genetic variants, which continue to be identified on a regular basis[9]. Moreover, they preclude convenient combinations of driver mutations that frequently co-occur in human cancer genomes.

Genome editing technologies like CRISPR-Cas9 can be used to rapidly engineer somatic mutations when delivered exogenously or when installed as germline alleles[10–14]. While these models have accelerated our ability to study putative cancer driver genes, they are most frequently used to induce DNA double-stranded breaks (DSBs), leading to inactivation of tumor suppressor genes via error-prone repair and frameshifting insertion/deletion (indel) formation. Although CRISPR-based homology-directed repair (HDR) has been used to model precise single nucleotide variants (SNVs) in Cas9-knockin mice, this method requires an exogenous DNA donor template and is limited by low efficiency and high rates of indel byproducts[15]. Furthermore, the requirement for DSBs to induce frameshifts or HDR-based precise edits can lead to confounding genotoxic effects, including on-target chromothripsis events[16].

Precision genome editing technologies like base editing[17] can be used to model cancer in mice by installing specific transition mutations with high efficiency and negligible indel byproducts[11]. Though precise and highly efficient, base editors also have limitations, including the requirement for different base editor enzymes depending on the type of mutation being studied (e.g., cytosine or adenine base editors), and their propensity for bystander editing, which can prohibit introducing the desired amino acid substitutions. While the recent development of C:G transversion base editors will expand the scope of cancer modeling[18,19], current base editing technology is not amenable to modeling the full spectrum of small somatic mutations.

In contrast to base editing and CRISPR-Cas9, prime editing enables engineering the full spectrum of single nucleotide substitutions and small indels with high product purity[20,21]. Prime editors consist of a Cas9-reverse transcriptase fusion that complexes with prime editing guide RNAs (pegRNAs) encoding mutations of interest within a reverse transcriptase template[20,21]. Given that most recurrent cancer driver mutations are not captured by current mouse models and are not readily amenable to base editing, prime editing theoretically offers a more rapid, precise, and versatile approach to study the full spectrum of driver mutations, their combinations, and the growing catalog of genetic lesions that confer resistance to targeted therapies[22–25]. Beyond editing versatility, prime editing also avoids the formation of indel byproducts associated with DSBs. This is particularly important for studying SNVs with putative neomorphic qualities in tumor suppressor genes, as HDR-directed mutations would be diluted by the higher rate of naturally selected indels. Finally, the lower rates of off-target activity associated with prime editing reduce the risk of confounding off-target effects[21].

With these considerations in mind, we developed a prime editing GEMM (PE GEMM) that eliminates the need for exogenous delivery of prime editor genes, which can be challenging given their current size of six kilobases[26,27]. Expressing the prime editing machinery from an endogenous locus also minimizes any concerns of confounding acute or chronic anti-tumor immune responses that could be induced by exogenous delivery of a Cas9-based fusion protein[28–30]. In conjunction with the development of the PE GEMM, we also developed a range of DNA vectors and engineered pegRNAs (epegRNAs) that promote efficient prime editing in a variety of cell lines and organoids derived from these mice. With this toolset, we established new organoid models harboring *Trp53* mutations frequently found in pancreatic cancer patients but not modeled by current GEMMs of the disease, as well as clinically relevant *Kras* mutations associated with resistance to KRAS^{G12C} inhibitors. Finally, we harnessed PE GEMMs to model cancer *in vivo* by initiating lung adenocarcinoma in an autochthonous fashion and pancreatic adenocarcinoma via orthotopic transplantation of prime edited organoids. We expect PE GEMMs to both expand the landscape of achievable cancer-associated mutations and accelerate techniques required to study their function and elucidate associated therapeutic vulnerabilities.

Results

Quantification of human cancer-associated mutations amenable to modeling by prime editing in mice. Recent work has shown that base editing can be used to elucidate the function of specific cancer-associated genetic variants[31] and to systematically probe a large fraction of all possible alleles for genes and proteins of interest[32]. Base editors are primarily capable of engineering transition SNVs[20] (A•T>G•C or G•C>A•T), though the base editor architecture has recently been adapted to produce C•G>G•C transversions with variable efficiency[18,19,33–35]. In contrast, prime editors are capable of engineering all transition and transversion SNVs[21], as well as indels[36,37], expanding the potential for rapid modeling of genetic variants even further. To define the expanded editing capacity afforded by prime editing, we quantified the abilities of both base and prime editing to install specific somatic mutations identified from a cohort of

43,035 genetically-profiled cancer patients from the Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) dataset (**Fig. 1a,b and Supplementary Fig. 1**)[9,31]. Out of 422,822 mutations identified from targeted exon sequencing of 594 cancer-associated genes, 82.3% are SNVs, while the remaining 17.7% are deletions (DEL), insertions (INS), and di/oligo-nucleotide variants (DNVs, ONVs), in descending order of frequency (**Fig. 1a**).

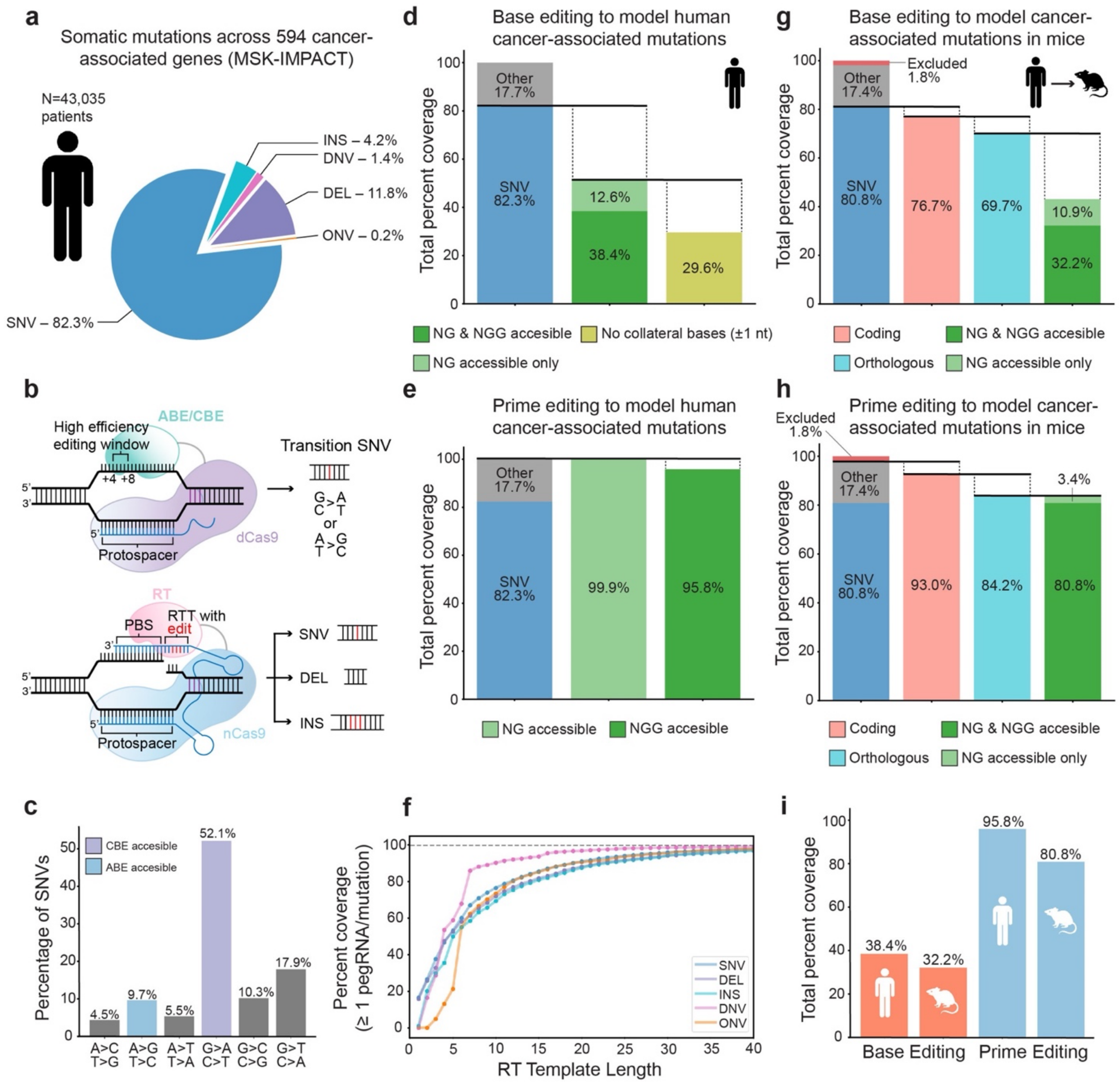


Figure 1. Quantification of human cancer-associated mutations amenable to modeling by base editing or prime editing in humans and mice.

a. Distribution of somatic variant types in a cohort of 43,035 patients with 422,822 mutations observed in 594 cancer-associated genes. Single nucleotide variants = SNV, deletions = DEL, insertions = INS, di-nucleotide variants = DNV, oligo-nucleotide variants = ONV.

b. Schematic of the modeling capabilities of base editing (top) and prime editing (bottom).

c. Quantification of somatic SNVs by type, illustrating enrichment for transition SNVs. Transition SNVs amenable to modeling by cytosine base editors (CBE) are shown in purple, while transition SNVs amenable to modeling by adenine base editors (ABE) are shown in blue. Transversions are shown in gray.

d. Quantification of mutations amenable to modeling with cytosine or adenine base editors that use an NG or NGG PAM. All percentages are given as a percentage of *all* mutations in the dataset. 38.4% of all mutations are amenable to base editing and fall within the protospacer of an NGG PAM (dark green), while an additional 12.6% of all mutations are amenable to base editing and fall within an NG PAM protospacer (light green). Of this subset of mutations that fall within either an NG or NGG PAM protospacer, only ~60%, or 29.6% of all mutations, lack matching collateral bases within one nucleotide (nt) of the mutation site.

e. 95.8% of all mutations in the dataset are potentially amenable to modeling by a prime editor using an NGG PAM (dark green) coupled with a pegRNA with a reverse transcription (RT) template length of 30 nucleotides. 99.9% of all mutations can be modeled by a prime editor using an NG PAM with the same pegRNA specifications.

f. Percentage of mutations with at least one suitable pegRNA as a function of the RT template length of the pegRNA, excluding the additional length of a homologous region in the RT template. Calculations assume the prime editor recognizes an NGG PAM.

g. Quantification of orthologous coding mutations potentially amenable to modeling by base editing in mice. Mutations are defined as orthologous if they derive from a wild-type amino acid conserved in the murine ortholog, as determined by pairwise protein alignment between human and mouse protein sequences. The rightmost bar indicates the fraction of orthologous coding mutations that can be modeled by base editors that recognize NG or NGG PAMs. "Excluded mutations" refers to mutations that fall in a gene lacking an ortholog. All percentages are given as a percentage of *all* mutations in the dataset.

h. Quantification of orthologous coding mutations potentially amenable to modeling by prime editing. Orthologous mutations are defined as in Fig. 1g. The rightmost bar indicates the ability of an NG or NGG prime editor to model these orthologous mutations, assuming an RT template greater than 30 nt. Excluded mutations are defined as in Fig. 1g.

i. Summary of the cancer mutation modeling capabilities of base and prime editing assuming an NGG PAM.

To estimate what fraction of common cancer-associated mutations are captured in currently available transgenic mouse models, we analyzed a dataset curated from the Mouse Genome Informatics database (see Methods)[38,39]. We found that 65 of the 100 most frequent SNVs in MSK-IMPACT (50/84 = 60% of missense SNVs) are not represented by published mouse cancer models (**Supplementary Tables 1 and 2**). Notably, the majority of these mutations are transition SNVs, which comprise 61.8% of all SNVs in the MSK-IMPACT dataset and are theoretically compatible with engineering using base editors (**Fig. 1c**). Still, only 51% of the total mutations are potentially amenable to base editing, depending on the positioning of requisite NG or NGG PAM sequences[20,31] (**Fig. 1d**). This value further drops to 29.6% when considering only mutations without identical bases present within one adjacent nucleotide (**Fig. 1d**). Adjacent identical nucleotides could be collaterally edited and potentially result in undesired editing outcomes. Moreover, this analysis does not account for the dependence of editing efficiency on a desired edit's location within the protospacer, which further reduces these theoretical base editing compatibility values (**Supplementary Fig. 2**).

We used a similar approach to quantify the modeling capabilities of prime editors that use an NGG or NG PAM coupled with variable reverse transcription template (RTT) lengths encoded within pegRNAs (**Supplementary Fig. 1**). Using an NGG PAM and RTT length of 30 base pairs (bp), excluding the additional length of a homologous region in the RTT, prime editing theoretically reaches 95.8% coverage of all mutations in this dataset — 2.5x the coverage of comparable base editors (**Fig. 1e**). This value increases to 99.9% for prime editors that use an NG PAM (**Fig. 1e**). Moreover, analysis of the relationship between RTT length and modeling capabilities reveals that ~85% of mutations in this dataset can be modeled by placing the mutation within the first 15 bp of the RTT (**Fig. 1f**). Importantly, these parameters are well within the recommended guidelines for pegRNA RTT length, even with the additional size required for a region of homology[20].

We also sought to determine the fraction of cancer-associated mutations that derive from protein sequences conserved in murine orthologs. We reasoned that only this subset of conserved sequences, when mutated in mouse systems, could be expected to mimic effects seen in human cancer. To quantify the ability of base and prime editors to model cancer-associated mutations in mice, we performed pairwise alignment on orthologous mouse and human proteins to define whether mutations derive from a conserved wild-type amino acid and reside in a region of homology (**Supplementary Fig. 1**). Of the 95% of SNVs that occur in coding sequences, 90.9% derive from codons that encode conserved amino acids between mouse and human (representing 69.7% of all mutations). Of these conserved, cancer-associated SNVs, 61.8% are amenable to base editing (NG or NGG PAM), which translates to 43.1% of all mutations in the dataset (**Fig. 1g**). In contrast, NG or NGG prime editors are capable of modeling 100% of coding mutations that occur at conserved amino acid residues in mice (84.2% of all mutations in the dataset) **(Fig. 1h**). In total, 80.8% of human cancer-associated mutations observed in this dataset could be modeled in mice with prime editors using a traditional NGG PAM, effectively doubling the modeling capabilities of base editors while also expanding to new classes of mutations beyond SNVs (**Fig. 1f,i**). This same pattern holds when filtering the dataset to only mutations that occur in multiple patients, and when considering various stringencies of homology in the regions flanking the mutations of interest (**Supplementary Figs. 2 and 3**). In total, these results demonstrate that prime editing could significantly broaden both the diversity and number of human cancer-associated mutations that can be rapidly modeled in murine orthologs.

Design and construction of a Cre-inducible prime editor allele. We sought to develop a transgenic system capable of precisely engineering the majority of cancer-associated mutations without requiring exogenous delivery of a prime editor enzyme. To accomplish this, we targeted a transgene expression cassette encoding the PE2 enzyme and the mNeonGreen (mNG)[40] fluorescent reporter, separated by the P2A ribosome skipping sequence, into the *Rosa26* locus[10,41] (**Fig. 2a**). Like previous Cre-inducible *Rosa26* alleles[10,42,43], transgene expression is driven by the CAGG promoter and is induced by Cre-mediated excision of a loxP-stop-loxP (LSL) cassette. A neomycin resistance gene was included to enable selection of cells containing the targeted allele. We also incorporated FRT/FRT3 sequences flanking the central construct to enable Flp recombinase-mediated replacement of the $Rosa26^{PE2}$ allele with future generations of prime editor enzymes or other desirable editors[44,45]. This vector was targeted to $Trp53^{flox/flox}$ C57BL/6J embryonic stem cells, where *Trp53* can be deleted upon expression of Cre recombinase (**Supplementary Fig. 4**). Chimeric mice were then crossed to wild-type C57BL/6J mice to generate pure strain heterozygous $Rosa26^{PE2/+}$;$Trp53^{flox/+}$ mice. These mice were subsequently crossed with $Rosa26^{PE2/+}$ $Trp53^{+/+}$ and $Trp53^{flox/flox}$ mice to generate $Rosa26^{PE2/+}$ mice on wild-type and $Trp53^{flox/flox}$ backgrounds.

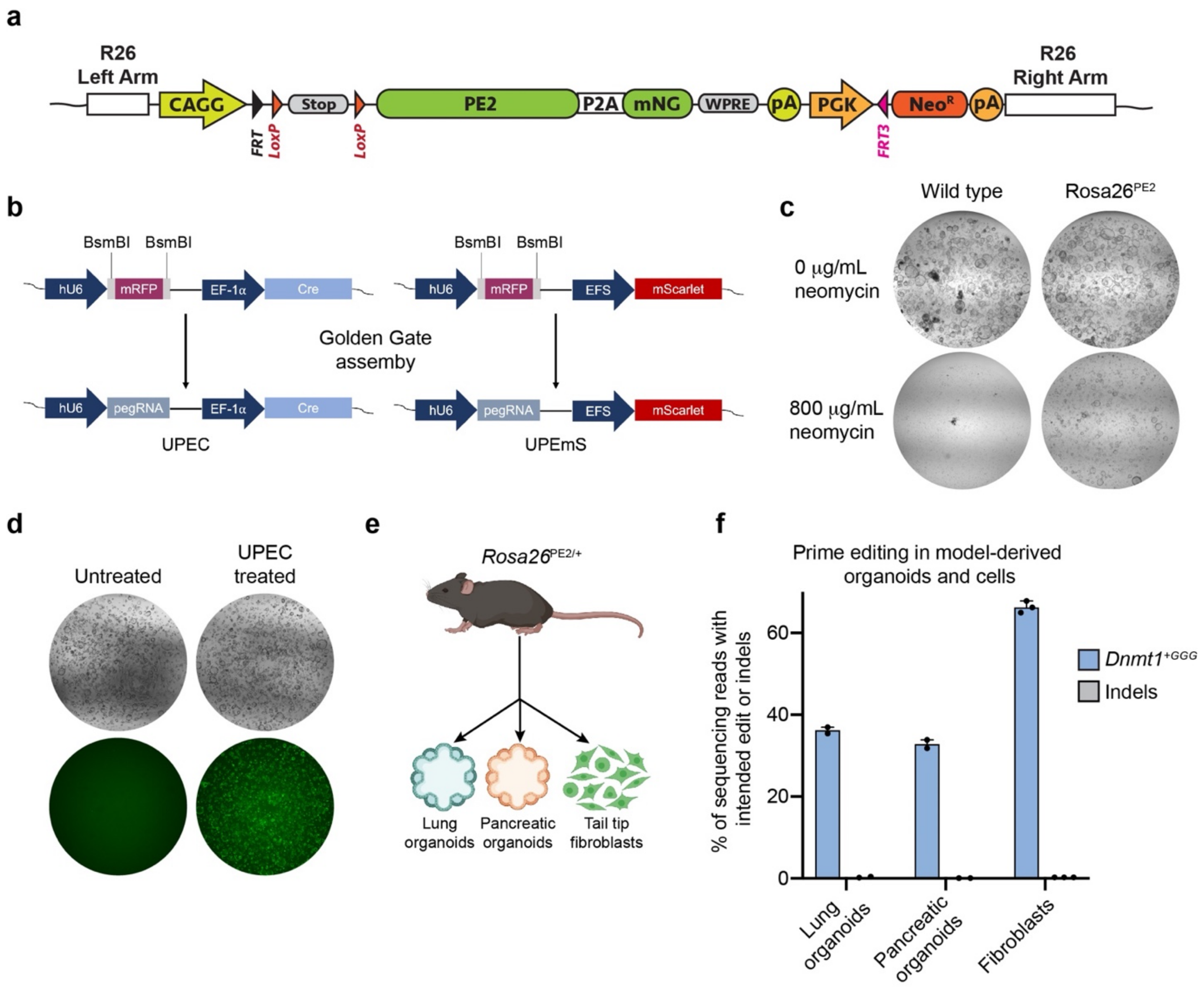


Figure 2. Design and functional validation of the *Rosa26^{PE2}* prime editor allele.

a. Schematic depicting the design of the Cre-inducible *Rosa26^{PE2}* allele.

b. Schematic depicting the formation of hU6-pegRNA-EF-1α-Cre (UPEC) and hU6-pegRNA-EFS-mScarlet (UPEmS) vectors from templates encoding a red fluorescent protein (RFP) by Golden Gate assembly.

c. Bright-field images of pancreatic organoids derived from chimeric prime editor mice and wild-type mice. With and without treatment with neomycin.

d. Bright-field and fluorescent images showing PE2-P2A-mNG expression only after exposure to Cre encoded by a UPEC vector.

e. Schematic depicting the derivation of multiple organoids and a fibroblast cell line from *Rosa26$^{PE2/+}$* prime editor mice.

Editing efficiency of a trinucleotide (+GGG) insertion located eight base pairs downstream of the start codon in *Dnmt1* in pancreatic organoids, lung organoids, and tail tip-derived fibroblasts (*n* = 2-3 biological replicates). Unintended indel byproducts in all conditions were present in <1% of sequencing reads.

Functional validation of the prime editor allele in organoids. To test the functionality of the *Rosa26*^{PE2} allele, we developed two lentiviral vectors that co-express a pegRNA and either Cre recombinase (UPEC) or the red fluorescent protein, mScarlet[46] (UPEmS) (**Fig. 2b**). We derived pancreatic organoids from chimeric *Trp53*$^{flox/flox}$*;Rosa26*$^{PE2/+}$ mice and developed a pure culture of transgene-containing cells via selection with neomycin (**Fig. 2c; Supplementary Fig. 5**). As expected, these pancreatic organoids displayed Cre-dependent mNG expression upon transduction with UPEC vectors (**Fig. 2d; Supplementary Fig. 5**). To test the prime editing functionality of this allele, we designed a single *Kras*-targeting pegRNA to introduce the common *Kras*^{G12D} mutation as a dinucleotide substitution (GGT>GAC), as well as a *Dnmt1*-targeting pegRNA encoding a +1 CCC insertion, which templates a trinucleotide insertion of a GGG glycine codon at residue 4 of *Dnmt1*. The *Dnmt1* pegRNA had been previously optimized and selected for its high activity, but the *Kras* pegRNA was assessed without prior optimization. UPEC-transduced organoids were selected using nutlin-3a, a mouse double minute 2 homolog (MDM2) inhibitor that induces cell cycle arrest in *Trp53*-proficient (but not deficient) cells[47]. After selection, we detected up to 33.8% editing efficiency at *Dnmt1*, 0.2% editing at *Kras*, and minimal indel byproducts in both cases (**Supplementary Fig. 5**). Together, these results validate the functionality of the *Rosa26*^{PE2} allele, including its ability to mediate prime editing of endogenous loci when using optimized pegRNAs.

Efficient prime editing at multiple loci in tissues derived from the *Rosa26*^{PE2} model. We next sought to evaluate prime editing across multiple tissues and target loci. To accomplish this, we derived lung organoids, pancreatic organoids, and tail tip-derived fibroblasts (TTFs) from multiple *Rosa26*$^{PE2/+}$ mice (**Fig. 2e**). Consistent with results using chimera-derived organoids, we observed highly efficient *Dnmt1* editing across all these tissues (**Fig. 2f**). Given the low prime editing efficiency of the initial *Kras* pegRNA, we next sought to empirically identify highly efficient pegRNAs that introduce the *Kras*^{G12D} transition as a single nucleotide variant (GGT>GAT). Based on previous work[48], we hypothesized that spacer sequences capable of producing the highest Cas9 indel efficiency in mouse N2A cells would serve as ideal scaffolds for high-efficiency pegRNA designs (**Supplementary Fig. 6 and Supplementary Table 3**). Using TTFs, we observed up to ~5% editing efficiency of *Kras*^{G12D} with spacer-optimized pegRNAs (**Fig. 3a and Supplementary Fig. 6**). To further amplify editing efficiency, we modified our best-performing pegRNA with a structured RNA pseudoknot motif, prequeosine$_1$-1 riboswitch aptamer (tevopreQ1), recently shown to enhance prime editing efficiency by more than 3-fold in cell lines[49]. This resulted in up to ~18.4% editing efficiency of *Kras*^{G12D} in pancreatic organoids and TTFs (**Fig. 3b**). We then modified this epegRNA to template the *Kras*^{G12C} transversion and observed reduced efficiency at ~0.5% in pancreatic organoids and ~5% in TTFs. We also generated *Kras*^{G12A} and *Kras*^{G12R} epegRNAs and observed up to ~30% editing efficiency with both epegRNAs in TTFs (**Fig. 3b**). Both *Kras*^{G12A} and *Kras*^{G12R} epegRNAs template G·C-to-C·G substitutions, which proceed from C·C mismatch intermediates. C·C mismatches are not efficiently repaired by mismatch repair and are thought to have higher basal prime editing rates as a consequence[44]. Collectively, these data demonstrate that the *Rosa26*^{PE2} allele enables efficient installation of both SNVs and insertions across a diverse array of cell lines and organoids.

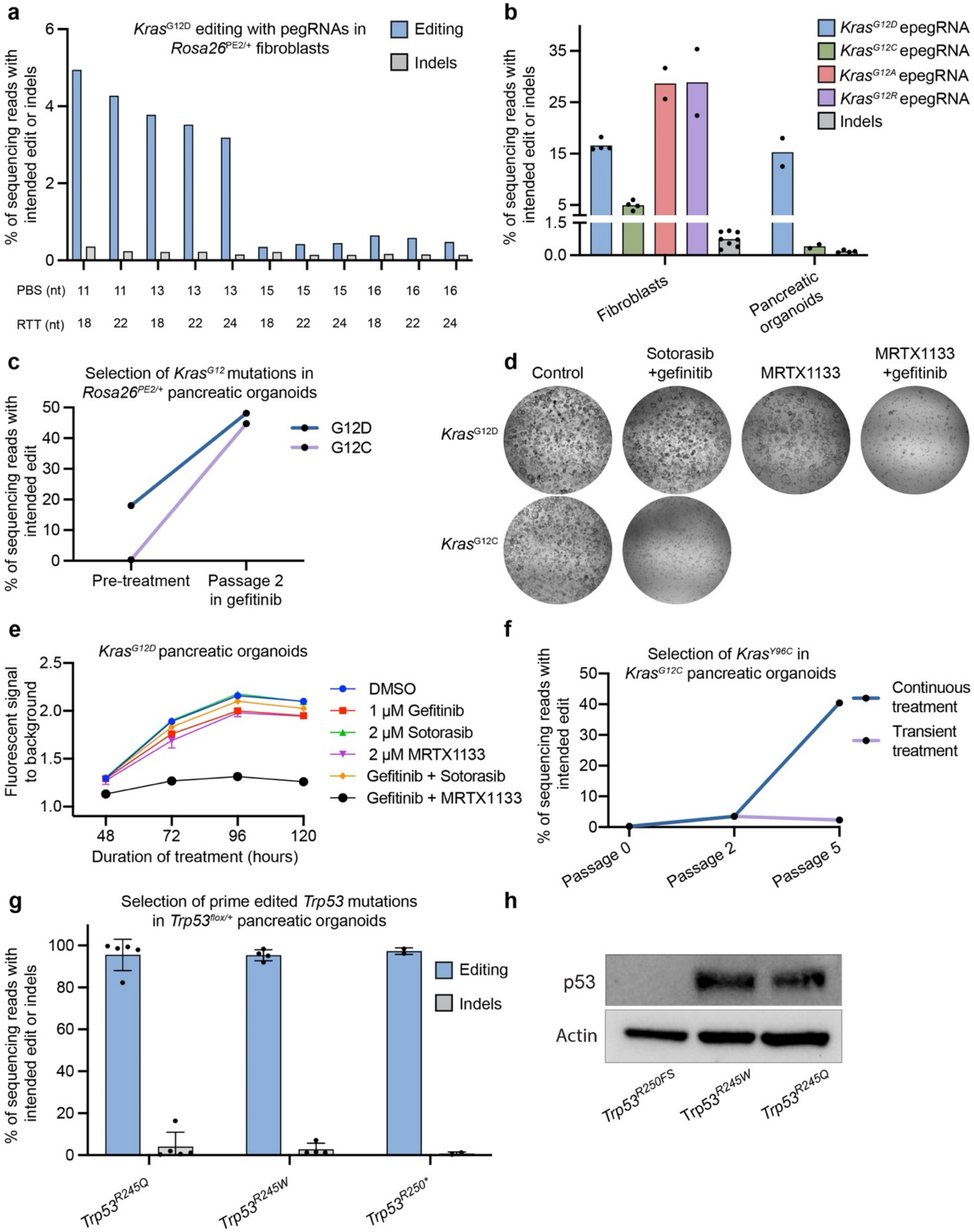


Figure 3. ***Ex vivo*** **prime editing and functional testing of** ***Kras*** **and** ***Trp53*** **mutations.**

a. Editing efficiency and indel byproduct frequency of the *Kras*^{G12D} transition mutation (G:C to A:T) templated by a cohort of pegRNAs based on a single Cas9 spacer (*n* = 1 for each pegRNA). pegRNAs are delineated by differences in the lengths of the primer binding site (PBS) and reverse transcriptase template (RTT).

b. Editing activity of four engineered pegRNAs (epegRNAs) templating either the *Kras*^{G12D} transition or the *Kras*^{G12C}, *Kras*^{G12A}, or *Kras*^{G12R} transversions in tail tip-derived fibroblasts or pancreatic organoids (*Kras*^{G12D} and *Kras*^{G12C}). *n* = 2-4 biological replicates. epegRNAs were generated by appending the trimmed evopreQ$_1$ motif after the primer binding site of the leftmost pegRNA depicted in Fig. 1a. Indel byproduct calculations were pooled from all conditions within each tissue.

c. Allele frequencies of *Kras*^{G12D} or *Kras*^{G12C} mutations in pancreatic organoids before and after 2 passages of treatment with gefitinib (1 μM) (*n* = 1). Gefitinib treatment selects for cells containing prime edited *Kras*^{G12D} or *Kras*^{G12C} mutations.

d. Bright-field images of prime edited *Kras*^{G12C} or *Kras*^{G12D} organoids treated for four days with either control DMSO, sotorasib (2 μM) and gefitinib (1 μM), MRTX1133 (5 μM), or MRTX1133 and gefitinib.

e. Viability of *Kras*^{G12D} pancreatic organoids under various treatment conditions. Viability was quantified using the alamarBlue HS Cell Viability Reagent, which is metabolized into a fluorescent derivative in living cells.

f. Allele frequency of *Kras*^{Y96C} in *Kras*^{G12C} organoids during and after treatment with sotorasib (2 μM) and gefitinib (*n* = 1). After two passages, organoids were split into two groups, which included continued treatment (continuous treatment) in one group and removal of treatment in a second group (transient treatment).

g. Allele frequencies of three *Trp53* mutations in *Trp53*$^{flox/+}$ pancreatic organoids treated with nutlin-3a for three to five passages after transduction with UPEC vectors (*n* = 2-5 biological replicates). Indel byproduct frequencies are also included. Note that the highest indel frequency depicted for *Trp53*^{R245W} derives primarily from a scaffold insertion in a single replicate. *Trp53*^{R245Q} and *Trp53*^{R245W} are homologous to mutations commonly observed in pancreatic cancer patients (*TP53*^{R248Q} and *TP53*^{R248W}), as described in Supplementary Figure 8. *Trp53*^{R250FS} denotes a dinucleotide deletion that induces a frameshift mutation.

h. Immunoblot indicating detectable levels of p53 protein in prime edited *Trp53*$^{flox/R245Q}$ and *Trp53*$^{flox/R245W}$ organoids and an absence of detectable protein in *Trp53*$^{flox/R250FS}$ organoids.

To confirm the functional effects of these mutations, we installed either *Kras*^{G12D} or *Kras*^{G12C} mutations in *Trp53*$^{flox/flox}$ *Rosa26*$^{PE2/+}$ pancreatic organoids and selected transduced cells with nutlin-3a. We then treated prime edited organoids with the epidermal growth factor receptor (EGFR) inhibitor, gefitinib, to select for the oncogenic *Kras* mutation[50] and evaluated the fraction of cells containing the intended edits before and after treatment. Consistent with receptor-independent signaling downstream of EGFR, cells transduced with *Kras*^{G12D} and

Kras^{G12C} epegRNAs survived treatment with gefitinib, while control cells infected with the template UPEC lacking a pegRNA did not (**Fig. 3c and Supplementary Fig. 7**). We then tested whether cells transduced with Kras^{G12C} epegRNAs were sensitive to sotorasib, a KRAS^{G12C}-specific inhibitor, alone or in combination with gefitinib. Consistent with previous work[51], we found that Kras^{G12C} pancreatic organoids were uniquely sensitive to the combination of sotorasib and gefitinib, while Kras^{G12D} organoids were unaffected by these treatments (**Fig. 3d** and **Supplementary Fig. 7**). While KRAS^{G12C} inhibition has shown promising signs of clinical efficacy in pancreatic cancer[5,52], current preclinical efforts focused on KRAS^{G12D} inhibition have the potential to benefit a broader fraction of patients with this disease (>38%)[53,54]. We therefore treated prime edited Kras^{G12D} pancreatic organoids with MRTX1133[55], a KRAS^{G12D} inhibitor, alone or in combination with gefitinib. Consistent with results using sotorasib, we found that Kras^{G12D} organoids were significantly more sensitive to the combination treatment compared with MRTX1133 alone (**Fig. 3d,e and Supplementary Fig. 7**), suggesting that concomitant EGFR inhibition may be a broadly effective strategy to augment the overall efficacy of KRAS mutant inhibitors in pancreatic cancer cells. Surprisingly, we also found that Kras^{G12C} organoids were sensitive to this combination and that the effect was comparable to that of sotorasib and gefitinib (**Supplementary Fig. 7**).

Prime editing enables rapid functional interrogation of putative resistance mutations.

While targeted therapies have revolutionized modern cancer treatment, therapy resistance is common and frequently arises through the acquisition of secondary missense mutations affecting the drugged driver[25,56,57]. Recent work by Awad and colleagues revealed a novel class of secondary *KRAS* mutations occurring in over 10% of non-small cell lung cancer and colorectal cancer patients that exhibited initial responses to treatment with adagrasib, a KRAS^{G12C}

inhibitor[56]. Intriguingly, several mutations occur in codons 95-96, which occupy the switch II pocket targeted by adagrasib and sotorasib.

To test the utility of the *Rosa26*^{PE2} model to functionally interrogate novel mutations associated with resistance, we developed an epegRNA designed to introduce the *Kras*^{Y96C} transversion and tested its capacity to promote resistance in prime edited *Kras*^{G12C} pancreatic organoids treated with gefitinib and sotorasib (**Supplementary Fig. 7)**. All organoids were treated with both inhibitors for an initial two passages, followed by continued treatment for three additional passages in one group (continuous treatment) and treatment removal in the second group (transient treatment). Consistent with patient data[56], organoids transduced with the *Kras*^{Y96C} epegRNA were resistant to combined treatment with gefitinib and sotorasib and exhibited increased allele frequency of the *Kras*^{Y96C} mutation over time (**Fig. 3f**). Positive selection for composite *Kras*$^{G12C;Y96C}$ mutant organoids was not observed in organoids following the removal of gefitinib and sotorasib, confirming the requirement of the selective pressure exerted by the treatment. Though initially discovered in lung cancer patients treated with sotorasib monotherapy, these data indicate that secondary *KRAS* mutations can also confer resistance in other tissue and combination treatment contexts. The above results demonstrate that the *Rosa26*^{PE2} allele can be harnessed for rapid preclinical evaluation of emerging mechanisms of resistance to targeted therapies in tissues of interest and, ultimately, for testing second-generation therapies designed to overcome resistance.

Rapid engineering of common p53 mutations using prime editing. A key advantage of PE GEMMs is the ability to mediate nearly any codon substitution in any accessible tissue, thereby enabling tissue-specific functional studies of genetic variants with putative effects on tumor progression. *TP53* is the most frequently mutated gene in human cancer and is often altered via missense mutations that can confer gain-of-function properties in certain contexts[58]. In an analysis of data from cBioPortal[59,60], we found that some of the most frequent p53 amino acid substitutions observed in lung (*TP53*^{R158L} and *TP53*^{R270L}) and pancreatic adenocarcinoma (*TP53*^{R248W} and *TP53*^{R248Q}) are not represented in published GEMMs of these diseases (**Supplementary Fig. 8**), despite having putative gain-of-function effects[61–63]. Notably, three of these mutations are transversions that cannot be modeled using base editing. Furthermore, the human amino acid (p53^{R248}) but not the human codon (CGG vs CGC) is conserved in murine *Trp53*. Thus, engineering the *Trp53*^{R245W} mutation in mice requires a dinucleotide substitution uniquely suitable to prime editing (**Supplementary Fig. 8**). We developed a suite of epegRNAs designed to introduce each of these mutations and a truncating deletion, *Trp53*^{R250FS}, using a *Trp53*$^{+/+}$ cell line derived from murine 3TZ cells (**Supplementary Fig. 9**). We achieved up to 6.4% basal editing efficiency for *Trp53*^{R245Q} and lower efficiencies (<3%) for other *Trp53* edits. After selection with nutlin-3a, most *Trp53*$^{flox/+}$;*Rosa26*$^{PE2/+}$ pancreatic organoids transduced with each of these epegRNAs exhibited a prime edited allele frequency near 100%. Remarkably, we also observed >90% editing purity in most of these organoids, demonstrating that each of these mutations can be installed with high precision using prime editing (**Fig. 3g; Supplementary Fig. 8**). While the ratio of prime edited reads to reads bearing random indels was typically high, we did observe a variable unintended single nucleotide substitution (0.24% - 11.34% of reads) attributable to partial transcription of the scaffold sequence when prime editing *Trp53*^{R245Q} (**Supplementary Fig. 8**). In one instance, we also observed an insertion of the scaffold sequence when prime editing *Trp53*^{R245W} (~7% of reads). Western blots of the resulting organoids confirmed retained p53 protein expression in the *Trp53*^{245Q} and *Trp53*^{R245W} conditions but not in the *Trp53*^{R250FS} condition (**Fig. 3h**). These results establish the utility of our approach for rapid installation of novel mutations using systems that can be easily translated to an *in vivo* setting.

Modeling lung and pancreatic adenocarcinoma *in vivo* using prime editing. To benchmark the utility of PE GEMMs to model cancer *in vivo*, we initiated lung and pancreatic adenocarcinomas using autochthonous and orthotopic transplantation strategies, respectively (**Fig. 4a**). To model lung cancer, we intratracheally infected *Trp53*$^{flox/flox}$;*Rosa26*$^{PE2/+}$ and *Trp53*$^{flox/flox}$;*Rosa26*$^{PE2/PE2}$ mice with UPEC lentiviruses encoding the template vector (n=4) or the optimized *Kras*^{G12D} (n=14), the less efficient *Kras*^{G12C} (n=13), or the neutral *Dnmt1*$^{+GGG}$ pegRNAs (n=6). Tumors initiated by UPEC-*Kras*^{G12D} were readily visible by μCT at 14 weeks post-injection (**Supplementary Fig. 10)**. At 16 weeks, we observed multifocal fluorescent lesions in 10/14 (71%) UPEC-*Kras*^{G12D} recipients and in none of the controls (**Fig 4b**).

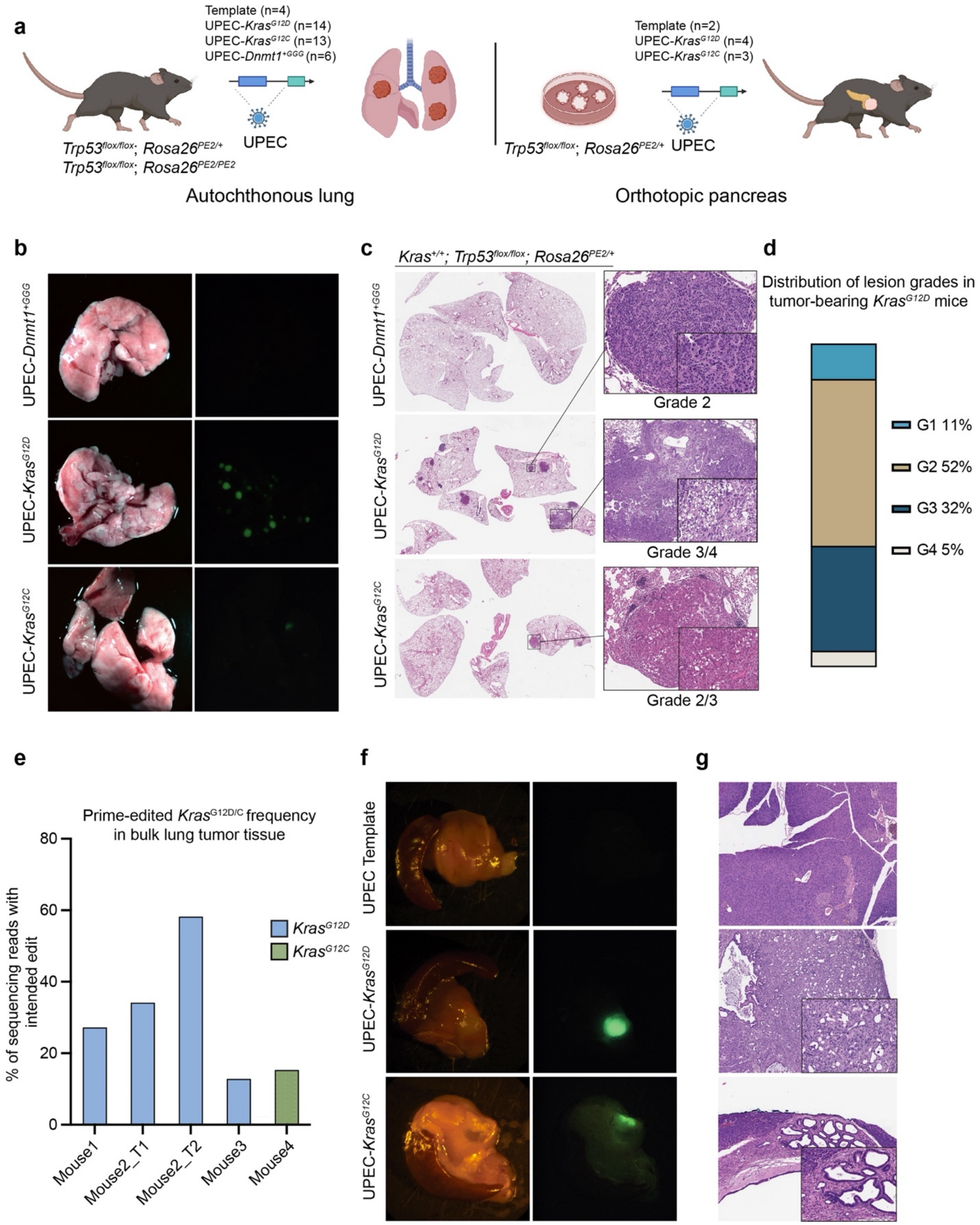

a
Template (n=4)
UPEC-$Kras^{G12D}$ (n=14)
UPEC-$Kras^{G12C}$ (n=13)
UPEC-$Dnmt1^{+GGG}$ (n=6)
$Trp53^{flox/flox}$; $Rosa26^{PE2/+}$
$Trp53^{flox/flox}$; $Rosa26^{PE2/PE2}$
UPEC
Autochthonous lung
Template (n=2)
UPEC-$Kras^{G12D}$ (n=4)
UPEC-$Kras^{G12C}$ (n=3)
$Trp53^{flox/flox}$; $Rosa26^{PE2/+}$
UPEC
Orthotopic pancreas
b
UPEC-$Dnmt1^{+GGG}$
UPEC-$Kras^{G12D}$
UPEC-$Kras^{G12C}$
c
$Kras^{+/+}$; $Trp53^{flox/flox}$; $Rosa26^{PE2/+}$
UPEC-$Dnmt1^{+GGG}$
UPEC-$Kras^{G12D}$
UPEC-$Kras^{G12C}$
Grade 2
Grade 3/4
Grade 2/3
d
Distribution of lesion grades in tumor-bearing $Kras^{G12D}$ mice
G1 11%
G2 52%
G3 32%
G4 5%
e
Prime-edited $Kras^{G12D/C}$ frequency in bulk lung tumor tissue
% of sequencing reads with intended edit
80
60
40
20
0
$Kras^{G12D}$
$Kras^{G12C}$
Mouse1
Mouse2_T1
Mouse2_T2
Mouse3
Mouse4
f
UPEC Template
UPEC-$Kras^{G12D}$
UPEC-$Kras^{G12C}$
g

Figure 4. PE GEMMs enable autochthonous and orthotopic modeling of lung and pancreatic cancer.

a. Schematic depicting the design of *in vivo* experiments. Autochthonous lung tumors were initiated with lentivirus encoding UPEC vectors. Pancreatic tumors were initiated by orthotopic transplantation of prime edited pancreatic organoids. "Template" refers to the template UPEC vector lacking a pegRNA.
b. Representative bright-field and fluorescent images of lungs derived from mice infected with the UPEC vector encoding the neutral *Dnmt1*$^{+GGG}$ pegRNA (top) or the *Kras*^{G12D} (middle) or *Kras*^{G12C} (bottom) epegRNAs described in Fig. 3b.
c. Hematoxylin and eosin (H&E) staining of representative tissue from a control mouse infected with UPEC-*Dnmt1*$^{+GGG}$ (top), and tumor-bearing mice infected with UPEC-*Kras*^{G12D} (middle) or UPEC-*Kras*^{G12C} (bottom). Callout boxes highlight a Grade 2 lesion (top) and a Grade 3/4 lesion (bottom) from *Kras*^{G12D}, as well as a Grade 2/3 lesion from *Kras*^{G12C}, at a greater magnification.
d. Stacked proportionality bar chart indicating the distribution of grades across lesions from nine tumor-bearing mice infected with UPEC-*Kras*^{G12D}. Lesion grades were called using the Aiforia algorithm as described in Methods.
e. Allele frequency of *Kras*^{G12D} or *Kras*^{G12C} in bulk lung tumors. Indel byproduct frequency was calculated as <1% in all cases. X-axis bar titles refer to specific mice. Mouse 2 had two tumors (T1 & T2) isolated for amplicon sequencing.
f. Representative bright-field and fluorescent images of pancreatic tissue in mice transplanted with *Trp53*$^{flox/flox}$ organoids treated with the parental UPEC vector (top) or organoids prime edited to contain *Kras*^{G12D} (middle) or *Kras*^{G12C} (bottom). Organoids were transplanted after selection of the prime edited allele using gefitinib, as described in Fig. 3.
g. H&E staining of representative pancreatic tissue from a control mouse (top), a mouse transplanted with *Kras*^{G12D} organoids (middle), and a mouse transplanted with *Kras*^{G12C} organoids (bottom). Histological features indicate a low-grade pancreatic intraepithelial neoplasia (PanIN) initiated by *Kras*^{G12C} and a more advanced pancreatic adenocarcinoma initiated by *Kras*^{G12D}.

Histological analysis confirmed that lesions induced by prime editing recapitulated the full spectrum of lung cancer progression, from grade 1 atypical adenomatous hyperplasia through grade 4 adenocarcinoma. By immunohistochemistry, prime edited tumors recapitulated the cellular and molecular evolution seen in the classical *Kras*$^{LSL\text{-}G12D/+}$;*Trp53*$^{flox/flox}$ (KP) GEMM model, demonstrating downregulation of lung lineage transcription factor *Nkx2-1* and expression of chromatin regulator *Hmga2* in poorly differentiated, advanced lesions[64–66] (**Fig. 4c-d, Supplementary Fig. 11**). We confirmed that tumors were initiated through on-target prime editing by sequencing genomic DNA derived from several bulk tumors (**Fig. 4e**).

In contrast to UPEC-*Kras*^{G12D} recipients, only 4/13 (31%) UPEC-*Kras*^{G12C} recipients presented tumors when harvested at 19 weeks. In these mice, we observed a single tumor lesion in three animals and two lesions in the fourth. The lower tumor efficiency and burden in this cohort is likely a consequence of the lower prime editing efficiency of the *Kras*^{G12C} epegRNA, although it could also be related to weaker oncogenicity of the *Kras*^{G12C} allele[51] in the context of our lung model. Regardless, these findings support the notion that PE GEMMs can be used to model cancer *in vivo*, even when using lower efficiency pegRNAs. Notably, across all UPEC-*Kras*^{G12D} and UPEC-*Kras*^{G12C} recipients, we observed tumor formation in 8/20 *Trp53*$^{flox/flox}$;*Rosa26*$^{PE2/+}$ mice (8/12 UPEC-*Kras*^{G12D} and 0/8 UPEC-*Kras*^{G12C}) and 6/7 *Trp53*$^{flox/flox}$;*Rosa26*$^{PE2/PE2}$ mice (2/2 UPEC-*Kras*^{G12D} and 4/5 UPEC-*Kras*^{G12C}). The difference in tumor formation between *Rosa26*$^{PE2/+}$ and *Rosa26*$^{PE2/PE2}$ recipients was significantly higher for UPEC-*Kras*^{G12C} ($P < 0.01$ for *Kras*^{G12C}, Fisher's Exact Test). Thus, two alleles of *Rosa26*^{PE2} likely facilitate higher prime editing efficiency, enabling tumor formation in the majority of animals infected with a low-efficiency epegRNA.

To further test the potential of PE GEMMs for cancer modeling *in vivo*, we transplanted prime edited *Kras*$^{G12D/+}$;*Trp53*$^{flox/flox}$;*Rosa26*$^{PE2/+}$ and *Kras*$^{G12C/+}$;*Trp53*^{floxflox};*Rosa26*$^{PE2/+}$ pancreatic organoids into immunocompetent mice harboring the *Rosa26*^{PE2} allele (to ensure immunological tolerance[67] to the prime editor enzyme). As controls, we transplanted *Trp53*$^{flox/flox}$;*Rosa26*$^{PE2/+}$ organoids infected with the template UPEC vector. Tumors were visible via ultrasound by five weeks, and fluorescent tumors that reflected the spectrum of pancreatic neoplasia were observed in 3/4 *Kras*$^{G12D/+}$ recipients six weeks post-transplantation (**Fig. 4f-g**). A small, low-grade pancreatic intraepithelial neoplasia (PanIN) lesion not detected by fluorescent microscopy was observed in the fourth *Kras*$^{G12D/+}$ recipient by histology. Notably, only one mouse from the cohort of animals transplanted with *Kras*$^{G12C/+}$ pancreatic organoids developed a lesion, specifically a low-grade PanIN, consistent with prior observations suggesting that *Kras*$^{G12C/+}$ may be less tumorigenic in the pancreas[51]. We did not observe tumor formation in control recipients by ultrasound, microscopy, or histology, consistent with prior work showing that *Trp53* knockout alone is insufficient for pancreatic tumorigenesis[43,68].

Discussion

Advances in genome editing technologies have accelerated functional genetic studies, yet many approaches modeling cancer mutations to date have relied on Cas9-mediated gene disruption via non-homologous end joining and, therefore, fail to recapitulate specific genetic lesions observed in human cancer. Emerging precision genome editing technologies like base editing and prime editing are poised to fill this gap by allowing the engineering of specific cancer-associated mutations. Nevertheless, the considerable size of base editors and prime editors makes delivery to most tissues and cell types challenging and poses significant limitations for *in vivo* studies. Prior studies have addressed this using split-prime editor systems that enable prime editing *in vivo* when delivered by dual adeno-associated virus (AAV) vectors. However, dual-AAV approaches are still hampered by challenges of delivery to many tissues and, importantly, they elicit an immune response against the prime editor enzyme[30,69]. The immunogenicity of genome editing reagents when delivered exogenously complicates cancer modeling experiments, particularly given the important role of the immune system in tumor surveillance and the landmark success of immunotherapies and targeted therapies that modulate immune recognition and clearance of tumor cells. With these challenges in mind, we developed a PE GEMM capable of rapidly installing a variety of genetic lesions with single nucleotide precision across *in vitro*, *ex vivo*, or *in vivo* contexts, as well as in an autochthonous, immunocompetent setting. By expressing the PE2 enzyme endogenously, we obviate the risk of a confounding immune response and significantly expand the capacity to deliver other functional cargo, such as Cre. We used this model to install a variety of cancer-associated mutations, including transversions, transitions, dinucleotide substitutions, and deletions across *Kras* and *Trp53*. In the context of our pancreatic orthotopic transplant experiments, we observed that different *Kras* mutations may exhibit variable *in vivo* tumor initiating potential, consistent with prior work comparing *Kras*^{G12C} and *Kras*^{G12D} autochthonous models in the pancreas[51]. Importantly, we also observed *Kras* allele-specific responses to mutant-specific targeted therapies. For example, similar to prior studies of KRAS$^{\mathrm{G12C}}$ inhibitors[51,70], we found that a KRAS$^{\mathrm{G12D}}$ inhibitor, MRTX1133, elicits a more powerful effect on prime edited *Kras*$^{\mathrm{G12D}}$ pancreatic organoids when combined with the EGFR inhibitor, gefitinib. Several other clinical agents targeting a broader spectrum of oncogene mutations are undergoing clinical evaluation, and the first KRAS$^{\mathrm{G12C}}$ inhibitor, sotorasib, has now been approved by the FDA[53,55]. PE GEMMs represent ideal systems for rapid interrogation of the effects of targeted therapies in the context of virtually any oncogenic mutation, including secondary resistance mutations, like *KRAS*^{Y96C}, that are now being identified in patients. Importantly, PE GEMMs also enable *in vivo* studies of these mutations in the context of syngeneic and immunocompetent mice, and our results show that autochthonous tumors can be generated efficiently in *Rosa26*$^{PE2/PE2}$ mice even with a suboptimal *Kras*^{G12C} epegRNA. This broad utility for modeling *Kras* mutations *in vivo* is critical, especially given that mutant KRAS inhibition has been shown to affect the immune milieu and tumor microenvironment in models of colon cancer[71] and may synergize with immune checkpoint blockade in other tissues not yet examined.

Beyond *KRAS*, we demonstrate in pancreatic organoids the precise installation and selection of two *Trp53* dinucleotide substitutions encoding two mutant amino acid residues frequently observed at the same codon in human pancreatic cancer, as well as a dinucleotide deletion at a nearby codon. Importantly, we observed over 90% editing purity after selection of all three mutations *in vitro*. Despite a high intended edit-to-unintended indel ratio, we observed an unintended single nucleotide substitution at variable frequency when prime editing *Trp53*^{R245Q} (**Supplementary Fig. 8**). We attribute this event to partial homology between the genomic region immediately following the reverse transcriptase template and the few nucleotides in the pegRNA scaffold that are commonly reverse-transcribed and excised during DNA repair, a prime editing intermediate noted by Nelson and colleagues[49]. Unintended edits such as this could be avoided by using an alternative pegRNA with a reverse transcriptase template ending a few nucleotides up or downstream to eliminate the homology, or could be reduced by introducing silent edits that prevent repeated editing of the same target site. However, this phenomenon merits additional caution during pegRNA design and may be exacerbated in long-term prime editing experiments such as when selecting cell lines over several passages with continuous expression of the prime editor and pegRNA.

The overall editing purity we observed highlights the utility of prime editing for precise engineering of mutations with negligible indel byproducts. This is a key advantage over Cas9 HDR-based approaches, in which the high rate of indel byproducts could dilute intended point mutations *in vitro* and *in vivo*. Low editing purity could especially limit the study of specific point mutations in tumor suppressor genes, as unintended indels in these genes can produce frameshift mutations subject to positive selection. This limitation is especially important when considering that many genes, including *TP53*, often harbor point mutations that confer different properties relative to loss-of-function truncations, including gain-of-function effects[61,72–74]. For instance, Schulz-Heddergott and colleagues demonstrated that *TP53*^{R248Q} exhibits a gain-of-function effect by hyperactivating the JAK2/STAT3 pathway, leading to more aggressive tumor progression in models of colon cancer[61]. These observations remain untested in models of pancreatic cancer *in vivo* due to a lack of suitable transgenic mouse models and human cell lines[73]. PE GEMMs are poised to fill critical gaps like this by allowing rapid and fine-tuned mutation control in a variety of tissue settings.

Though we did not explore them here, a variety of techniques are available to optimize prime editing efficiency, such as PE3 and PE3b editing strategies that combine nicking guides to bias DNA repair toward incorporation of prime edited nucleotides. Nevertheless, strategies based on single pegRNAs are more straightforward, have better multiplexing capacity because they rarely cause indels, and are better suited for high-throughput studies like genetic screens. Recent work has also detailed the use of silent or benign mutations close to the intended edit as a means to avoid mismatch repair of prime edited alleles and thereby enhance overall efficiency[44]. While we found that spacer optimization and testing of up to 15 guides was sufficient to identify epegRNAs suitable for our experiments, future users should consider these and other strategies, including the co-delivery of an MLH1 dominant negative gene (PE4/5)[44] or sensor-based pegRNA library approaches[31], to maximize overall prime editing efficiencies, which may be especially helpful for *in vivo* applications. Future models in development based on next-generation prime editor enzymes, such as the optimized PEmax prime editor protein[44], will further facilitate these applications.

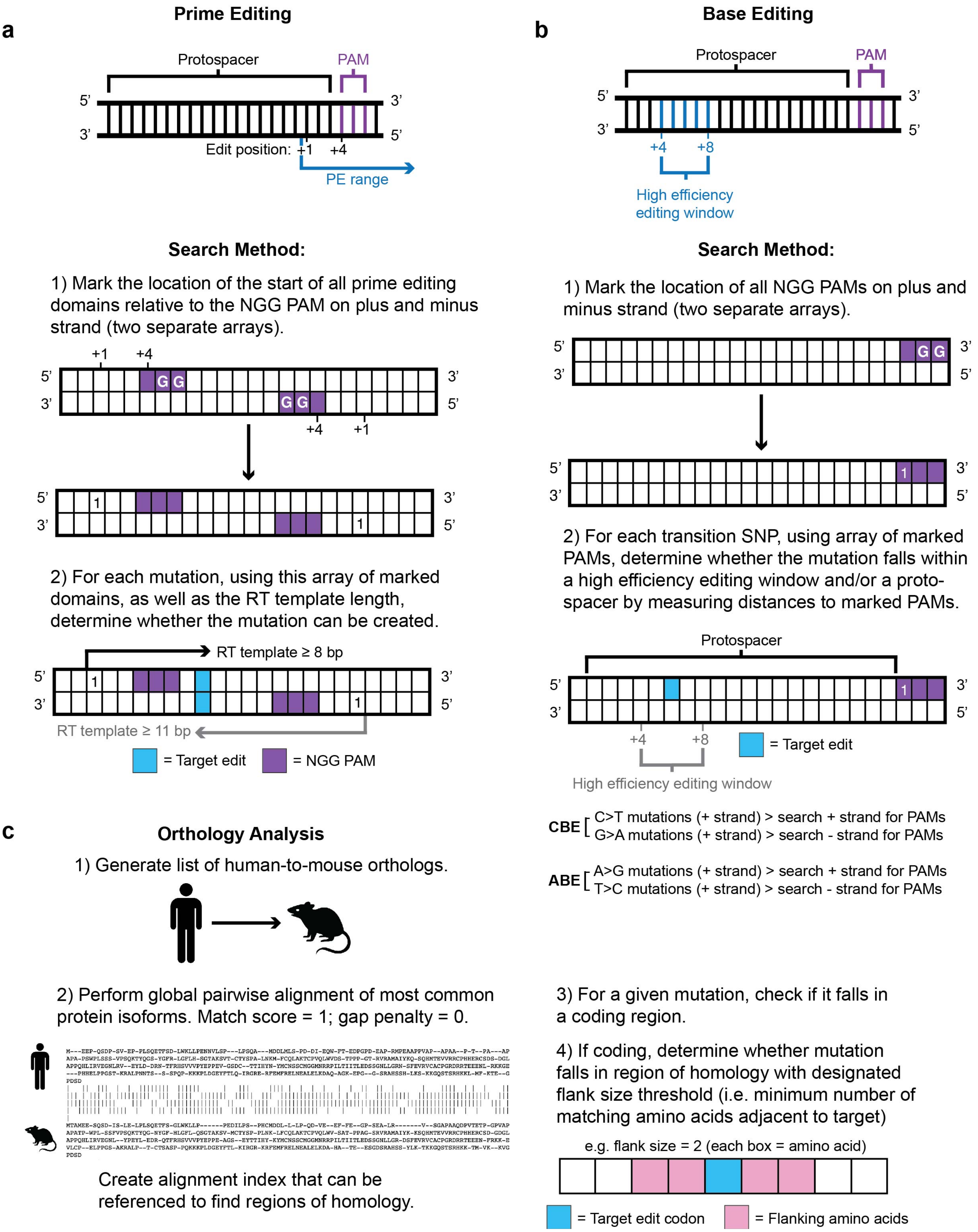
While we focused on installing cancer driver mutations, we note that PE GEMMs could be employed for broader applications, including modeling genetic diseases beyond cancer. We also envision applications involving the insertion of custom neoepitopes and other functional genetic sequences. These would enable investigators to address key questions in cancer genetics and immunology, while further reducing the need to generate, genotype, and otherwise maintain traditional GEMMs. Finally, the combination of multiple epegRNAs in the context of a modified UPEC vector should enable autochthonous generation of tumors defined by custom sets of multiple driver mutations in wild-type prime editor mice. This would enable increasingly complex studies of cooperating driver mutations. In general, PE GEMMs can provide a rapid preclinical avenue to complement clinical investigations aimed at treating cancer with novel precision treatment paradigms.

Acknowledgments:

We thank the entire Jacks Laboratory with particular thanks to Demi Sandel, Carla Concepcion, Megan Burger, William Hwang, and Alex Jaeger for useful discussions. We also thank Karen Yee, Judy Teixeira, Katherine Anderson, and Margaret Magendantz for administrative support. We thank Steve Murray, Roderick Bronson, and Milton Cornwall-Brady for helpful advice. S.I.G. and F.J.S.R. thank Edward Kastenhuber for useful advice on human and mouse ortholog analyses, as well as David Solit, Niki Schultz, Michael Berger, and Benjamin Gross for access to MSK-IMPACT data. Z.A.E. acknowledges support from the Ludwig Center at MIT and the National Cancer Institute of the National Institutes of Health under Award Number F31CA268835. F.J.S.R. is a HHMI Hanna Gray Fellow and was supported by NCI Cancer Center Support Grant P30-CA1405, the Ludwig Center at MIT (2036636), Koch Institute Frontier Awards (2036648 and 2036642), and the MIT Research Support Committee (3189800). G.A.N. was supported by a Helen Hay Whitney Postdoctoral Fellowship and K99 award

HL163805. We also thank the Koch Institute Swanson Biotechnology Center for technical support, especially the Histology core facility. This work was also supported by the Howard Hughes Medical Institute, David H. Koch Graduate Fellowship Fund, NCI Cancer Center Support Grant P30-CA1405, NIH Pre-Doctoral Training Grant T32GM007287, NIH grants U01 AI142756, RM1 HG009490, R35 GM118062, and the Lustgarten Foundation for Pancreatic Cancer Research.

Supplementary Figures and Legends


a
Prime Editing
Protospacer
PAM
5'
3'
3'
5'
Edit position: +1
+4
PE range
b
Base Editing
Protospacer
PAM
5'
3'
3'
5'
+4
+8
High efficiency editing window

Search Method:

1) Mark the location of the start of all prime editing domains relative to the NGG PAM on plus and minus strand (two separate arrays).

+1
+4
G G
G G
+4
+1
1
1

2) For each mutation, using this array of marked domains, as well as the RT template length, determine whether the mutation can be created.

RT template ≥ 8 bp
1
1
RT template ≥ 11 bp
= Target edit
= NGG PAM

Search Method:

1) Mark the location of all NGG PAMs on plus and minus strand (two separate arrays).

2) For each transition SNP, using array of marked PAMs, determine whether the mutation falls within a high efficiency editing window and/or a protospacer by measuring distances to marked PAMs.

Protospacer
1
+4
+8
= Target edit
High efficiency editing window

CBE [C>T mutations (+ strand) > search + strand for PAMs
G>A mutations (+ strand) > search - strand for PAMs

ABE [A>G mutations (+ strand) > search + strand for PAMs
T>C mutations (+ strand) > search - strand for PAMs

Orthology Analysis

1) Generate list of human-to-mouse orthologs.

2) Perform global pairwise alignment of most common protein isoforms. Match score = 1; gap penalty = 0.

Create alignment index that can be referenced to find regions of homology.

3) For a given mutation, check if it falls in a coding region.

4) If coding, determine whether mutation falls in region of homology with designated flank size threshold (i.e. minimum number of matching amino acids adjacent to target)

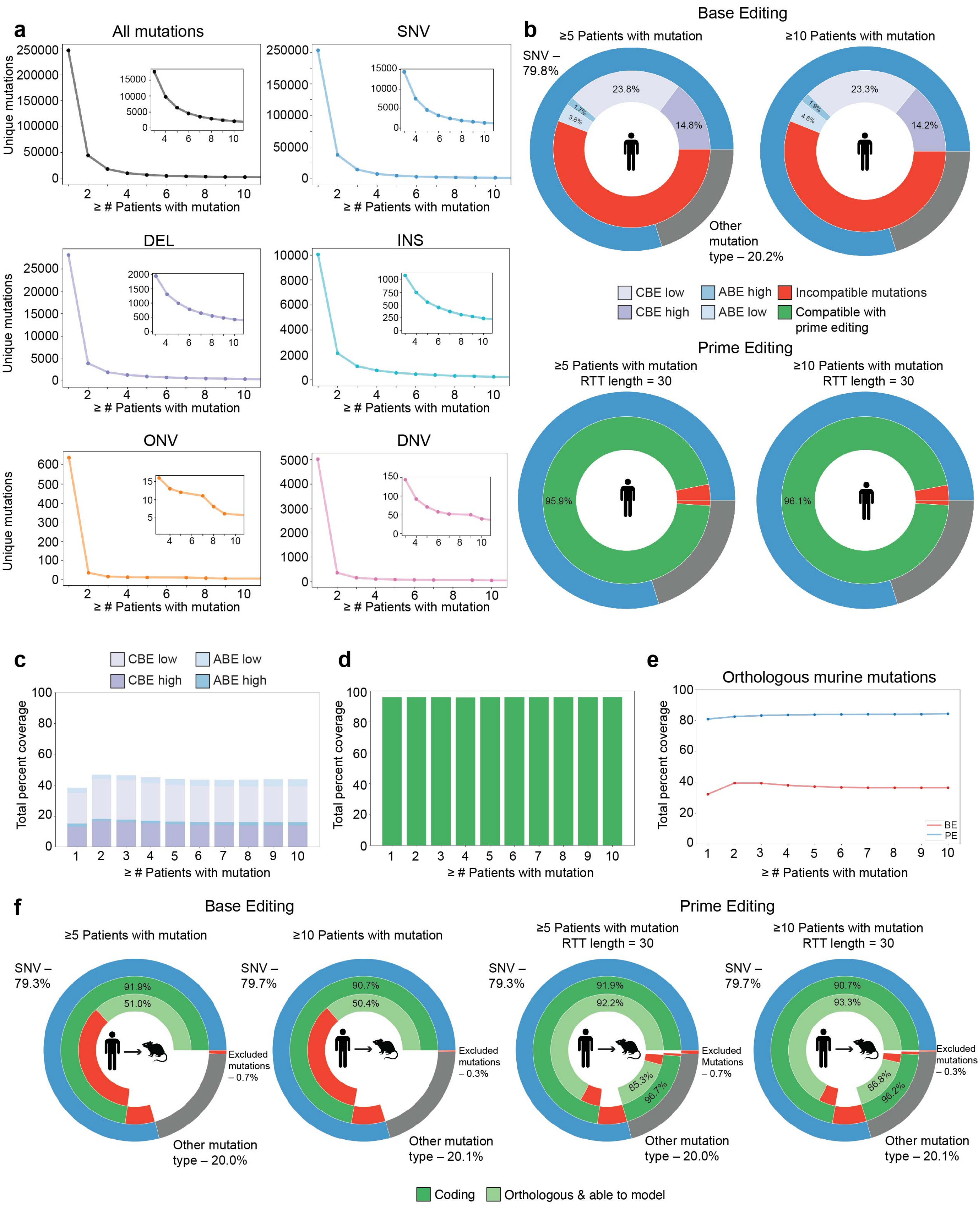
e.g. flank size = 2 (each box = amino acid)
= Target edit codon
= Flanking amino acids

Supplementary Figure 1. Computational methods for determining mutations accessible to base and prime editing and for evaluating conservation in murine protein sequences.

a. Method for determining whether a mutation is amenable to editing by a prime editor with an NGG PAM. Prime editing domains (the start sites of editing) are marked relative to the location of NGG PAMs. These domain markings are used, in combination with information about the RT template length, to determine if a mutation can be modeled.

b. Method for determining whether a mutation is amenable to editing by a base editor (cytosine or adenine) with an NGG PAM. The location of NGG PAMs is marked on the plus and minus strand, and for each transition SNV, it is determined based on distances to marked PAMs whether the mutation falls in a high efficiency editing window (positions +4 to +8 in a protospacer), within a protospacer, or outside any protospacer. Search rules for CBE and ABE are shown below.

c. Method for determining whether mutations fall in a conserved region between mouse and human orthologous protein sequences.


a
All mutations
SNV
DEL
INS
ONV
DNV
Unique mutations
≥ # Patients with mutation
b
Base Editing
≥5 Patients with mutation
≥10 Patients with mutation
SNV – 79.8%
23.8%
1.7%
3.8%
14.8%
23.3%
1.9%
4.6%
14.2%
Other mutation type – 20.2%
CBE low
CBE high
ABE high
ABE low
Incompatible mutations
Compatible with prime editing
Prime Editing
≥5 Patients with mutation RTT length = 30
≥10 Patients with mutation RTT length = 30
95.9%
96.1%
c
CBE low
ABE low
CBE high
ABE high
Total percent coverage
≥ # Patients with mutation
d
Total percent coverage
≥ # Patients with mutation
e
Orthologous murine mutations
Total percent coverage
BE
PE
≥ # Patients with mutation
f
Base Editing
Prime Editing
≥5 Patients with mutation
≥10 Patients with mutation
≥5 Patients with mutation RTT length = 30
≥10 Patients with mutation RTT length = 30
SNV – 79.3%
91.9%
51.0%
Excluded mutations – 0.7%
Other mutation type – 20.0%
SNV – 79.7%
90.7%
50.4%
Excluded mutations – 0.3%
Other mutation type – 20.1%
SNV – 79.3%
91.9%
92.2%
85.3%
96.7%
Excluded Mutations – 0.7%
Other mutation type – 20.0%
SNV – 79.7%
90.7%
93.3%
86.8%
96.2%
Excluded mutations – 0.3%
Other mutation type – 20.1%
Coding
Orthologous & able to model

Supplementary Figure 2. Prime editing has a greater capacity to model recurrent cancer-associated mutations than base editing.

a. Quantification of mutations that are detected in multiple patients in cancer-associated genes, depicted for each mutant variant type (Single nucleotide variants = SNV, deletions = DEL, insertions = INS, di-nucleotide variants = DNV, oligo-nucleotide variants = ONV). The y-axis in all plots indicates the total number of unique mutations per variant type.

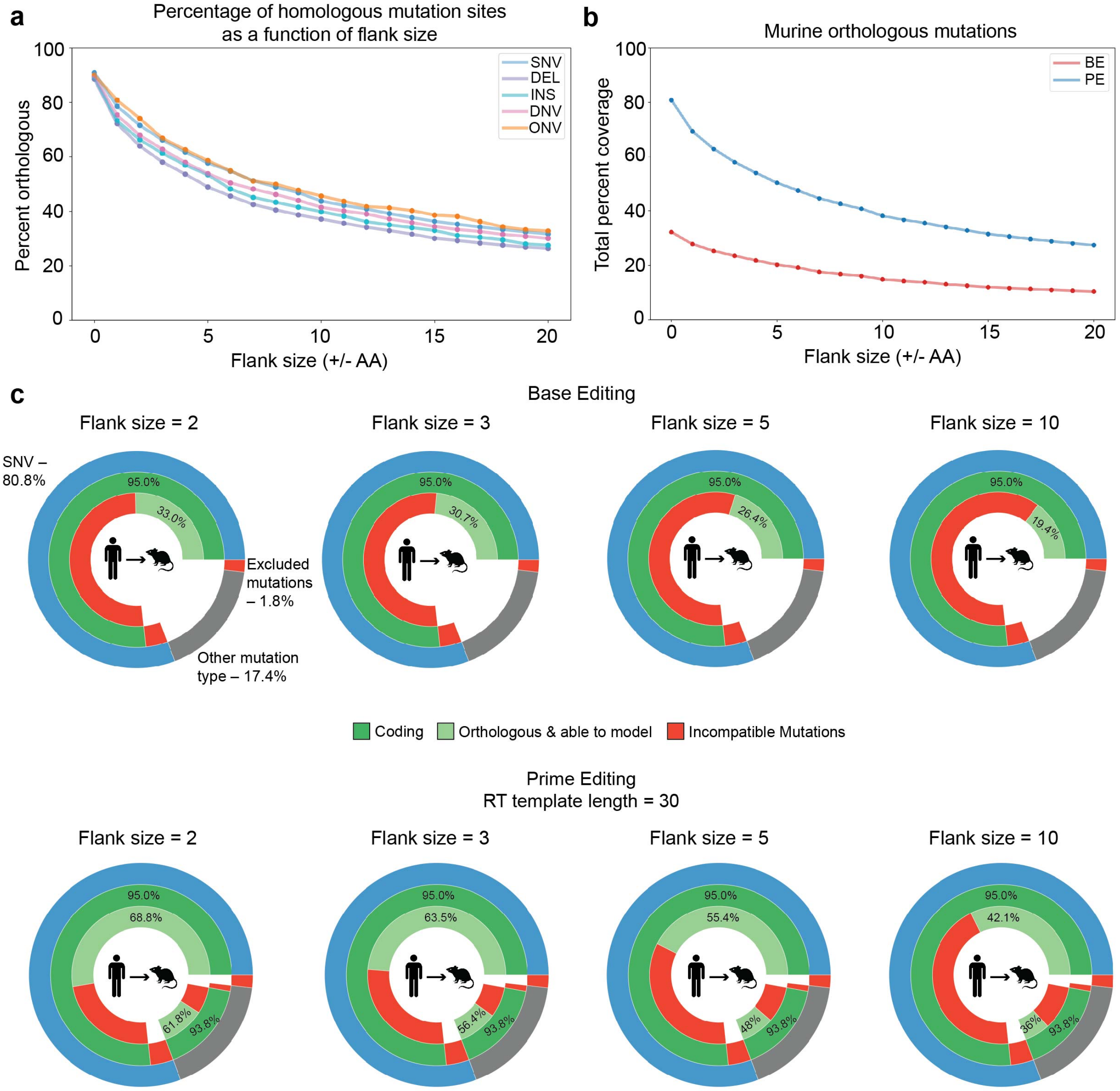
b. Quantification of recurrent mutations potentially amenable to modeling by a base editor with an NGG PAM (top) or a prime editor with an NGG PAM and a 30 base pair RT template (bottom). The columns show results considering mutations that occur in ≥5 patients (left) or ≥10 patients (right). "CBE/ABE high" indicates that the SNV falls in the high efficiency editing window (position +4 to +8 in the protospacer), while "CBE/ABE low" indicates the SNV falls within the protospacer but outside the high efficiency window. The data include SNVs (blue outer circle; 78.8% of mutations) and other mutation types (gray outer circle; 20.2% of mutations). All calculations assume a base or prime editor that recognizes only NGG PAMs.

c. Total percentage of recurrent mutations amenable to modeling by a base editor with an NGG PAM, quantified at multiple thresholds of mutation frequency, from mutations that occur ≥1 patient to those that occur ≥10 patients.

d. Total percentage of recurrent mutations amenable to modeling by a prime editor with an NGG PAM and a 30 base pair RT template at multiple thresholds of mutation frequency, from mutations that occur ≥1 patient to those that occur ≥10 patients.

e. Capabilities of prime versus base editing to model orthologous mutations in mice at multiple thresholds of mutation frequency, from ≥1 to ≥10 patients with each mutation. Prime editing can model approximately double the number of orthologous mutations in mice at all thresholds of mutation frequency. Base editing (BE) shown in red. Prime editing (PE) shown in blue.

f. Quantification of the ability of base editing (left) and prime editing (right) to model orthologous mutations in mice for mutations that occur in ≥5 or ≥ 10 patients.

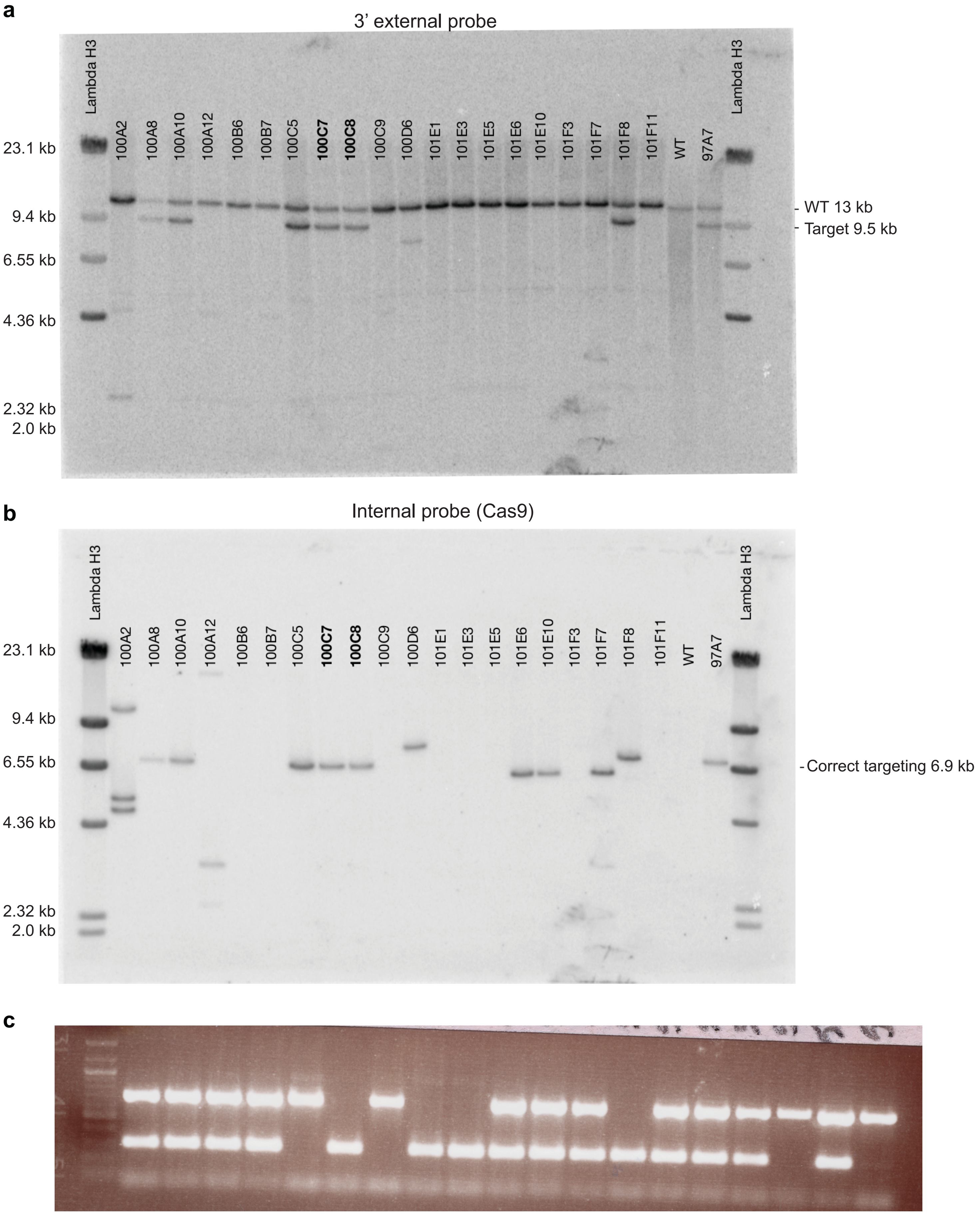


Supplementary Figure 3. Prime editing enables modeling a broader scope of cancer-associated mutations from residues conserved in mice at various homology stringencies.

a. The percentage of mutations, categorized by variant type, that fall in a region of homology as a function of flank size. Flank size is defined as the number of amino acids on either side of the mutant codon that must match between the human and mouse orthologs for the mutated codon to be considered orthologous (e.g., a flank size of two would mean that a total of five amino acids–two upstream and two downstream of the mutant codon–would need to match in the human and mouse protein to be considered orthologous.)

b. Capabilities of prime editing (blue) and base editing (red) to model mutations that derive from a wild-type amino acid residue conserved in mice as a function of conserved flank size (i.e., the stringency of homology). All calculations assume a base or prime editor that recognizes only NGG PAMs.

c. Quantification of the ability of base editing (top) and prime editing (bottom) to model orthologous mutations in mice at different flank size values (2, 3, 5, and 10 from left to right). These plots correspond with the data points in panel (b).

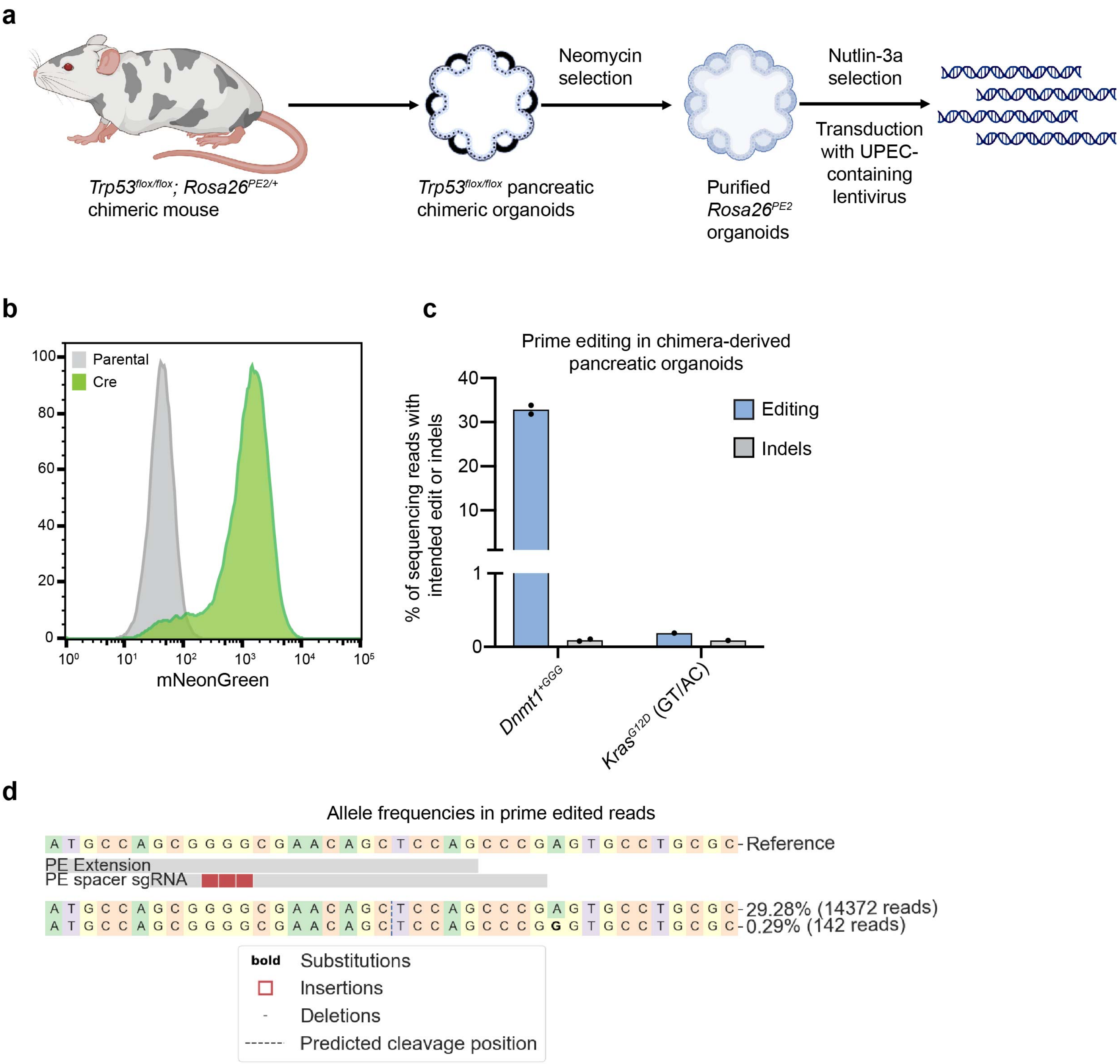

a
3’ external probe
Lambda H3
100A2
100A8
100A10
100A12
100B6
100B7
100C5
100C7
100C8
100C9
100D6
101E1
101E3
101E5
101E6
101E10
101F3
101F7
101F8
101F11
WT
97A7
Lambda H3
23.1 kb
9.4 kb
6.55 kb
4.36 kb
2.32 kb
2.0 kb
WT 13 kb
Target 9.5 kb
b
Internal probe (Cas9)
Lambda H3
100A2
100A8
100A10
100A12
100B6
100B7
100C5
100C7
100C8
100C9
100D6
101E1
101E3
101E5
101E6
101E10
101F3
101F7
101F8
101F11
WT
97A7
Lambda H3
23.1 kb
9.4 kb
6.55 kb
4.36 kb
2.32 kb
2.0 kb
Correct targeting 6.9 kb
c

Supplementary Figure 4. Genetic validation of *Rosa26* targeting.

a. Southern blots depicting the external probe used to evaluate *Rosa26* targeting.

b. Southern blots depicting the internal probe used to evaluate *Rosa26* targeting. Successfully targeted 100C7 and 100C8 ESC clones were transmitted through the germline.

c. Gel image of a genotyping PCR of the *Rosa26*^{PE2} allele. The larger band (~717 base pairs) reflects the presence of the *Rosa26*^{PE2} allele, while the smaller band indicates a wild-type sequence. See Supplementary Table 3 for the primer sequences used for genotyping.



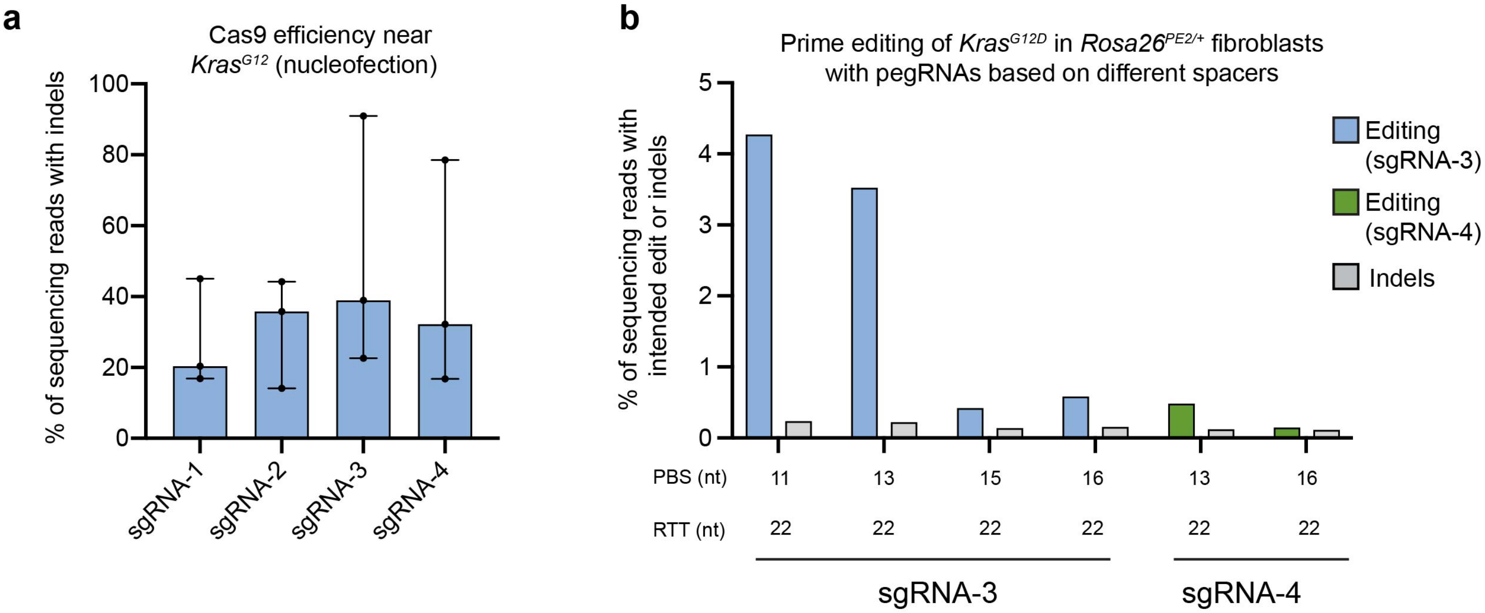
Supplementary Figure 5. Functional validation of the *Rosa26*^{PE2} allele in pancreatic organoids derived from chimeric mice.

a. Schematic depicting the experimental workflow. Cells containing the *Rosa26*^{PE2} allele were selected using neomycin treatment. Purified *Trp53*$^{flox/flox}$;*Rosa26*$^{PE2/+}$ organoids were then treated with nutlin-3a after lentiviral infection to ensure selection of successfully transduced cells.

b. mNG expression in purified pancreatic organoids quantified by the Guava easyCyte flow cytometer. Untreated cells ("parental") are depicted in gray and cells treated with the UPEC vector encoding Cre recombinase ("Cre") are depicted in green.

c. Editing efficiency and indel byproduct frequency of pegRNAs targeting *Dnmt1*$^{+GGG}$ (*n* = 2) and *Kras*^{G12D} (*n* = 1). The *Kras*-targeted pegRNA mediates a dinucleotide substitution encoding *Kras*^{G12D} at low efficiency.

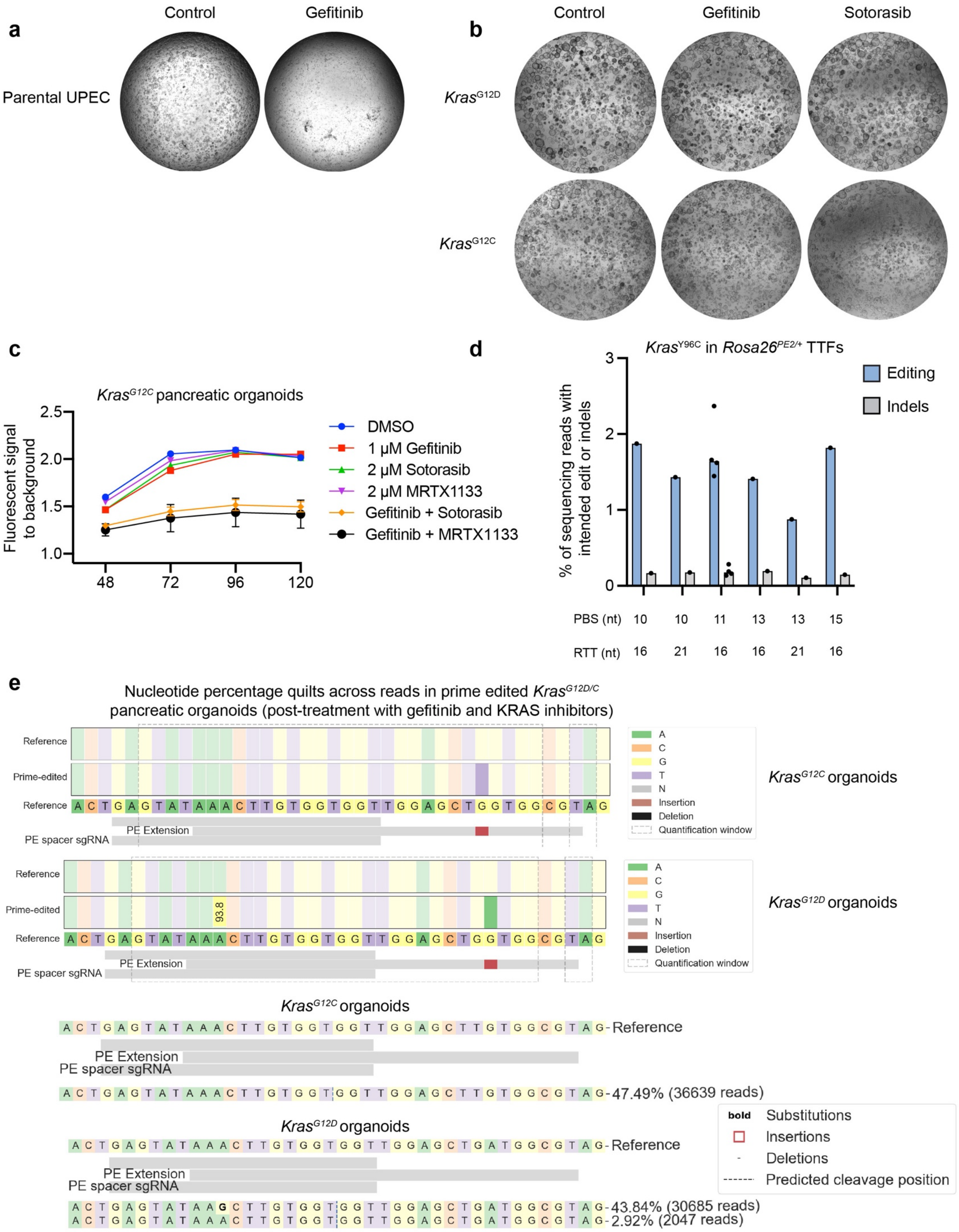
d. Allele frequency table output by CRISPResso2 depicting the percentages of reads encoding specific sequences. The three red bars represent the position of the encoded trinucleotide insertion. 29.28% of all reads contain the prime edited insertion. Note that "reference" refers to the DNA sequence including the intended prime edit without modification.



Supplementary Figure 6. Selection of an optimal spacer for pegRNAs templating *Kras*^{G12} mutations.

a. Indel frequencies associated with Cas9 activity from four single guide RNAs (sgRNAs) targeted near the *Kras*^{G12} locus in mouse N2A cells (*n* = 3 biological replicates). See Supplementary Table 2 for guide RNA sequences.

b. Editing efficiency of representative pegRNAs based on spacers corresponding to sgRNAs-3 and -4 (*n* = 1 replicate per pegRNA). The data for the four leftmost pegRNAs based on sgRNA-3 are also presented in Fig. 3a. pegRNAs based on sgRNA-3 generally exhibited higher prime editing efficiency than pegRNAs based on sgRNA-4.



a
Control
Gefitinib
Parental UPEC
b
Control
Gefitinib
Sotorasib
Kras^{G12D}
Kras^{G12C}
c
Kras^{G12C} pancreatic organoids
Fluorescent signal to background
2.5
2.0
1.5
1.0
48
72
96
120
DMSO
1 μM Gefitinib
2 μM Sotorasib
2 μM MRTX1133
Gefitinib + Sotorasib
Gefitinib + MRTX1133
d
Kras^{Y96C} in Rosa26$^{PE2/+}$ TTFs
% of sequencing reads with intended edit or indels
3
2
1
0
Editing
Indels
PBS (nt) 10 10 11 13 13 15
RTT (nt) 16 21 16 16 21 16
e
Nucleotide percentage quilts across reads in prime edited Kras$^{G12D/C}$ pancreatic organoids (post-treatment with gefitinib and KRAS inhibitors)
Reference
Prime-edited
Reference A C T G A G T A T A A A C T T G T G G T G G T T G G A G C T G G T G G C G T A G
PE Extension
PE spacer sgRNA
A
C
G
T
N
Insertion
Deletion
Quantification window
Kras^{G12C} organoids
Reference
Prime-edited
93.8
Reference A C T G A G T A T A A A C T T G T G G T G G T T G G A G C T G G T G G C G T A G
PE Extension
PE spacer sgRNA
A
C
G
T
N
Insertion
Deletion
Quantification window
Kras^{G12D} organoids
Kras^{G12C} organoids
A C T G A G T A T A A A C T T G T G G T G G T T G G A G C T T G T G G C G T A G - Reference
PE Extension
PE spacer sgRNA
A C T G A G T A T A A A C T T G T G G T G G T T G G A G C T T G T G G C G T A G - 47.49% (36639 reads)
Kras^{G12D} organoids
A C T G A G T A T A A A C T T G T G G T G G T T G G A G C T G A T G G C G T A G - Reference
PE Extension
PE spacer sgRNA
A C T G A G T A T A A **G** C T T G T G G T G G T T G G A G C T G A T G G C G T A G - 43.84% (30685 reads)
A C T G A G T A T A A A C T T G T G G T G G T T G G A G C T G A T G G C G T A G - 2.92% (2047 reads)
bold Substitutions
Insertions
- Deletions
------- Predicted cleavage position

Supplementary Figure 7. Treatment of prime edited *Kras*^{G12D} and *Kras*^{G12C} pancreatic organoids with EGFR and KRAS inhibitors in the presence or absence of a secondary resistance mutation.

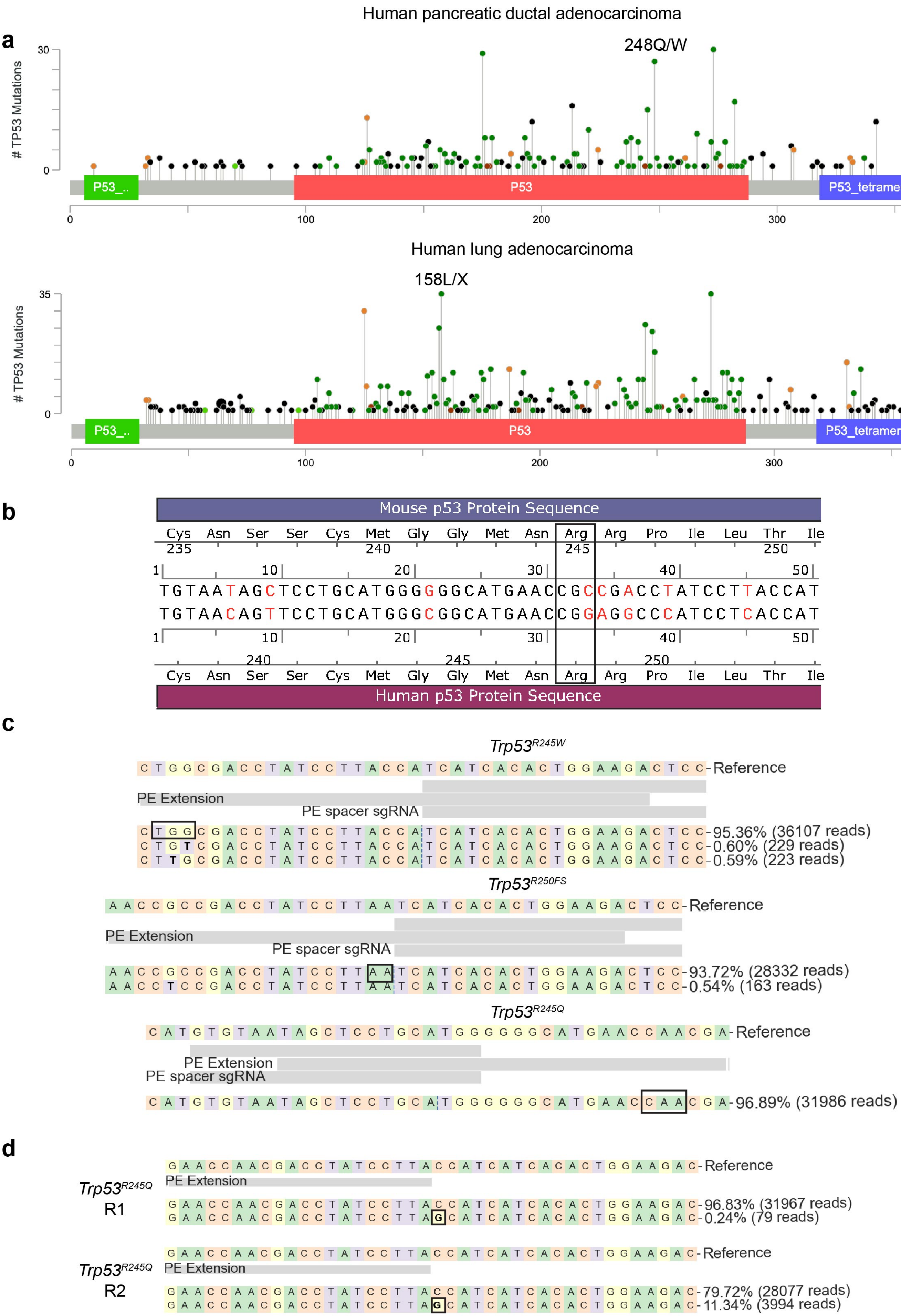
a. Bright-field images of unedited *Rosa26*$^{PE2/+}$ pancreatic organoids treated with gefitinib after infection with the template UPEC vector (i.e., lacking a pegRNA).

b. Additional bright-field images supplementing Fig. 3d and depicting prime edited *Kras*^{G12D} and *Kras*^{G12C} organoids treated either with DMSO, gefitinib, or sotorasib.

c. Viability of *Kras*^{G12C} organoids under various treatment conditions quantified using the alamarBlue HS Cell Viability Reagent.

d. Editing efficiency and indel frequencies of epegRNAs templating the *Kras*^{Y96C} transversion in *Rosa26*$^{PE2/+}$ TTFs (n = 1-4 biological replicates per pegRNA).

e. Nucleotide percentage quilts and allele frequency tables output by CRISPResso2 depicting sequencing reads derived from prime edited *Kras*^{G12C} and *Kras*^{G12D} pancreatic organoids. Prime edited alleles were previously selected with gefitinib, and these sequencing data derive from organoids after treatment with gefitinib in combination with mutant KRAS inhibitors.


a
Human pancreatic ductal adenocarcinoma
248Q/W
# TP53 Mutations
30
0
P53_..
P53
P53_tetramer
0
100
200
300
393aa
Human lung adenocarcinoma
158L/X
# TP53 Mutations
35
0
P53_..
P53
P53_tetramer
0
100
200
300
393aa
b
Mouse p53 Protein Sequence
Cys Asn Ser Ser Cys Met Gly Gly Met Asn Arg Arg Pro Ile Leu Thr Ile
235 240 245 250
1 10 20 30 40 50
TGTAATAGCTCCTGCATGGGGGGCATGAACCGCCGACCTATCCTTACCAT
TGTAACAGTTCCTGCATGGGCGGCATGAACCGGAGGCCCATCCTCACCAT
1 10 20 30 40 50
240 245 250
Cys Asn Ser Ser Cys Met Gly Gly Met Asn Arg Arg Pro Ile Leu Thr Ile
Human p53 Protein Sequence
c
Trp53R245W
CTGGCGACCTATCCTTACCATCATCACACTGGAAGACTCC - Reference
PE Extension
PE spacer sgRNA
CTGGCGACCTATCCTTACCATCATCACACTGGAAGACTCC - 95.36% (36107 reads)
CTGTCGACCTATCCTTACCATCATCACACTGGAAGACTCC - 0.60% (229 reads)
CTTGCGACCTATCCTTACCATCATCACACTGGAAGACTCC - 0.59% (223 reads)
Trp53R250FS
AACCGCCGACCTATCCTTAATCATCACACTGGAAGACTCC - Reference
PE Extension
PE spacer sgRNA
AACCGCCGACCTATCCTTAATCATCACACTGGAAGACTCC - 93.72% (28332 reads)
AACCTCCGACCTATCCTTAATCATCACACTGGAAGACTCC - 0.54% (163 reads)
Trp53R245Q
CATGTGTAATAGCTCCTGCATGGGGGGCATGAACCAACGA - Reference
PE Extension
PE spacer sgRNA
CATGTGTAATAGCTCCTGCATGGGGGGCATGAACCAACGA - 96.89% (31986 reads)
d
Trp53R245Q
R1
GAACCAACGACCTATCCTTACCATCATCACACTGGAAGAC - Reference
PE Extension
GAACCAACGACCTATCCTTACCATCATCACACTGGAAGAC - 96.83% (31967 reads)
GAACCAACGACCTATCCTTAGCATCATCACACTGGAAGAC - 0.24% (79 reads)
Trp53R245Q
R2
GAACCAACGACCTATCCTTACCATCATCACACTGGAAGAC - Reference
PE Extension
GAACCAACGACCTATCCTTACCATCATCACACTGGAAGAC - 79.72% (28077 reads)
GAACCAACGACCTATCCTTAGCATCATCACACTGGAAGAC - 11.34% (3994 reads)

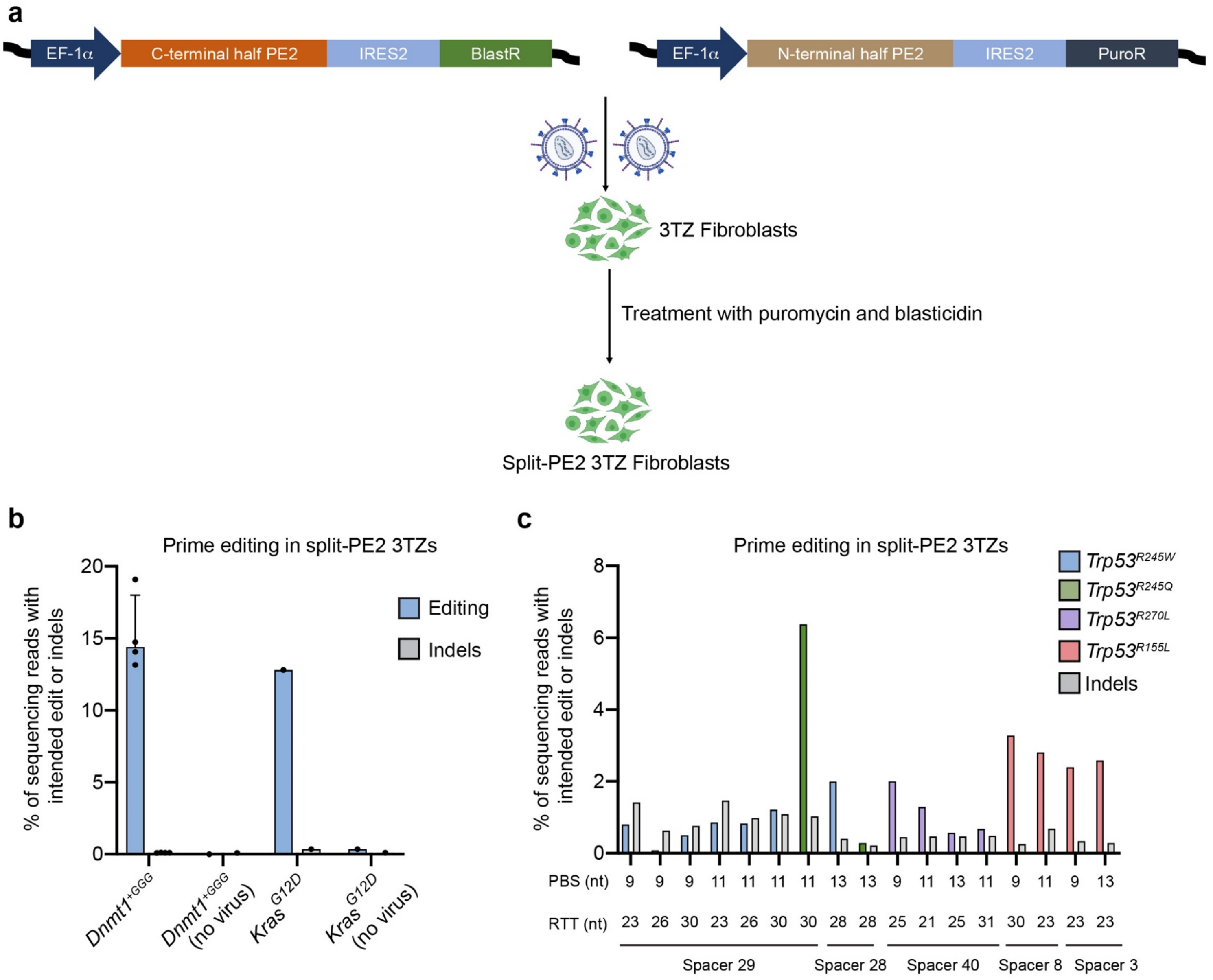
Supplementary Figure 8. Frequency distribution of *TP53* mutations in lung and pancreatic cancer patients and prime editing *Trp53* mutations in pancreatic organoids.

a. Lolliplot frequency distributions retrieved from cBioPortal as described in the methods section for pancreatic cancer patients (top) and lung cancer patients (bottom).

b. Multiple sequence alignment of murine (*Trp53*) and human (*TP53*) DNA and protein sequences surrounding the *Trp53*^{R245} locus. Substitutions are highlighted in red font. A single nucleotide substitution between these orthologs creates a requirement for a dinucleotide substitution to model *Trp53*^{R245W} in mice.

c. Allele frequency tables output by CRISPResso2 depicting the sequencing reads derived from *Trp53*$^{flox/+}$;*Rosa26*^{PE2} pancreatic organoids after selection with nutlin. These are derived from the same sequencing data underlying the plots described in Fig. 3h. Edited codons for *Trp53*^{R245Q} and *Trp53*^{R245W} are indicated within a black box. *Trp53*^{R250FS} involves a dinucleotide deletion of ACCA -> AA.

d. Allele frequency tables depicting the region immediately after the 3' extension sequence. The plots are derived from two biological replicates (R1 and R2) of *Trp53*$^{flox/+}$;*Rosa26*^{PE2} organoid lines infected with UPEC-*Trp53*^{R245Q} and treated with nutlin. The top plot is derived from the same data for *Trp53*^{R245Q} depicted in (c). A black box indicates an unintended nucleotide modification occurring at variable efficiency immediately after the 3' extension sequence.



Supplementary Figure 9. Construction of the split-PE2 3TZ fibroblast line for testing epegRNAs targeted to *Trp53*.

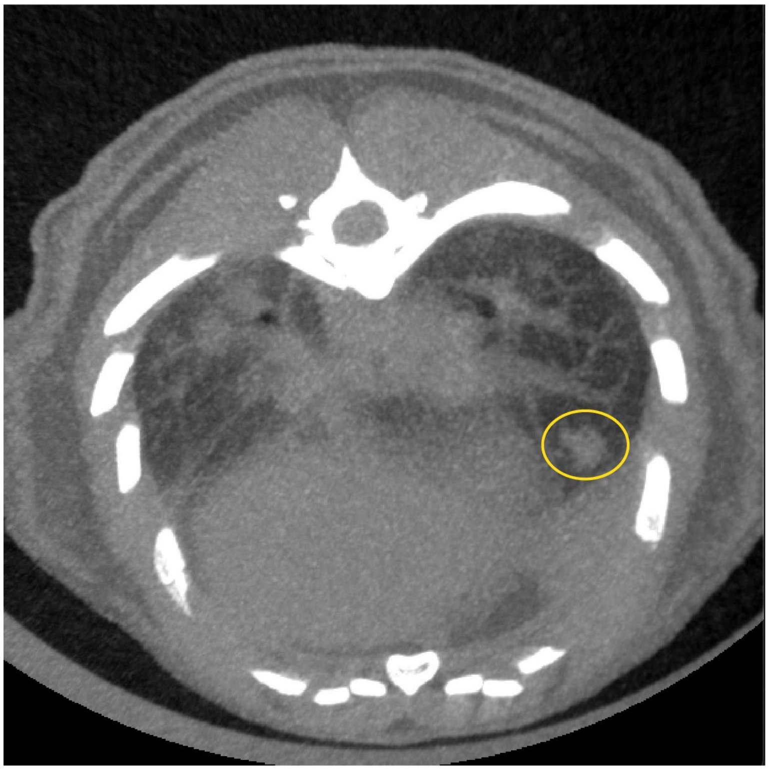
a. Schematic depicting the installation of split halves of the PE2 enzyme using lentivirus and a combination of puromycin and blasticidin resistance genes.

b. Editing efficiency of a pegRNA targeted to *Dnmt1* (*n* = 4 biological replicates) and an epegRNA (the same depicted in Fig. 3b) targeted to *Kras* (*n* = 1) in split-PE2 3TZ cells confirms their ability to mediate prime editing. Estimates of $Dnmt1^{+GGG}$ and $Kras^{G12D}$ frequency in cells not transduced with virus are included as controls (*n* = 1).

c. Editing efficiency and indel byproduct frequency of 3-4 representative epegRNAs tested for each of the *Trp53* mutations described in the Fig. 3, as well as two *Trp53* mutations ($Trp53^{R155L}$ and $Trp53^{R270L}$) homologous to those commonly observed in human lung cancer. epegRNA and spacer sequences are listed in Supplementary Table 2.

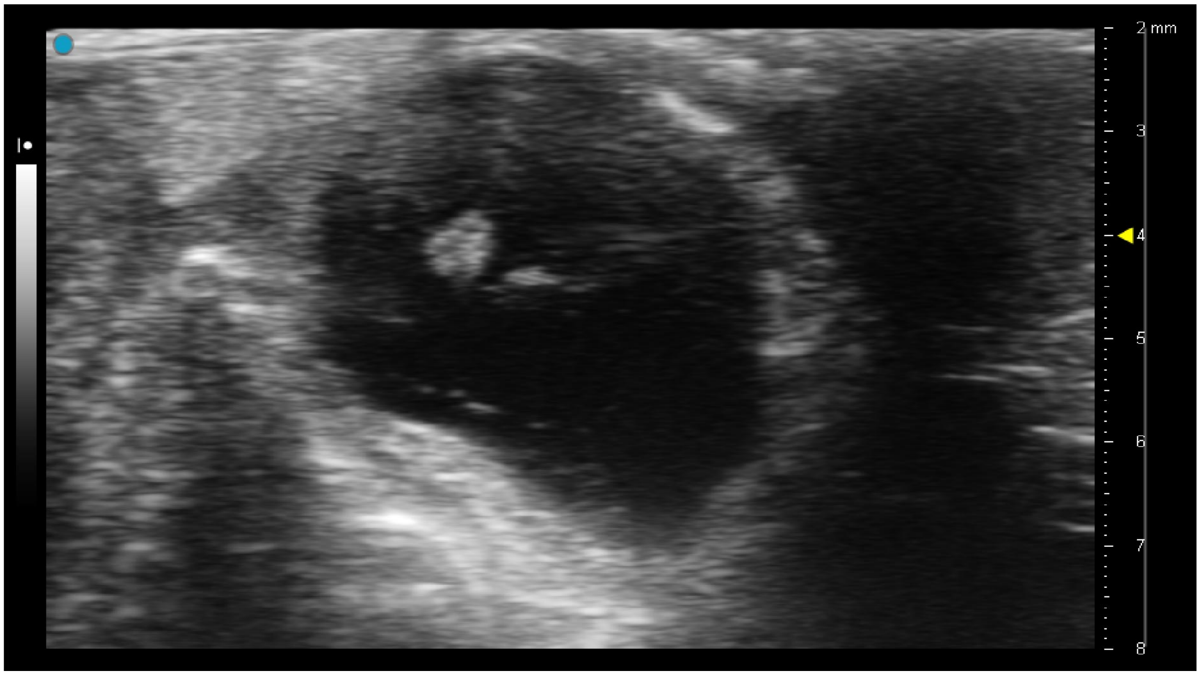
a

Visible lung tumor by micro-CT
(14 weeks post-infection)



b

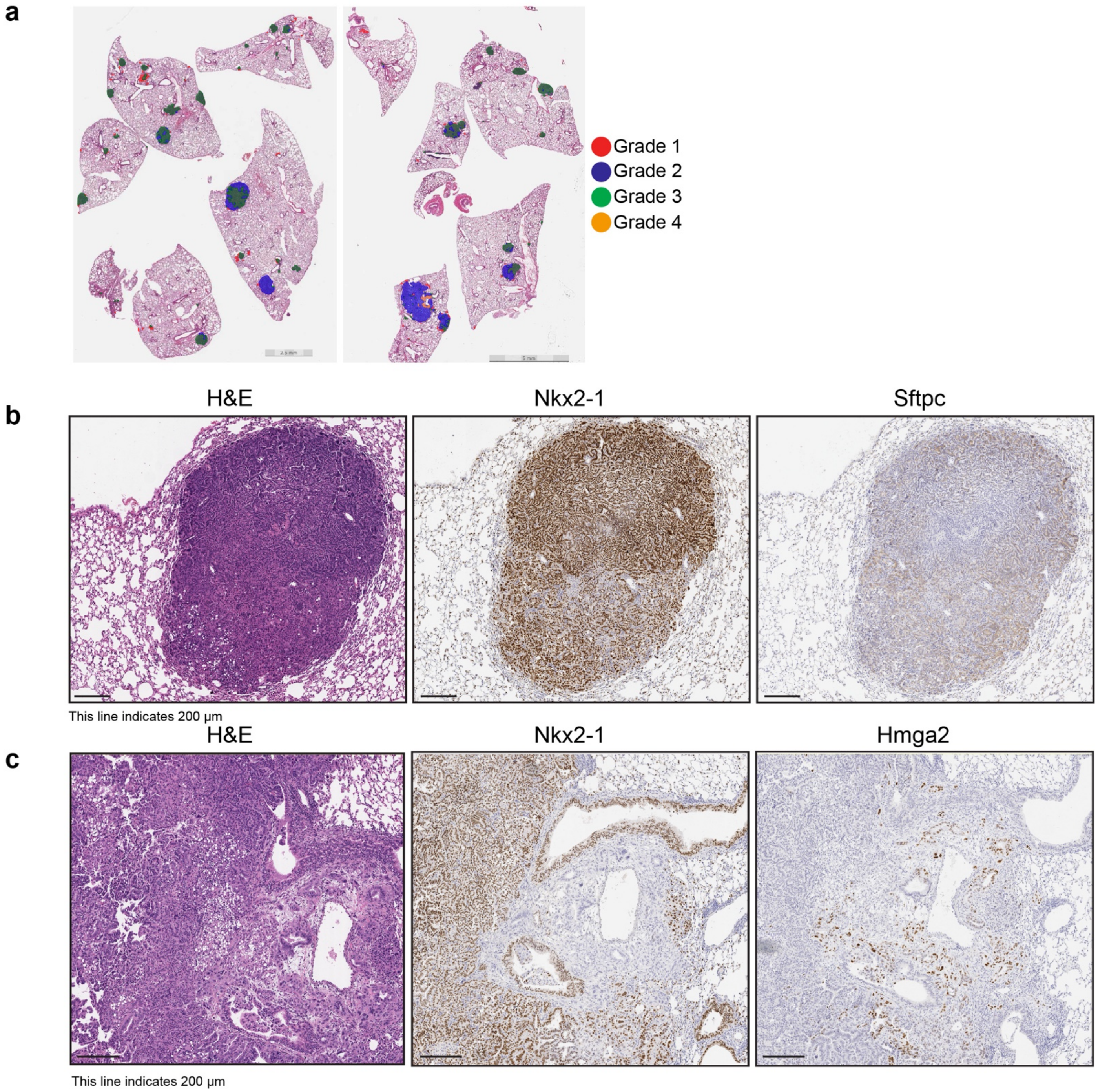
Visible pancreatic tumor by ultrasound
(6 weeks post-transplantation)



Supplementary Figure 10. Micro-CT and ultrasound imaging of lung and pancreatic tumors *in vivo*.

a. Micro-CT of the lung in a *Trp53*$^{flox/flox}$*;Rosa26*$^{PE2/+}$ mouse 14 weeks after infection with lentiviral vectors encoding UPEC-*Kras*^{G12D} (using the epegRNA described in Fig. 3b). A visible tumor is highlighted by a yellow circle.

b. Ultrasound imaging depicting a pancreatic tumor derived from orthotopic transplantation of *Trp53*$^{flox/flox}$*;Rosa26*$^{PE2/+}$ organoids, which were prime edited using the *Kras*^{G12D} epegRNA. The *Kras*^{G12D} allele was selected with gefitinib *in vitro* prior to transplantation, as shown in Fig. 3c. All organoids were transplanted into *Rosa26*^{PE2} recipients.


a
Grade 1
Grade 2
Grade 3
Grade 4
b
H&E
Nkx2-1
Sftpc
This line indicates 200 µm
c
H&E
Nkx2-1
Hmga2
This line indicates 200 µm

Supplementary Figure 11. Prime edited tumors recapitulate histological features of gold standard GEMMs of lung adenocarcinoma.

a. Aiforia grade analysis of 16-week tumor-bearing lungs initiated by *in vivo* prime editing (red = Grade 1, green = Grade 2, blue = Grade 3, orange = Grade 4).

b. H&E and immunohistochemical (IHC) staining of a representative Grade 2 tumor 16 weeks post-initiation. IHC staining was performed with antibodies specific to Nkx2.1, a lung lineage transcription factor, or Surfactant Protein C (Sftpc), a marker of AT2 cells. Scale bars: 200 µm.

c. H&E and IHC staining of a Grade 3/4 tumor 16 weeks post-initiation. IHC staining was performed with antibodies specific to Nkx2-1 or HMGA2, a lung embryonic chromatin regulator. Scale bars: 200 µm.

Supplementary Tables are available upon request.

Methods

Bioinformatic analysis of prime and base editor capabilities for modeling cancer-associated mutations.

We constructed a Python-based computational pipeline to compare the abilities of prime and base editors to model cancer-associated mutations. Data were retrieved from MSK-IMPACT datasets previously described[31]. The pipeline, all related scripts, and intermediate data needed to reproduce our results are available at **https://github.com/samgould2/prime-vs-base-editing**.

Analysis of recurrent cancer mutations incorporated in published transgenic mouse models.

To estimate the fraction of frequent cancer driver mutations captured by currently available transgenic mouse models, we used the MouseMine tool from the Mouse Genome Informatics database[38,39] and obtained a comprehensive list of published transgenic alleles. We initiated our search using the mammalian phenotype code MP:0002006 (“neoplasm”) to retrieve all mouse models related to the study of cancer. We then modified the search by adding the following parameters: “Allele Type,” “Mutations (Name)”, “Alleles (Name and Molecular Note and Attribute string),” and “Subjects (synonyms → names).” We then filtered the results to retain only allele types annotated as “Targeted,” “Transgenic,” or “Endonuclease-mediated.”

After exporting these data (Supplementary Table 2), we then identified the 100 most frequent single nucleotide variants present in the MSK-IMPACT dataset. We then manually cross-referenced these two lists using standard Bash commands (e.g., grep to isolate relevant genes and codons) to identify available models representing specific mutations. In cases where models were absent in the MouseMine list, we performed a manual literature search to confirm an absence of models in the published literature. Using this approach, we designated for each mutation whether 1) any transgenic allele exists that can be used to model cancer in mice and 2) whether any existing models enable selective expression in a tissue of interest (e.g., through Cre recombinase-induced removal of a LoxP-STOP-LoxP cassette). The manually annotated results for the 100 most frequent mutations, as well as the entire MouseMine list are included as Supplementary Tables 1 and 2.

Design and cloning of the Cre-inducible prime editor allele.

The PE2-P2A-mNG *Rosa26* targeting vector was generated with a backbone formed via BstBI and AscI restriction enzyme digestion of the *Sp*Cas9-NLS-P2A-EGFP *Rosa26* targeting vector[10,41]. A fragment encoding the PE2 enzyme was generated by PCR amplification from the pCMV-PE2 plasmid obtained from Addgene[21] (Addgene no. 132775), and a fragment containing the P2A-mNG sequence was amplified from a lentivirus plasmid encoding Cre-P2A-mNG. Two additional fragments containing WPRE-pA-PGK (Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element-poly(A)-PGK promoter) and a neomycin resistance gene (NeoR-pA) were PCR-amplified from the *Sp*Cas9-NLS-P2A-EGFP vector. A FRT3 site was installed by incorporating overlapping portions of this motif into the PCR primers. All primers used are listed and described in Supplementary Table 4. A 5-part Gibson assembly reaction generated the final targeting vector using these components[75]. Successful assembly was confirmed with diagnostic restriction enzyme digestions and Sanger sequencing. The plasmid encoding the PE2-P2A-mNG *Rosa26* targeting vector will be deposited at Addgene for distribution.

Embryonic stem cell targeting, validation, and chimera generation.

KP4*, a C57BL/6J *Trp53*$^{flox/flox}$ (P) murine embryonic stem (ES) cell line, was generated by crossing a hormone primed C57Bl6J *Trp53*$^{flox/flox}$ female with a C57Bl6J *Kras*$^{LSL\text{-}G12D}$;*Trp53*$^{flox/flox}$ male. At 3.5 days post coitum, blastocysts were flushed from the uterus, isolated, and cultured individually on a mouse embryonic fibroblast (MEF) feeder layer in ESCM+LIF+2i (Knockout DMEM (GIBCO), 15% FBS (Hyclone), 1% NEAA (Sigma), 2mM Glutamine (GIBCO), 0.1mM b-mercaptoethanol (SigmaAldrich) 50 IU Penicillin, 50 IU Streptomycin, 1000U/ml LIF (Amsbio), 3µM CHIR99021 (Abmole), and 1µM PD0325901 (Abmole)). After 5-7 days in culture, the outgrown inner cell mass was isolated, trypsinized and replated on a fresh MEF layer. ES cell lines were genotyped for *Kras*$^{LSL\text{-}G12D}$ (KP4* was wildtype +/+ for *Kras*), *Trp53*$^{flox/flox}$, and Zfy (Y-chromosome specific). Primer sequences are available upon request. ES cell lines were tested for pluripotency by injection into host blastocysts from albino mice to generate chimeric mice.

Briefly, 36 µg of the prime editor targeting vector (R26–CAGG-LoxStopLox-Cas9$^{(H840A)}$-MMLVRT-P2A-mNeonGreen-WPRE-bHGpA; PGK-Neo-PGKpA) was linearized with PvuI, phenol/$CHCl_3$ extracted, and then was ethanol precipitated. After resuspending the DNA in 150µl of PBS it was mixed with 3 x 10^6 KP4* ES cells in 650 µl of PBS in a 4mm electroporation cuvette. The cell/DNA mixture was pulsed once in a Biorad Genepulsar 2 (600V and 25µF) followed by re-plating of the cells on irradiated mouse embryonic fibroblasts (MEFs). After 48hrs, the ES cell cultures were placed under selection with Geneticin (GIBCO) at 350µg/mL. When colonies were large enough to pick, 45 colonies were manually picked using a stereomicroscope. Each clone was expanded and evaluated for correct integration by PCR with primers spanning the 5' homology arm. Eleven PCR–positive clones were further evaluated using southern blot analysis. Briefly, genomic DNA was digested with EcoRV-HF (NEB) overnight. Digestions were electrophoresed on 0.7% agarose gels and blotted to Amersham Hybond XL nylon membranes (GE Healthcare). Samples were probed with ^{32}P-labeled "Rosa26 3' "external" and Cas9 "internal" probes applied in Church buffer (probe sequences available on request). Five clones (100A8, 100A10, 100C5, 100C7, and 100C8) showed correct integration. Correctly targeted clones verified by both PCR and southern blot analysis were injected into albino C57BL/6J blastocysts. High degree chimeras (visually assessed by coat color percentage) from the 100C7 and 100C8 ES cell clones successfully transmitted the prime editor construct through the germline.

Nucleofection of Neuro-2a cells and genomic DNA preparation

To evaluate spacers near the genetic locus encoding G12 in *Kras*, Neuro-2a cells were nucleofected using the SF Cell Line 4D-Nucleofector X Kit (Lonza) with 2 x 10^5 cells per sample (program DS-137). 800 ng of SpCas9-expressing plasmid and 200 ng of single guide RNA(sgRNA)-expressing plasmid were used according to the manufacturer's protocol. Following nucleofection, samples were brought to 100 µL total volume with DMEM supplemented with GlutaMax (Thermo Fisher Scientific) and 10% (vol/vol) FBS (Gibco, qualified) at 37 °C. After incubation for 10 minutes at room temperature, the cells were seeded on 48-well poly-D-lysine coated plates (Corning). 3 days following nucleofection, the cells were washed with PBS after removing the media and then lysed by addition of 150 µL of freshly prepared lysis buffer (10 mM Tris-HCl, pH 8 at 23 °C; 0.05% SDS; 25 µg ml^{-1} of proteinase K (Qiagen)). Lysates were incubated at 37 °C for one hour and then heat-inactivated for 30 min at

80 °C. The *Kras* amplicon was amplified from the genomic DNA samples, sequenced on an Illumina MiSeq, and analyzed with CRISPResso2[76] for indel quantification as previously described[33]. Primers used for amplification of the *Kras* amplicon are listed in Supplementary Table 4.

pegRNA design and cloning.

pegRNAs were designed in part using the pegRNA design tool, Prime Design[77]. In some cases (e.g., editing at *Kras*^{G12D}), CRISPR sgRNAs were tested prior to pegRNA design to select spacers that exhibited the highest level of Cas9 activity. For some designs (e.g., all *Trp53*-targeted pegRNAs), the trimmed evopreQ$_1$ motif was included to form epegRNAs and optimize editing efficiency within a limited cohort of initial candidates[49]. pegRNAs and their sequences are provided in Supplementary Table 3. Generally, the editing efficiency of all *Trp53*-targeted pegRNAs was tested first in single replicates in split-PE2 *Trp53*$^{+/+}$ 3TZ cells, while *Kras*-targeted and other pegRNAs were tested in *Rosa26*$^{PE2/+}$ tail tip-derived fibroblasts. pegRNAs with the highest editing efficiencies were selected for later assays in organoids and mice.

All pegRNAs were tested within the context of hU6-pegRNA-EF-1α-Cre (UPEC) or hU6-RFP/pegRNA-EFS-mScarlet (UPEmS) vectors. All pegRNA-expressing vectors were assembled via Golden Gate Assembly[78] using the uncut template plasmid and three annealed oligo pairs consisting of the spacer sequence, the scaffold, and the 3’ extension, all with compatible overhangs. Assembly was facilitated using the Golden Gate Assembly Kit (BsmBI-v2) from New England BioLabs. Plasmid DNA was extracted from transformed bacterial cultures using QIAprep Miniprep or Midiprep kits according to the manufacturer’s instructions.

The UPEmS template vector was generated via Gibson assembly of three insert fragments and a linearized backbone. Two fragments were formed by PCR amplification from the “pU6 pegRNA

GG acceptor" plasmid (Addgene plasmid no.132777)[21]. Specifically, the hU6 promoter was amplified using primers modified to install a BsmBI recognition site and the pAF Gibson adapter sequence on either side of the promoter (pAF-hU6-BsmBI), and the red fluorescent protein (RFP) component was also amplified in part using a primer that installed another BsmBI recognition site (forming BsmBI-RFP-BsmBI-pAR/gBF). A third fragment, gAR/pBF-EFS-mScarlet-gBR, was amplified from a separate lentiviral plasmid containing U6-sgRNA-EFS-mScarlet. All fragments were designed to contain compatible overhangs for Gibson assembly. All vectors with detailed maps and sequences will be deposited into Addgene. All primer names and sequences are listed in Supplementary Table 4.

The UPEC template plasmid (hU6-RFP-EF-1α-Cre; hU6-pegRNA-EF-1α-Cre after cloning) was developed by Gibson assembly of two insert fragments and the same backbone used to clone pUPEmS. The pBF-EF-1alpha-Cre-gBR fragment was generated using pBF and gBR PCR primers targeting the pUSEC (U6-sgRNA-EF-1alpha-Cre) vector[75,79]. The pAF-U6-RFP-gAR fragment was amplified from the UPEmS vector. The final UPEC vector enables concomitant assembly of the pegRNA spacer, scaffold, and 3' extension. Plasmids encoding the template UPEC and UPEmS vectors used for Golden Gate cloning will be deposited at Addgene.

Generation of tail tip-derived *Rosa26*$^{PE2/+}$ *Trp53*$^{flox/flox}$ fibroblasts.

To generate a *Rosa26*$^{\mathrm{PE2}}$ cell line for convenient testing of pegRNAs, a 2 cm piece was excised from the tail tip of an anesthetized, 3.5-week-old male. The sample was sprayed with ethanol and then dipped in PBS several times. A lengthwise incision was made, and outside skin and hair were removed. The sample was then incubated at 37 °C in digestion buffer comprised of 5 mL DMEM, 25 μL penicillin-streptomycin, 5 μL Amphotericin B, 10 μL DNase (40U / mL -20 °C; 1:500), 50 μL collagenase (100 mg / mL; 1:100), and 50 μL $CaCl_2$ (36 mM; 1:100). Samples were then washed twice with PBS, and dissociated chunks were added to a 6cm dish with minimal media to promote adherence. Additional media containing Amphotericin B was added the following day.

HEK293 and fibroblast cell culture conditions.

HEK293-FS, split-PE2 3TZ and tail tip-derived *Rosa26*$^{PE2/+}$ fibroblast cells were cultured in standard media consisting of Dulbecco's Modified Eagle's Medium (DMEM) (Corning), penicillin-streptomycin, and 10% (volume/volume) fetal bovine serum (FBS). All cultured cells were incubated at 37 °C and 5% CO_2.

Pancreatic ductal organoid isolation and culture.

Pancreata from mice of the desired genotype were dissected manually and minced with a razor blade. Pancreas tissue was then dissociated by 20 minutes of gentle agitation in pancreas digestion buffer [1X PBS (Corning), 125 U/ml collagenase IV (Worthington)] at 37 ˚C. Tissue suspensions were then strained through 70 µM filters, washed with 1X PBS, and pelleted with slow deceleration by centrifugation. Cells were resuspended in 100% Matrigel (Corning) and plated as 50µL domes into 24-well plates (GenClone). Upon solidification of domes, cells were cultured in organoid complete media as previously described[43], or alternatively, in complete medium as follows: AdDMEM/F12 medium supplemented with HEPES (1x, Invitrogen), GlutaMAX (1x, Invitrogen), penicillin/streptomycin (1x, VWR), B27 (1x Invitrogen), R-Spondin1-Conditioned Medium (10% v/v), A83-01 (0.5µM, Tocris), murine Epidermal Growth Factor (mEGF, 0.05µg/mL, PeproTech), Fibroblast Growth Factor 10 (FGF-10, 0.1µg/mL, PeproTech), Gastrin I (0.01µM, Tocris), recombinant murine Noggin (0.1µg/mL, PeproTech), N-acetyl-L-cysteine (1.25mM, Sigma-Aldrich), Nicotinamide (10mM, Sigma-Aldrich) and Y-27632 (10.5µM, Cayman Chemical Company).

Organoids were passaged using TrypLE Express (Life Technologies) for Matrigel digestion for 15-30 minutes at 37 °C. Organoids were passaged for at least four passages prior to treatment with lentivirus. Organoids were infected at a high multiplicity of infection to ensure 100% recombination. Briefly, concentrated lentivirus (either diluted 1:9 or undiluted) was introduced to cell suspensions at the time of passage. Fluorescence microscopy confirmed Cre-mediated expression of mNG in organoids treated with UPEC, or mScarlet expression in organoids treated with UPEmS. For *Trp53*$^{\text{flox/flox}}$ lines, Nutlin-3a was added to organoid media (10 μM, Sigma-Aldrich) to ensure purification of recombined organoids. For prime edited organoids harboring *Kras*^{G12D} or *Kras*^{G12C} mutations, organoids were cultured in the presence of 1 μM Gefitinib in full organoid media (Cayman) to select for the intended edit. Sotorasib (Selleck) was added to media at 1, 2, and 5 μM for experiments as indicated in the main text. MRTX1133 (MedChem) was added to the media at 2 or 5 μM as indicated in the main text. Selection of prime edited mutations was confirmed by deep amplicon sequencing of organoids several days after the initial infection with lentivirus, and then again after several passages under treatment with drug. For selection of transgene-containing cells from chimera-derived pancreatic organoids, organoids were treated with 800 μg / mL of Geneticin (GIBCO). All drug stocks were dissolved in DMSO (Sigma-Aldrich). Pancreatic organoids were maintained in culture for <30 passages.

Organoid viability and proliferation were quantified using the alamarBlue HS Cell Viability Reagent (ThermoFisher Scientific). Viability reagent was directly added to organoid culture at 1/10 media volume. After 24 hours, 200 μl of reagent-containing media was removed and assayed in replicate in a Tecan Infinite Pro m200 using the manufacturer's parameters. For time course experiments, media containing 10% viability reagent was replenished after 1 PBS wash.

Lung organoid culture.

Lung organoids were cultured as previously described[80]. Briefly, lungs were derived from 8–20-week-old mice and transferred into 500µL dispase and minced. 3-5 mL digestion buffer containing Advanced DMEM/F-12, Penicillin/Streptomycin, Amphotericin B, 1 mg/mL Collagenase (Sigma, C9407-500MG), 40 U/mL DNase I (Roche, 10104159001), 5µM HEPES, and 0.36 mM $CaCl_2$ was added for a 20-60 minute incubation at 37 ºC in a rotating oven. The resulting suspension was incubated in 1 mL ACK Lysis Buffer (Thermo, A1049201) for 3-5 minutes at room temperature to lyse red blood cells. Samples were then washed two times with fluorescence-activated cell sorting (FACS) buffer (1X PBS with 1 mM EDTA and 0.1% BSA) and filtered through 40µm mesh to remove chunks. Samples were resuspended in 150 µL FACS buffer and CD45 cells were depleted using the EasySep^{TM} Mouse CD45 Positive Selection kit (STEMCELL technologies, 18945) and preserving the flow through. Cells were then resuspended in FACS buffer containing 1x PBS, 0.1% BSA, and 2 mM EDTA and were stained with anti-mouse CD31-APC (1:500, Biolegend, 102507), CD45-APC (1:500, BD Biosciences, 559864), EpCAM-PE (1:500, Biolegend, 118206), MHCII-APC-eFluor-780 (1:500, Thermo, 47-5321-82) on ice for 30 minutes and then resuspended in FACS buffer containing DAPI (1µg/mL, Thermo, D1306). The suspensions were then sorted for DAPI-, CD31-, CD45-, EpCAM+, MHCII+ cells. Approximately 20,000 sorted AT2 cells were mixed with Growth Factor Reduced Matrigel (Corning) at a ratio of 1:9 and seeded onto multi-well plates as 20µL drops. The drops were incubated at 37 ºC for 15 minutes to solidify, and then overlaid with F^7NHCS medium supplemented with Y-27632 (Cayman). The cultures were maintained in a humidified 37 ºC / 5% CO_2 incubator at ambient O_2 pressure. Media was replenished every 3-4 days using F^7NHCS medium without Y-27632 and organoids were passaged 6-12 days after plating. For passaging, matrigel drops were dissolved in TrypLE Express (Sigma, 12604-013) and incubated at 37 ºC for

7-15 minutes. The organoid suspensions were then dissociated into single cells by vigorous pipetting, washed twice, resuspended in 1x PBS, and plated as described above. We typically plated 10,000 cells per drop.

Generation of a 3TZ fibroblast cell line with the PE2 enzyme expressed from split inteins. A cell line based on murine 3TZs cells was developed to test *Trp53*-targeted pegRNAs on a *Trp53*$^{+/+}$ background. Two plasmids containing halves of the PE2 enzyme and distinct antibiotic resistance genes were generated via Gibson Assembly. The split intein-based constructs described by Anzalone *et al*. (2019) were utilized to enable post-translational splicing of the intein motifs and subsequent joining of the halves to form the full PE2 enzyme[21,81]. Specifically, the N-terminal half of PE2 (the first 573 amino acids of the Cas9 nickase joined to the Npu N-intein) was PCR-amplified from the U6-DNMT1-hSynapsin-PE2-Nterminal-P2A-EGFP-KASH-lenti plasmid (Addgene no. 135955) and then cloned into a puromycin resistance gene-containing backbone. A blasticidin resistance gene-containing backbone was assembled into a second vector with a PCR-amplified DNA fragment encoding the C-terminal half of PE2 (Npu C-intein joined to the remaining C-terminal half of PE2), amplified from the hSynapsin-PE2-Cterminal-lenti plasmid (Addgene no. 135956).

After Gibson assembly and subsequent plasmid validation by Sanger sequencing, the two constructs were incorporated into lentivirus, which was used to transduce murine 3TZ fibroblast cells. Transduced cells were then incubated for 48 hours. After 48 hours, both infected and control cells were plated into 12-well plates and treated with a range of dilutions of both puromycin and blasticidin, maximally delivered at 10 µg / mL and 20 µg / mL, respectively. After confirming lentivirus-mediated resistance to both antibiotics in infected cells, a polyclonal cell line was maintained generally with standard media containing 8 µg / mL of puromycin and

12 µg / mL of blasticidin. Prime editing in the resulting split-PE2 3TZ cells was confirmed using the *Dnmt1* pegRNA described in Figs. 2-3 as a positive control (Supplementary Figure 9). This cell line was then used to test cohorts of pegRNAs targeted to *Trp53*.

Production of lentivirus and transduction.

Lentivirus was produced by transfection of the expression vector into 293FS* cells along with psPAX2 (psPAX2 was a gift from Didier Trono - Addgene plasmid # 12260 ; http://n2t.net/addgene:12260 ; RRID:Addgene_12260) and pMD2.G (pMD2.G was a gift from Didier Trono - Addgene plasmid # 12259 ; http://n2t.net/addgene:12259 ; RRID:Addgene_12259) packaging plasmids at a 4:3:1 ratio using polyethylenimine or Mirus transfection reagent. For small-scale transfections, 4 x 10^5 cells were seeded per well in a 6-well plate (Corning). 24 hours after seeding, cells were transfected using 4 µL PEI or Mirus (Mirus Bio) according to the manufacturer's specifications. For large-scale transfections, 7.5 x 10^6 cells were seeded in 15 cm plates (Corning). 24 hours after seeding, cells were transfected using 80 µl Mirus or PEI. Media was replaced 24 hours after transfection. 48-72 hours post-transfection, viral supernatant was collected and filtered through 0.45 µm filters. Large-scale virus was concentrated by ultracentrifugation at 25,000 rpm at 4˚C and then resuspended in Opti-MEM (Thermo-Fisher). 1 mL of small-scale viral supernatant was added directly to 1 x 10^5 cells at seeding in a 6-well plate (Corning) for transduction. Small-scale transductions were supplemented with polybrene (10 mg/mL, 1:1000, Sigma). Concentrated large-scale lentivirus and small-scale virus were stored at -80˚ C if not used immediately. Generally, cell lines were infected with small-scale virus while organoids were infected with large-scale virus.

Lentivirus titer assay.

Quantification of lentiviral titer was performed as previously described using a GFP Cre reporter 3TZ cell line[14]. Briefly, concentrated lentivirus was serially diluted and added to 10,000 cells. 48 hours post-infection, cells were trypsinized and analyzed to calculate the percentage of GFP positivity by flow cytometry.

Intratracheal delivery of lentivirus into the lung.

Intratracheal lentiviral infection has been previously described[82]. Mice were anesthetized in an isoflurane chamber. 6×10^4 TU and 1×10^5 TU of lentivirus containing UPEC vectors encoding pegRNAs and Cre recombinase were injected intratracheally into *Trp53*$^{\text{flox/flox}}$;*Rosa26*$^{\text{PE2/+}}$ or *Trp53*$^{\text{flox/flox}}$;*Rosa26*$^{\text{PE2/PE2}}$ mice. Animals were monitored after injection to confirm recovery. Mice were at least six weeks old at the time of infection. Mice were sex and age matched within 4 weeks across experimental arms.

Orthotopic transplantation of pancreatic organoids.

Orthotopic transplants of pancreatic organoids were performed with minor modifications from previously described protocols[43,83,84]. Briefly, animals were anesthetized with Isoflurane, the left abdominal side was depilated with Nair, and the surgical region was disinfected with Chloraprep swabstick (BD). A small incision (~1.5cm) was made in the left subcostal area, and the spleen and pancreas were exteriorized with ring forceps. The organoid suspension (containing 1×10^5 organoid cells in 100μL of 50% PBS + 50% Matrigel) was injected using a 30-gauge needle into the pancreatic parenchyma parallel to the main pancreatic artery. Successful injection was verified by the appearance of a fluid bubble without signs of intraperitoneal leakage. The pancreas and spleen were gently internalized, and the peritoneal and skin layers were sutured independently using 4/0 PGA suture and 4/0 silk suture respectively (AD Surgical). All mice received pre-operative analgesia Buprenorphine Sustained-Release (Bup-SR, 0.5mg/kg) and were followed post-operatively for any signs of distress. Organoid/Matrigel mixtures were kept on ice throughout the whole procedure to avoid solidification. For orthotopic transplantation, syngeneic C57BL/6J *Rosa26*^{PE2} mice (aged 6-17 weeks) were used as recipients. Male pancreatic organoids were only transplanted into male recipients.

Animal studies.

All mouse experiments described in this study were approved by the Massachusetts Institute of Technology Institutional Animal Care and Use Committee (IACUC) (institutional animal welfare assurance no. **A-3125-01**. *Kras*$^{LSL\text{-}G12D}$ and *Trp53*$^{flox/flox}$ mice have previously been described[85,86]. All mice were maintained on a pure C57Bl/6 background. Mice with the appropriate genotypes and between 7-20 weeks were chosen for *in vivo* experiments. Mice of both sexes were used for autochthonous lung tumor initiation, and male mice were chosen for orthotopic pancreatic organoid experiments as the transplanted organoid line was male-derived. Mice were assessed for morbidity according to guidelines set by the MIT Division of Comparative Medicine and were humanely sacrificed prior to natural expiration.

Ultrasound imaging.

Observation of murine pancreatic tumors with high resolution ultrasound was performed as described[43,87]. Animals were anesthetized with Isoflurane and the left subcostal region of animals was depilated with Nair. Animals were imaged with a Vevo3100/LAZRX ultrasound and photoacoustic imaging system (Fujifilm-Visualsonics). Anesthetized animals were positioned supine and angled on an imaging platform for visualization of peritoneal organs. Landmark organs including kidney and spleen were first identified prior to imaging. A thin layer of ultrasound gel was applied over the depilated region of the abdomen. The transducer (VisualSonics 550S) was positioned above the abdomen and set at the scanning midpoint of the healthy pancreas or tumor. Approximately 10 mm of scanning area were used to capture the entirety of pancreas tumors, using a Z-slick thickness of 0.04 mm. Ultrasound scans were uploaded to Vevo Lab Software, from which representative images were exported.

Rodent µCT.

Control and tumor bearing mice were anesthetized by Isoflurane (3%, then maintained at 2.0-2.5% in oxygen – VetEquip, Livermore, CA) and scanned in a prone position using a Skyscan 1276 (Bruker Corp, Billerica, MA) with the following parameters: 100kVp source voltage, 200 uA current, 0.5mm Aluminum X-ray filter, 108ms exposure time, 0.65 degree rotational step size over 360 degrees in a continuous rotation. With 4x4 detector binning, the nominal pixel size after reconstruction (Bruker NRecon software) was 40.16 micron. Data were visualized using ImageJ.

Histology and immunohistochemistry.

Upon sacrifice, pancreata from control and tumor-bearing animals were manually dissected from the peritoneal cavity. A tissue dye (Davidson) was used to mark tumor regions for histology. Tumor-bearing lung was flushed with 1X PBS and separated into separate lobes. Tissue was fixed in Zinc Formalin overnight, transferred to 70% Ethanol, and then embedded in paraffin. Hematoxylin and eosin (H&E) staining was performed per a standard protocol by the Hope Babette Tang Histology Facility at the Koch Institute. Digitally scanned images of H&E slides were obtained with an Aperio ScanScope at 20X magnification. Digital slides were uploaded to the Aiforia™ image analysis software. Histologic quantification of tumor grade was performed by an automated deep neural network and in consultation with veterinary pathologist Dr. Roderick Bronson. Training of the Aiforia algorithm has been previously described[66]. For grade calling, the nsclc_v25 algorithm was used.

Lungs were removed and inflated with Zinc formalin, then left in approximately 5 mL Zinc formalin for 24 hours. Afterwards, samples were switched to 70% ethanol, and later processed for paraffin embedding. For IHC, sections were cut at 4 μm and dewaxed. Antigen retrieval was performed using citrate buffer (pH = 6.0). Endogenous peroxidase was blocked using DAKO Dual Endogenous Enzyme Block, and endogenous species protein was blocked using 2.5% Horse Serum (Vecta). Slides were incubated at 4 ˚C overnight with the following antibodies: anti-NKX2-1 (1:1000, abcam, Cat# ab76013; RRID:AB_1310784), anti-SFTPC (1:5000, Millipore Sigma, Cat# AB3786; RRID:AB_91588), and anti-HMGA2 (1:1000, Cell Signaling Technologies, Cat# 8179S; RRID:AB_11178942). ImPRESS anti-Rabbit horseradish peroxidase and DAB Peroxidase Substrate Kit (Vector) were used to develop slides. Tissues were counterstained with hematoxylin. Slides were digitally scanned and analyzed using QuPath[88].

Isolation of genomic DNA.

Genomic DNA was extracted from organoids or tissue using either the KAPA Express Extract kit (Sigma) or the Qiagen Puregene Core Kit A. Briefly, for extraction with KAPA Express, dissociated cells were pelleted, washed with PBS, and pelleted again prior to resuspension (per condition) in 100 μL consisting of 88 μL PCR-grade water, 10 μL of 10X KAPA Express Extract Buffer, and 2 μL of KAPA Express Extract enzyme. This and subsequent steps were followed per the manufacturer's instructions. For the Qiagen Puregene Core Kit A, briefly, minced tumors or dissociated cells were pelleted and resuspended in Qiagen Cell Lysis Solution and vortexed. Protein and ethanol precipitation were then performed to isolate pure genomic DNA, which was used for PCR amplification, all steps executed according to the manufacturer's instructions. Genomic DNA was then stored at 4 ºC or -20 ºC until use.

DNA sequencing and analysis of genomic DNA samples.

Target loci were amplified from genomic DNA using PCR primers listed and described in Supplementary Table 4. Amplicons were then purified using either agarose gel extraction or using a QIAquick PCR purification kit (Qiagen). Purified amplicons were typically then submitted to the Massachusetts General Hospital Center for Computational and Integrative Biology's DNA Core for next-generation sequencing (samples prepared according to guidelines provided for the CRISPR Sequencing service).

Amplicons prepared for evaluating prime editing efficiency of the initial *Trp53*^{245}- and *Kras*^{96}-targeted pegRNAs described in Supplementary Figs. 7 and 9 were sequenced as previously described[17,89]. Following amplification from genomic DNA, samples were given unique Illumina TruSeq barcodes for pooled sequencing. Barcoded PCR products were pooled and purified by electrophoresis with a 2% agarose gel using a Gel Extraction Kit (QIAGEN), eluting with 30 µL H_2O. DNA concentration was quantified using a Qubit dsDNA High Sensitivity Assay Kit (ThermoFisher Scientific) and sequenced on an Illumina MiSeq instrument (single-end read, 250–300 cycles) according to the manufacturer's protocols.

Sequencing reads stored in FASTQ files were aligned to reference amplicons and analyzed using the deep sequencing analysis program, CRISPResso2[76], V2.2.6, as a command-line tool in prime editing mode. General CRISPResso2 parameters employed for each target are described in Supplementary Note 1. The --split_interleaved_input parameter was included for paired-end data. Generally, prime editing efficiency was calculated as the percentage of reads aligning to the prime edited amplicon (excluding indels) relative to all reads aligning to both the prime edited and reference amplicons (including indels). Only reads with an average phred score >30 were considered[44]. Indel percentages were calculated in similar fashion using the total number of indel-bearing reads designated as "discarded" by CRISPResso2. In rare cases containing contaminant reads from off-target amplicons, contaminating reads were removed *in silico* by requiring every analyzed read to possess 60-90% homology to the reference amplicon. For experiments involving pegRNAs that alter multiple nucleotides, the allele frequency tables output by CRISPResso2 were consulted to confirm that the majority of prime edited reads contained all of the intended nucleotide alterations.

Immunoblotting.

Pancreas organoids were dissociated with TrypLE for 30 minutes at 37 ºC, washed 6x with PBS, and then lysed in cell lysis buffer (RIPA with 100X HALT protease and phosphatase inhibitors). Protein concentration was quantified with the bicinchoninic acid (BCA) assay (Pierce) and boiled in 4x Laemmli Buffer + 355μM 2-mercaptoethanol for five minutes. 15 μg of reduced samples were loaded onto 4-12% Bis-Tris SDS-PAGE gels (Invitrogen) and run at 100V. Gels were transferred onto a PVDF membrane (60 V, 3 hours) and blocked with PBST (PBS + 0.5% Tween20) + 5% milk for one hour. Blots were incubated with primary antibody (p53 clone 1C12, Cell Signaling Technology (CST), β-Actin clone 13E5, CST, 1:5000) overnight at 4 ºC, washed 3 times in PBST, incubated with HRP-conjugated secondary antibody in 5% milk for one hour at room temperature, and then washed 3 times in PBST. Blots were developed with Clarity or Clarity-Max ECL substrate (BioRad) and imaged on a ChemiDoc Gel Imaging System (BioRad).

Mutation frequency analysis using cBioPortal.

Somatic mutation frequencies of *TP53* in human pancreatic cancer were retrieved from cBioPortal[59,60] by querying this gene from four patient cohorts: CPTAC, TCGA (Firehose Legacy), QCMG, and ICGC. Overall mutation frequency was calculated as the fraction of patients possessing mutations at a specific codon divided by the total number of patients for which *TP53* mutation status had been profiled. The same analysis was applied for human lung adenocarcinoma, with the following patient cohorts: Non-Small Cell Lung Cancer[90] (TRACERx, NEJM and Nature 2017) and Pan-Lung Cancer[91] (TCGA, Nat Genet 2016).

References

1. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science* **339**, 1546–1558 (2013).
2. Garraway, L. A. & Lander, E. S. Lessons from the cancer genome. *Cell* **153**, 17–37 (2013).
3. Hyman, D. M., Taylor, B. S. & Baselga, J. Implementing Genome-Driven Oncology. *Cell* **168**, 584–599 (2017).
4. Chang, M. T. *et al.* Accelerating Discovery of Functional Mutant Alleles in Cancer. *Cancer Discov* **8**, 174–183 (2018).
5. Hong, D. S. *et al.* KRASG12C Inhibition with Sotorasib in Advanced Solid Tumors. *New England Journal of Medicine* **383**, 1207–1217 (2020).
6. Lynch, T. J. *et al.* Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. *New England Journal of Medicine* **350**, 2129–2139 (2004).
7. Hill, W., Caswell, D. R. & Swanton, C. Capturing cancer evolution using genetically engineered mouse models (GEMMs). *Trends in Cell Biology* **31**, 1007–1018 (2021).
8. Kersten, K., de Visser, K. E., van Miltenburg, M. H. & Jonkers, J. Genetically engineered mouse models in oncology research and cancer medicine. *EMBO Mol Med* **9**, 137–153 (2017).
9. Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* **23**, 703–713 (2017).
10. Platt, R. J. *et al.* CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440–455 (2014).
11. Annunziato, S. *et al.* In situ CRISPR-Cas9 base editing for the development of genetically engineered mouse models of breast cancer. *EMBO J* **39**, e102169 (2020).
12. Han, T. *et al.* R-Spondin chromosome rearrangements drive Wnt-dependent tumour initiation and maintenance in the intestine. *Nat Commun* **8**, 15945 (2017).
13. Dow, L. E. *et al.* Inducible in vivo genome editing with CRISPR-Cas9. *Nat Biotechnol* **33**, 390–394 (2015).
14. Sánchez-Rivera, F. J. *et al.* Rapid modelling of cooperating genetic events in cancer through somatic genome editing. *Nature* **516**, 428–431 (2014).
15. Winters, I. P. *et al.* Multiplexed in vivo homology-directed repair and tumor barcoding enables parallel quantification of Kras variant oncogenicity. *Nat Commun* **8**, 2053 (2017).
16. Leibowitz, M. L. *et al.* Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nat Genet* **53**, 895–905 (2021).
17. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).

18. Kurt, I. C. *et al.* CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat Biotechnol* **39**, 41–46 (2021).

19. Koblan, L. W. *et al.* Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nat Biotechnol* **39**, 1414–1425 (2021).

20. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* **38**, 824–844 (2020).

21. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).

22. Liu, P. *et al.* Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. *Nat Commun* **12**, 2121 (2021).

23. Gainor, J. F. *et al.* Molecular Mechanisms of Resistance to First- and Second-Generation ALK Inhibitors in ALK-Rearranged Lung Cancer. *Cancer Discov* **6**, 1118–1133 (2016).

24. Gainor, J. F. *et al.* Patterns of Metastatic Spread and Mechanisms of Resistance to Crizotinib in ROS1-Positive Non-Small-Cell Lung Cancer. *JCO Precis Oncol* **2017**, (2017).

25. Kobayashi, S. *et al.* EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med* **352**, 786–792 (2005).

26. Canté-Barrett, K. *et al.* Lentiviral gene transfer into human and murine hematopoietic stem cells: size matters. *BMC Res Notes* **9**, 312 (2016).

27. Zinn, E. & Vandenberghe, L. H. Adeno-associated virus: fit to serve. *Current Opinion in Virology* **8**, 90–97 (2014).

28. Wang, D. *et al.* Adenovirus-Mediated Somatic Genome Editing of Pten by CRISPR/Cas9 in Mouse Liver in Spite of Cas9-Specific Immune Responses. *Hum Gene Ther* **26**, 432–442 (2015).

29. Annunziato, S. *et al.* Modeling invasive lobular breast carcinoma by CRISPR/Cas9-mediated somatic genome editing of the mammary gland. *Genes Dev.* **30**, 1470–1480 (2016).

30. Böck, D. *et al.* In vivo prime editing of a metabolic liver disease in mice. *Science Translational Medicine* **14**, eabl9238 (2022).

31. Sánchez-Rivera, F. J. *et al.* Base editing sensor libraries for high-throughput engineering and functional analysis of cancer-associated single nucleotide variants. *Nat Biotechnol* 1–12 (2022) doi:10.1038/s41587-021-01172-3.

32. Erwood, S. *et al.* Saturation variant interpretation using CRISPR prime editing. *Nat Biotechnol* 1–11 (2022) doi:10.1038/s41587-021-01201-1.

33. Chen, L. *et al.* Programmable C:G to G:C genome editing with CRISPR-Cas9-directed base excision repair proteins. *Nat Commun* **12**, 1384 (2021).

34. Grünewald, J. *et al.* A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing. *Nat Biotechnol* **38**, 861–864 (2020).

35. Yuan, T. *et al.* Optimization of C-to-G base editors with sequence context preference predictable by machine learning methods. *Nat Commun* **12**, 4902 (2021).

36. Anzalone, A. V. *et al.* Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nat Biotechnol* **40**, 731–740 (2022).

37. Jiang, T., Zhang, X.-O., Weng, Z. & Xue, W. Deletion and replacement of long genomic sequences using prime editing. *Nat Biotechnol* **40**, 227–234 (2022).

38. Bult, C. J. *et al.* Mouse Genome Database (MGD) 2019. *Nucleic Acids Res* **47**, D801–D806 (2019).

39. MouseMine, Mouse Genome Informatics Web Site, The Jackson Laboratory, Bar Harbor, Maine. World Wide Web. (2022).

40. Shaner, N. C. *et al.* A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat Methods* **10**, 407–409 (2013).
41. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* **13**, 133–140 (2010).
42. Ng, S. R. *et al.* CRISPR-mediated modeling and functional validation of candidate tumor suppressor genes in small cell lung cancer. *Proc Natl Acad Sci U S A* **117**, 513–521 (2020).
43. Freed-Pastor, W. A. *et al.* The CD155/TIGIT axis promotes and maintains immune evasion in neoantigen-expressing pancreatic cancer. *Cancer Cell* **39**, 1342-1360.e14 (2021).
44. Chen, P. J. *et al.* Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635-5652.e29 (2021).
45. Schlake, T. & Bode, J. Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry* **33**, 12746–12751 (1994).
46. Bindels, D. S. *et al.* mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat Methods* **14**, 53–56 (2017).
47. Vassilev, L. T. *et al.* In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844–848 (2004).
48. Kim, H. K. *et al.* Predicting the efficiency of prime editing guide RNAs in human cells. *Nat Biotechnol* **39**, 198–206 (2021).
49. Nelson, J. W. *et al.* Engineered pegRNAs improve prime editing efficiency. *Nat Biotechnol* **40**, 402–410 (2022).
50. Drost, J. *et al.* Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43–47 (2015).
51. Zafra, M. P. *et al.* An In Vivo Kras Allelic Series Reveals Distinct Phenotypes of Common Oncogenic Variants. *Cancer Discov* **10**, 1654–1671 (2020).
52. Strickler, J. H. *et al.* First data for sotorasib in patients with pancreatic cancer with KRAS p.G12C mutation: A phase I/II study evaluating efficacy and safety. *JCO* **40**, 360490–360490 (2022).
53. Hofmann, M. H., Gerlach, D., Misale, S., Petronczki, M. & Kraut, N. Expanding the Reach of Precision Oncology by Drugging All KRAS Mutants. *Cancer Discovery* **12**, 924–937 (2022).
54. Waters, A. M. & Der, C. J. KRAS: The Critical Driver and Therapeutic Target for Pancreatic Cancer. *Cold Spring Harb Perspect Med* **8**, a031435 (2018).
55. Wang, X. *et al.* Identification of MRTX1133, a Noncovalent, Potent, and Selective KRASG12D Inhibitor. *J Med Chem* **65**, 3123–3133 (2022).
56. Awad, M. M. *et al.* Acquired Resistance to KRASG12C Inhibition in Cancer. *N Engl J Med* **384**, 2382–2393 (2021).
57. Pao, W. *et al.* Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med* **2**, e73 (2005).
58. Freed-Pastor, W. A. & Prives, C. Mutant p53: one name, many proteins. *Genes Dev* **26**, 1268–1286 (2012).
59. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery* **2**, 401–404 (2012).
60. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1 (2013).

61. Schulz-Heddergott, R. *et al.* Therapeutic Ablation of Gain-of-Function Mutant p53 in Colorectal Cancer Inhibits Stat3-Mediated Tumor Growth and Invasion. *Cancer Cell* **34**, 298-314.e7 (2018).

62. Barta, J. A., Pauley, K., Kossenkov, A. V. & McMahon, S. B. The lung-enriched p53 mutants V157F and R158L/P regulate a gain of function transcriptome in lung cancer. *Carcinogenesis* **41**, 67–77 (2020).

63. Shakya, R. *et al.* Mutant p53 upregulates alpha-1 antitrypsin expression and promotes invasion in lung cancer. *Oncogene* **36**, 4469–4480 (2017).

64. Jackson, E. L. *et al.* The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res* **65**, 10280–10288 (2005).

65. Winslow, M. M. *et al.* Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature* **473**, 101–104 (2011).

66. LaFave, L. M. *et al.* Epigenomic State Transitions Characterize Tumor Progression in Mouse Lung Adenocarcinoma. *Cancer Cell* **38**, 212-228.e13 (2020).

67. Aoyama, N. *et al.* Transgenic mice that accept Luciferase- or GFP-expressing syngeneic tumor cells at high efficiencies. *Genes Cells* **23**, 580–589 (2018).

68. Hingorani, S. R. *et al.* Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).

69. Liu, B. *et al.* A split prime editor with untethered reverse transcriptase and circular RNA template. *Nat Biotechnol* (2022) doi:10.1038/s41587-022-01255-9.

70. Amodio, V. *et al.* EGFR Blockade Reverts Resistance to KRASG12C Inhibition in Colorectal Cancer. *Cancer Discov* **10**, 1129–1139 (2020).

71. Canon, J. *et al.* The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* **575**, 217–223 (2019).

72. Escobar-Hoyos, L. F. *et al.* Altered RNA Splicing by Mutant p53 Activates Oncogenic RAS Signaling in Pancreatic Cancer. *Cancer Cell* **38**, 198-211.e8 (2020).

73. Klemke, L. *et al.* The Gain-of-Function p53 R248W Mutant Promotes Migration by STAT3 Deregulation in Human Pancreatic Cancer Cells. *Front Oncol* **11**, 642603 (2021).

74. Schoenfeld, A. J. *et al.* The Genomic Landscape of SMARCA4 Alterations and Associations with Outcomes in Patients with Lung Cancer. *Clin Cancer Res* **26**, 5701–5708 (2020).

75. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–345 (2009).

76. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* **37**, 224–226 (2019).

77. Hsu, J. Y. *et al.* PrimeDesign software for rapid and simplified design of prime editing guide RNAs. *Nat Commun* **12**, 1034 (2021).

78. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE* **3**, e3647 (2008).

79. Akama-Garren, E. H. *et al.* A Modular Assembly Platform for Rapid Generation of DNA Constructs. *Sci Rep* **6**, 16836 (2016).

80. Naranjo, S. *et al.* Modeling diverse genetic subtypes of lung adenocarcinoma with a next-generation alveolar type 2 organoid platform. 2021.12.07.471632 Preprint at https://doi.org/10.1101/2021.12.07.471632 (2021).

81. Zettler, J., Schütz, V. & Mootz, H. D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett* **583**, 909–914 (2009).
82. DuPage, M., Dooley, A. L. & Jacks, T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nat Protoc* **4**, 1064–1072 (2009).
83. Boj, S. F. *et al.* Organoid models of human and mouse ductal pancreatic cancer. *Cell* **160**, 324–338 (2015).
84. Kim, M. P. *et al.* Generation of orthotopic and heterotopic human pancreatic cancer xenografts in immunodeficient mice. *Nat Protoc* **4**, 1670–1680 (2009).
85. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev* **15**, 3243–3248 (2001).
86. Marino, S., Vooijs, M., van Der Gulden, H., Jonkers, J. & Berns, A. Induction of medulloblastomas in p53-null mutant mice by somatic inactivation of Rb in the external granular layer cells of the cerebellum. *Genes Dev* **14**, 994–1004 (2000).
87. Sastra, S. A. & Olive, K. P. Quantification of murine pancreatic tumors by high-resolution ultrasound. *Methods Mol Biol* **980**, 249–266 (2013).
88. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878 (2017).
89. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
90. Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non–Small-Cell Lung Cancer. *N Engl J Med* **376**, 2109–2121 (2017).
91. Campbell, J. D. *et al.* Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet* **48**, 607–616 (2016).

CHAPTER 4: DISCUSSION

In the work described in Chapter 2, my colleagues and I studied the immunopeptidome of pancreatic cancer, aiming to uncover new targets for immunotherapy. We were motivated by a burst of publications in recent years that described abundant noncanonical sources of cancer-restricted peptide:MHC complexes [1,2]. We sought to extend this search to pancreatic cancer, as such targets remain largely unexplored in this disease. Harnessing patient-derived organoids and matched whole-genome and RNA-sequencing, we found a wide variety of peptides derived from canonical and noncanonical sources in the immunopeptidomes of these samples. Critically, many peptides derived from open reading frames in transcripts that are functionally unexpressed in normal tissues, suggesting they may represent safe targets for immunotherapy. In other cases, we observed peptides from open reading frames that did not generate detectable peptides in the immunopeptidomes of normal tissues, establishing a second class of potentially safe targets that may be *translated* in a cancer-restricted fashion.

We also uncovered mutation-derived neoepitopes in most samples, counter to expectations in the literature suggesting that most pancreatic tumors do not present neoepitopes. We established evidence that sequencing of bulk pancreatic tumors may obscure the presence of several neoepitopes in a subset of low-purity pancreatic tumors. Our results may have implications for improving bioinformatic prediction in ongoing clinical efforts that aim to vaccinate pancreatic cancer patients with predicted neoepitopes.

Going forward, more work is needed to (1) identify additional sources of pancreatic cancer-restricted epitopes, (2) establish rigorous methods to ascertain their off-tumor expression patterns, and (3) develop therapeutic agents that can effectively target them. In the sections ahead, I will explore each of these challenges in turn.

Identifying the full scope of pancreatic cancer-restricted epitopes.

To my knowledge, our work was the first to incorporate a proteogenomics approach to examine the immunopeptidome of human PDAC samples. I accomplished this in part with a computational pipeline that incorporates phased germline and somatic variants into patient-specific protein databases. Most individuals harbor ~4 million germline variants compared to the reference genome [3]. Prior work from Malachi Griffith's laboratory has shown that ~5% of neoepitope-encoding somatic variants occur in-phase with proximal germline variants that would affect the predicted neoepitope sequence [4]. I hypothesized such variants would also affect epitopes derived from retained introns, as intronic sequences generally do not experience the purifying selective pressure that constrains variation in exons [5]. I was optimistic that a variant-recoded search space would enable identification of more targets, but we did not identify any additional PDAC-restricted neoepitopes as a result of incorporating germline variants. We did identify hundreds of germline variant-derived epitopes, but none coincided with a somatic mutation.

Retained introns may be an exception, as several detected peptides were mapped specifically to a shifted reading frame of a retained intron. Anecdotally, it is worth mentioning that I conducted a search for germline indels present within these introns and found numerous variants overlapping the coordinates of retained introns. However, additional validation is still required, including rigorous review of the candidate indels and comparison of their size to the reading frames from which potential intronic epitopes were detected.

In general, incorporating germline variants did enable a slight overall increase (<1%) in the number of peptides identified in canonical protein-coding genes. Thus, germline variants may slightly shrink the large fraction of high-quality spectra that are often left unmapped in immunopeptidomics datasets [6]. I did not extend this analysis to include germline variation within nuORFs or somatic frameshift sequences, which should be considered in future search spaces informed by proteogenomics.

In general, our work was the first to expand the search for pancreatic cancer-restricted epitopes to include retained introns, somatic indels in tumor suppressor genes, and nuORFs derived from lncRNAs, pseudogenes, and untranslated regions. Work from the laboratories of Catherine Wu, Aviv Regev, and Steven Carr recently showed broad translation and MHC-I presentation of peptides from several classes of "noncoding" sequences in cancers cells. Notably, these included several transcripts that are transcribed and/or translated in a cancer-specific fashion, i.e., not in healthy tissues. A prior study from Elaine Fuch's laboratory also showed broad translation of 5' UTRs induced by the transcription factor SOX2, acting downstream of oncogenic RAS in mouse models of squamous cell carcinoma (SCC) [7]. These findings are reminiscent of long-standing observations that cancer cells can ectopically express typically silenced genes as they undergo de-differentiation [8]. Such genes include those whose expression is only seen in embryonic, placental, or reproductive tissues; mesothelin is one such example long-established in pancreatic cancer [9]. Broad, cancer-specific dysregulation of the transcriptome and translation machinery appears to induce a similar effect for a cohort of nuORF-associated transcripts. This is supported by the apparent absence of some of these transcripts in healthy tissues, and by their potential induction by oncogenic RAS [7]. Our findings furnish a large set of additional noncanonical peptides that seem to be restricted to PDAC or select healthy tissues, underscoring their therapeutic potential.

While our work represents a significant step forward in the exploration of immunotherapy targets for pancreatic cancer, a vast array of additional sources remains unexplored. Most promising among these may be other classes of epitopes derived from cancer-specific splicing events.

Splicing dysregulation is a frequent feature of cancer [10]. Studies have shown that cancer-specific splicing events, including exon shuffling, exon skipping, and alternative 3' or 5' splice site usage, can give rise to novel isoforms not previously observed in normal tissues [11,12]. We did not explore these possibilities in Chapter 2, though the data we publish will provide an excellent resource for future researchers to search for epitopes derived from mis-splicing events in pancreatic cancer. To this end, future work should consider *de novo* transcript assembly based on RNA-Seq of PDAC organoids. Our analysis, while extensive, was only based on annotated transcripts [2], and it is conceivable that pancreatic cancer cells may produce novel transcript isoforms not currently captured in reference transcriptomes.

Finally, it is worth noting that the empirical methods employed in Chapter 2 almost certainly do not capture most peptides presented on pancreatic cancer cells. While we achieved a high depth of more than 15,000 peptides for each sample, it is thought that a typical human cell presents 100,000-750,000 peptide:MHC-I complexes *per HLA gene* [13]. While many complexes will be duplicates harboring the same peptide, this statistic implies that current immunopeptidomics technology may only identify a small fraction of peptides presented by MHC-I in a given population of cells. Thus, improvements in the sensitivity of peptide:MHC-I mass spectrometry profiling will complement expanded search spaces, enabling identification of more epitopes still undiscovered.

Methods to ascertain off-tumor expression patterns of candidate epitopes.

For all classes of candidate immunotherapy targets, it is important to rigorously evaluate normal tissue expression patterns. This helps designate which epitopes are therapeutically prohibitive, such as those expressed in vital tissues like the heart. Several clinical studies highlight the critical importance of anticipating off-tumor expression. For example, clinical trials conducted with TCRs engineered to target peptides derived from two supposed cancer-restricted antigens, MART-1 and MAGE-A3, have encountered numerous safety issues. In the worst cases, these have resulted in treatment-induced fatalities [14,15]; in others, they have spawned adverse events such as hearing loss, neurological toxicities, and severe colitis. Some of these toxicities were attributed to unanticipated expression of the antigen on normal cell types (e.g., MART-1 expressed in melanocytes of the middle ear) [11], while others were attributed to cross-reactivity of the TCR to self-antigens distinct from the target [14]. Severe toxicities may be an acceptable tradeoff for therapies that efficiently treat the disease, but they should be limited and managed as much as possible.

Consortium efforts like GTEx and the Human Protein Atlas afford researchers an unprecedented ability to examine transcript expression patterns throughout healthy tissues. For clinical application, it is imperative to examine these datasets as deeply and exhaustively as possible. While this will go a long way, it should be noted that these databases were historically restricted to profiling bulk tissues, though there is an increasing incorporation of single cell datasets. While a candidate target may not be identified as expressed in a bulk tissue, it remains possible that the target is expressed highly in a specific cell subtype, a signal that can be diluted in bulk sequencing datasets. Thus, in ideal practice, researchers should examine target expression in both large bulk sequencing and single-cell sequencing datasets. These efforts will better curate the expression patterns of candidate targets, including within rare tissue subtypes for which toxicity should be anticipated. This will inevitably shrink the list of viable, cancer-restricted targets, but it will better inform clinical trial design and enable management of toxicities that can be addressed.

Finally, the search for cancer-specific epitopes should not end with those naturally presented. Lu and colleagues have shown that inhibition of factors that regulate splicing proteins can unearth dozens of additional cancer epitopes [12]. To this point, our work has shown that pancreatic cancer cells harbor hundreds of retained introns not identified in normal tissues. While these did not seem to generate frequent epitopes in our dataset, they represent a wellspring of potential protein-coding RNA species. Pharmacological efforts could focus on specifically enhancing the expression and stability of these transcripts, thereby promoting their translation and MHC-I presentation. This could serve to dramatically amplify the load of cancer-specific epitopes presented by pancreatic cancer cells.

Developing therapeutic agents to effectively target PDAC-restricted epitopes.

Target identification is a critical first step toward clinical translation. The next is development of an applicable therapeutic agent. The targets identified in my work could be amenable to many immunotherapeutic approaches. This includes the adoptive transfer of T cells specific to one or more of these targets, as well as cancer vaccines, which could be augmented to include both traditional neoepitopes and noncanonical epitopes described in Chapter 2. I consider these the most likely and currently practical therapeutic applications. In addition to adoptive cell therapy and cancer vaccines, I also imagine other applications that could be considered as new T cell engineering technologies evolve. I have outlined a few speculative examples below.

MAP-specific CAR-T cells based on single-chain fragment variables

The immune system may frequently fail to respond to epitopes presented by pancreatic tumors. Hedge and colleagues used a neoantigen-expressing mouse model of pancreatic cancer to show that a key reason may be a lack of T cell priming [16]. This is explained by a scarcity of dendritic cells in the tumor microenvironment, which diminishes immunosurveillance and may prevent activation of naïve T cells specific to tumor antigens. Other work from Stromnes *et al*. (2015) shows that adoptively transferred, tumor antigen-specific CD8$^+$ cells can mediate tumor regressions in mouse models of PDAC. These studies suggest we can overcome the lack of an endogenous immune response through delivery of antigen-specific T cells established exogenously.

In Chapter 2, my colleagues and I demonstrated a method to generate T cells specific to candidate PDAC-restricted epitopes using PBMCs from healthy donors. We are in the process of sequencing the TCRs at single cell resolution. Once sequenced, we will select the most frequent TCR clones and clone them into lentiviral constructs. Finally, we will transduce CD8$^+$ T cell lines with the TCR constructs, followed by coculture of the engineered T cells with patient-derived organoids. This will enable us to confirm and optimize antigen-mediated killing of pancreatic cancer cells. This procedure could also be harnessed clinically, by introducing validated, PDAC antigen-specific TCRs into patient-derived T cells, followed by adoptive cell transfer back into the patient.

However, our study indicates that some cancer-restricted targets may not be amenable to this procedure. For example, *IGF2BP3*-encoded peptides represent a promising cohort of PDAC-associated epitopes, but we were unable to identify T cells that recognize them. We attributed this to MHC-I presentation of IGF2BP3 peptides in thymic tissue, which probably eliminates IGF2BP3 peptide-specific T cells during central tolerance. However, as discussed in Chapter 1, new approaches enable researchers to engineer single chain fragment variables (scFvs) that bind with high affinity to specific peptide:MHC complexes. These can then be engineered as chimeric antigen receptors to form CAR-T cells, or they can be used as bi-specific antibodies that redirect CD8$^+$ T cells against the antigen. Both approaches would bypass the limitations of native T cell repertoires and enable the development of therapeutics that effectively target antigen-bearing pancreatic tumors.

Dual TCR-engineered T cells conditional on a logical gate

While a fair number of peptide:MHC complexes seem to be restricted to pancreatic cancer cells, this increases dramatically if we consider combinations of peptide:MHC pairs. For example, 290 single peptide:MHC complexes were found in 100% of analyzed patient samples, and a subset of these complexes are PDAC-restricted. However, the commutative property dictates that these targets co-occur in up to 41,905 peptide:MHC complex *pairs*. A pair of complexes is more likely to be cancer-specific than a single complex, since the underlying transcripts will often exhibit different normal tissue expression patterns. Thus, in cancer cells, it is possible to observe the co-expression of two transcripts that are rarely or never expressed together in normal tissues. With this in mind, one could envision an engineered cytotoxic T cell that would only kill its target upon the simultaneous recognition of two peptide:MHC complexes.

Choe *et al*. demonstrated proof of this concept with synthetic notch (synNotch) CAR-T cells. In these cells, a synNotch receptor first recognizes its cognate antigen on a target cell, stimulating a downstream signaling pathway that directly activates transcription of the CAR. If the CAR then binds its cognate antigen on the same target cell, the T cell unleashes its effector response. Thus, target cell killing requires recognition of two antigens. Importantly, both synNotch receptors and CARs can be engineered to recognize custom targets with scFvs [17]. With these considerations in mind, I hypothesize that the synNotch CAR-T approach could be repurposed to target two peptide:MHC complexes from the catalog of those established by the work in Chapter 2. By targeting two MAPs derived from different source transcripts with limited, mutually exclusive healthy tissue expression patterns, it may be possible to develop a large suite of novel T cell therapies with a reduced risk of off-tumor, on-target toxicity concerns. In this way, researchers could convert PDAC-associated MAPs into PDAC-specific MAP pairs, as many of the PDAC-associated MAPs identified in Chapter 3 seem to be solely expressed in PDAC cells and one normal tissue type.

Next-generation animal models of human cancer.

Cancer modeling reached an inflection point soon after the advent of CRISPR, when pioneering work from our laboratory and others showed that CRISPR technology can install somatic cancer driver mutations *in vivo*. These studies highlighted a new paradigm of rapid cancer modeling unconstrained by the construction of new germline alleles in mice. The work described in Chapter 3 further extended this paradigm with prime editing, a more versatile iteration of CRISPR technology. Prime editing enables researchers to model a broader spectrum of driver mutations than we could with prior technologies based on CRISPR-Cas9 and base editing. We introduced the first prime editing GEMM, inducing both lung and pancreatic tumors *in vivo* with custom mutations including both transition and transversion mutations. But the full potential of this system remains untapped. For example, future studies could employ UPEC derivatives that accommodate two or more pegRNAs for multiplexed prime editing. Combinations of point mutations, especially those including putative gain-of-function p53 mutations, have been largely unexplored by models based on Cas9 and base editing. In principle, our system could model simultaneous point mutations, provided the availability of two pegRNAs of sufficient efficiency. This would enable a variety of interesting questions. For example, our work reinforced the notion that $\text{KRAS}^{\text{G12C}}$ mutations are far less potent in generating pancreatic tumors than $\text{KRAS}^{\text{G12D}}$. However, it remains unclear what additional changes $\text{KRAS}^{\text{G12C}}$-driven pancreatic tumors require to bypass this limited oncogenic potential. Our study shows that homozygous deletion of *Trp53* is insufficient to promote tumorigenesis in combination with $\mathrm{KRAS}^{\mathrm{G12C}}$. Nevertheless, 2% of human pancreatic tumors are driven by G12C mutations [18]. The prime editing GEMM could be used to test a variety of other hypotheses. For example, custom combinations of mutations in *Trp53*, *Smad4*, *Cdkn2a*, and/or other candidate driver genes could be studied in the presence of Kras^{G12C} to identify genotypes that enable progression to pancreatic adenocarcinoma. These studies should be guided by examination of patient sequencing data, which will help prioritize candidate driver mutations that may frequently co-occur with KRAS^{G12C}.

In addition to modeling primary mutations, this system represents a valuable co-clinical tool for studies of secondary mutations associated with targeted therapy resistance. As we show with the Kras^{Y96C} resistance mutation, prime editing GEMMs can mediate composite mutations within the same gene. This feature can be used to rapidly establish tumors harboring both primary and secondary mutations, thereby enabling *in vivo* studies that seek to identify novel therapeutic vulnerabilities unique to the resistant oncogene. This also allows preclinical trials of second-generation therapies targeted against the resistant oncogene.

Prime editing GEMMs also provide a versatile system for testing long-standing hypotheses regarding gain-of-function mutations in tumor suppressor genes, particularly *Trp53*. As discussed in Chapter 1, comparisons of the KPC mouse model, based on $\mathit{Trp53}^{R172H}$, and models based on $\mathit{Trp53}^{flox}$ have shown that $\mathit{Trp53}^{R172H}$ promotes accelerated tumor progression, consistent with a gain-of-function effect. Similar effects have been posited or demonstrated for other p53 mutations (see Chapter 3), including $\mathit{TP53}^{R248Q/W}$, but these remain untested in many cancer types that commonly exhibit these mutations. By affording broad editing versatility, prime editing GEMMs will facilitate rapid functional studies of these mutations. Ultimately, through dissection of individual driver mutations and their combinations, prime editing GEMMs will promote new discoveries that may reveal additional mutation-specific vulnerabilities, informing ever more precise treatment paradigms tailored to individual cancer genotypes.

References

1. Smart, A. C. *et al.* Intron retention is a source of neoepitopes in cancer. *Nat. Biotechnol.* **36**, 1056–1058 (2018).
2. Ouspenskaia, T. *et al.* Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat. Biotechnol.* **40**, 209–217 (2022).
3. Hiltemann, S., Jenster, G., Trapman, J., van der Spek, P. & Stubbs, A. Discriminating somatic and germline mutations in tumor DNA samples without matching normals. *Genome Res.* **25**, 1382–1390 (2015).
4. Hundal, J. *et al.* Accounting for proximal variants improves neoantigen prediction. *Nat. Genet.* **51**, 175–179 (2019).
5. Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* **4**, e08890 (2015).
6. Mommen, G. P. M. *et al.* Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc. Natl. Acad. Sci.* **111**, 4507–4512 (2014).
7. Sendoel, A. *et al.* Translation from unconventional 5′ start sites drives tumour initiation. *Nature* **541**, 494–499 (2017).
8. Weinberg, R. A. & Weinberg, R. A. *The Biology of Cancer*. (W.W. Norton & Company, 2013). doi:10.1201/9780203852569.
9. Kleeff, J. *et al.* Pancreatic cancer. *Nat. Rev. Dis. Primer* **2**, 1–22 (2016).
10. Stanley, R. F. & Abdel-Wahab, O. Dysregulation and therapeutic targeting of RNA splicing in cancer. *Nat. Cancer* **3**, 536–546 (2022).
11. Frankiw, L., Baltimore, D. & Li, G. Alternative mRNA splicing in cancer immunotherapy. *Nat. Rev. Immunol.* **19**, 675–687 (2019).
12. Lu, S. X. *et al.* Pharmacologic modulation of RNA splicing enhances anti-tumor immunity. *Cell* **184**, 4032-4047.e31 (2021).
13. Kumari, S. *et al.* Alloreactive cytotoxic T cells provide means to decipher the immunopeptidome and reveal a plethora of tumor-associated self-epitopes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 403–408 (2014).
14. Cameron, B. J. *et al.* Identification of a Titin-derived HLA-A1-presented peptide as a cross-reactive target for engineered MAGE A3-directed T cells. *Sci. Transl. Med.* **5**, 197ra103 (2013).
15. van den Berg, J. H. *et al.* Case Report of a Fatal Serious Adverse Event Upon Administration of T Cells Transduced With a MART-1-specific T-cell Receptor. *Mol. Ther. J. Am. Soc. Gene Ther.* **23**, 1541–1550 (2015).
16. Hegde, S. *et al.* Dendritic Cell Paucity Leads to Dysfunctional Immune Surveillance in Pancreatic Cancer. *Cancer Cell* **37**, 289-307.e9 (2020).

17. Morsut, L. *et al.* Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell* **164**, 780–791 (2016).
18. Zafra, M. P. *et al.* An In Vivo Kras Allelic Series Reveals Distinct Phenotypes of Common Oncogenic Variants. *Cancer Discov.* **10**, 1654–1671 (2020).

APPENDIX I: THE CD155/TIGIT AXIS PROMOTES AND MAINTAINS IMMUNE EVASION IN NEOANTIGEN-EXPRESSING PANCREATIC CANCER

Authors: William A Freed-Pastor[1,2], Laurens J Lambert[1,3], Zackery A Ely[1,3], Nimisha B Pattada[1], Arjun Bhutkar[1], George Eng[1,4], Kim L Mercer[1,5], Ana P Garcia[1], Lin Lin[1], William M Rideout III[1], William L Hwang[1,6,7], Jason M Schenkel[1,8], Alex M. Jaeger[1], Roderick T Bronson[1], Peter MK Westcott[1], Tyler Hether[9], Prajan Divakar[9], Jason W Reeves[9], Vikram Deshpande[4], Toni Delorey[7], Devan Phillips[7], Omer H Yilmaz[1,3,4], Aviv Regev[1,3,5,7,10], Tyler Jacks[1,3,5,11]

Author Contributions: W.F.P. and T.J. conceived of, designed and directed the study; Z.A.E. designed and constructed the neoepitope prediction pipeline and conducted all scRNA-Seq bioinformatic analyses and neoepitope predictions; W.F.P., L.J.L., N.B.P., A.P.G., K.L.M., performed all types of experiments reported in the study; A.B. conducted TCGA bioinformatic analyses and provided conceptual advice; G.E., V.D., O.H.Y. provided human biospecimens and conceptual advice; W.F.P., L.L., N.B.P. performed murine surgeries; G.E., R.T.B. provided pathology expertise and review; W.F.P., A.P.G., W.M.R. conducted mESC targeting and chimera generation; W.L.H., T.D., D.P. performed scRNA-seq. T.H., P.D., and J.W.R. provided biostatistical support for Nanostring GeoMx profiling. J.M.S., A.M.J., P.M.K.W., O.H.Y., A.R., and A.B. provided conceptual advice; W.F.P., L.J.L., Z.A.E., and T.J. wrote the manuscript with comments from all authors.

---

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

[2] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

[3] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[4] Department of Pathology, Massachusetts General Hospital, Boston, MA 02114, USA

[5] Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

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

[7] Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[8] Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA

[9] NanoString Inc., Seattle, WA, 98109, USA

[10] Current address: Genentech, 1 DNA Way, South San Francisco, CA, USA

[11] Correspondence: tjacks@mit.edu

Abstract

The CD155/TIGIT axis can be co-opted during immune evasion in chronic viral infections and cancer. Pancreatic adenocarcinoma (PDAC) is a highly lethal malignancy, and immune-based strategies to combat this disease have been largely unsuccessful to date. We corroborate prior reports that a substantial portion of PDAC harbors predicted high affinity MHC class I-restricted neoepitopes and extend these findings to advanced/metastatic disease. Using multiple preclinical models of neoantigen-expressing PDAC, we demonstrate that intratumoral neoantigen-specific $CD8^+$ T cells adopt multiple states of dysfunction, resembling those in tumor-infiltrating lymphocytes of human PDAC patients. Mechanistically, genetic and/or pharmacologic modulation of the CD155/TIGIT axis was sufficient to promote immune evasion in autochthonous neoantigen-expressing PDAC. Finally, we demonstrate that the CD155/TIGIT axis is critical to maintain immune evasion in PDAC and uncover a combination immunotherapy (TIGIT/PD-1 co-blockade plus CD40 agonism) that elicits profound anti-tumor responses in preclinical models, now poised for clinical evaluation.

Introduction

Pancreatic cancer is the third leading cause of cancer related deaths in the United States [1] and, despite considerable progress in improving chemotherapeutic regimens [2,3], advanced/metastatic pancreatic adenocarcinoma (PDAC) continues to carry a dismal prognosis with fewer than 5% of patients surviving to five years. While insights regarding the molecular and cellular mechanisms of immune escape have fueled tremendous clinical successes in a range of tumor types, microsatellite-stable PDAC, which represents greater than 98% of all patients [4], has been largely refractory to available immune checkpoint blockade [5–7]. While harboring an intermediate mutational burden [8], recent whole-exome sequencing efforts have demonstrated that the majority of early-stage human PDAC contains predicted neoepitopes [9,10]. In fact, neoantigen-specific tumor-infiltrating lymphocyte (TIL) clones can be isolated from human PDAC [11–14] and $CD8^+$ TILs express multiple co-inhibitory receptors, consistent with chronic antigen stimulation and T cell exhaustion/dysfunction [15]. Extensive profiling of the immune landscape in human PDAC has uncovered a complex immune microenvironment, characterized by numerous immune-suppressive cell populations and an enrichment in exhausted/dysfunctional $CD8^+$ T cells, marked by elevated surface expression of the co-inhibitory receptor TIGIT [16,17]. However, the tumor-reactivity of dysfunctional/exhausted $CD8^+$ T cells in PDAC is currently unknown.

Previous work has investigated neoantigen expression in PDAC using transplantation of monolayer cell lines [18,19] or autochthonous genetically-engineered mouse models [20]. However, the results of these studies have thus far been contradictory, with monolayer-based models leading to preponderant T cell-mediated rejection, while neoantigen-expression paradoxically leads to tumor acceleration in autochthonous PDAC. Organotypic culture has emerged as a robust complementary platform to traditional genetically-engineered animals for preclinical modeling [21–23] as orthotopic transplantation of pancreatic organoids closely recapitulates the genetics and histopathology of human PDAC [22,24].

Here we demonstrate that PDAC, including advanced/metastatic lesions, harbors predicted high affinity neoepitopes with novel MHC class I binding ability relative to their wild-type peptide counterparts. Using multiple preclinical models of neoantigen-expressing murine PDAC paired with co-clinical profiling of human PDAC, we uncover the CD155/TIGIT axis as both necessary and sufficient to maintain immune evasion in PDAC. Finally, we reveal a combination immunotherapy (TIGIT/PD-1/CD40a), which leverages this dependency, capable of eliciting profound anti-tumor responses in preclinical models.

RESULTS

Both localized and advanced/metastatic human PDAC harbor predicted high affinity neoepitopes

Recent sequencing and computational studies have challenged the claim that pancreatic cancer harbors few predicted neoantigens [9,10]. However, to date, neoepitope prediction in PDAC has primarily focused on missense mutations, which may be significantly underestimating the total neoantigen burden in PDAC. In addition, these analyses have largely been performed on early-stage/resectable pancreatic cancer, which represents a minority of all PDAC patients [25] and, importantly, fails to encompass the patients most likely to be considered for immunotherapy clinical trials.

To address the broader neoepitope landscape in PDAC, we developed a neoepitope prediction pipeline incorporating HLA allele typing, mutation calling, variant effect prediction, and peptide:MHC class I binding predictions; we also expanded the search space to consider variants derived from missense, frameshift, and in-frame insertion/deletion mutations (**Figure**

S1A, STAR Methods). Consistent with prior neoepitope prediction studies in PDAC, our pipeline uncovered a wealth of putative neoepitopes in early-stage resectable tumors, profiled as part of The Cancer Genome Atlas (TCGA) [26] (**Figure 1A and Table S1**). Even after excluding one patient with mismatch repair deficiency from the TCGA dataset, the majority of patients (86%, 127/148) harbored putative neoepitopes, with 73% (108/148) harboring one or more neoepitopes with predicted high affinity (<50 nM) for MHC class I (**Figure 1A**). We also examined instances of novel predicted MHC class I binding, in which variants were predicted to confer MHC class I binding (affinity <500 nM) or strong binding (<50 nM) ability relative to corresponding wild-type sequences predicted to have low affinity for MHC class I (>1000 nM) or frameshift-derived variants, which had no corresponding wild-type sequence. Using this 'non-binding-to-binding' analysis, we observed a considerable proportion of patients (81%, 120/148) whose tumors harbored one or more of these potentially immunogenic neoepitopes (**Figure 1B**).

Figure 1

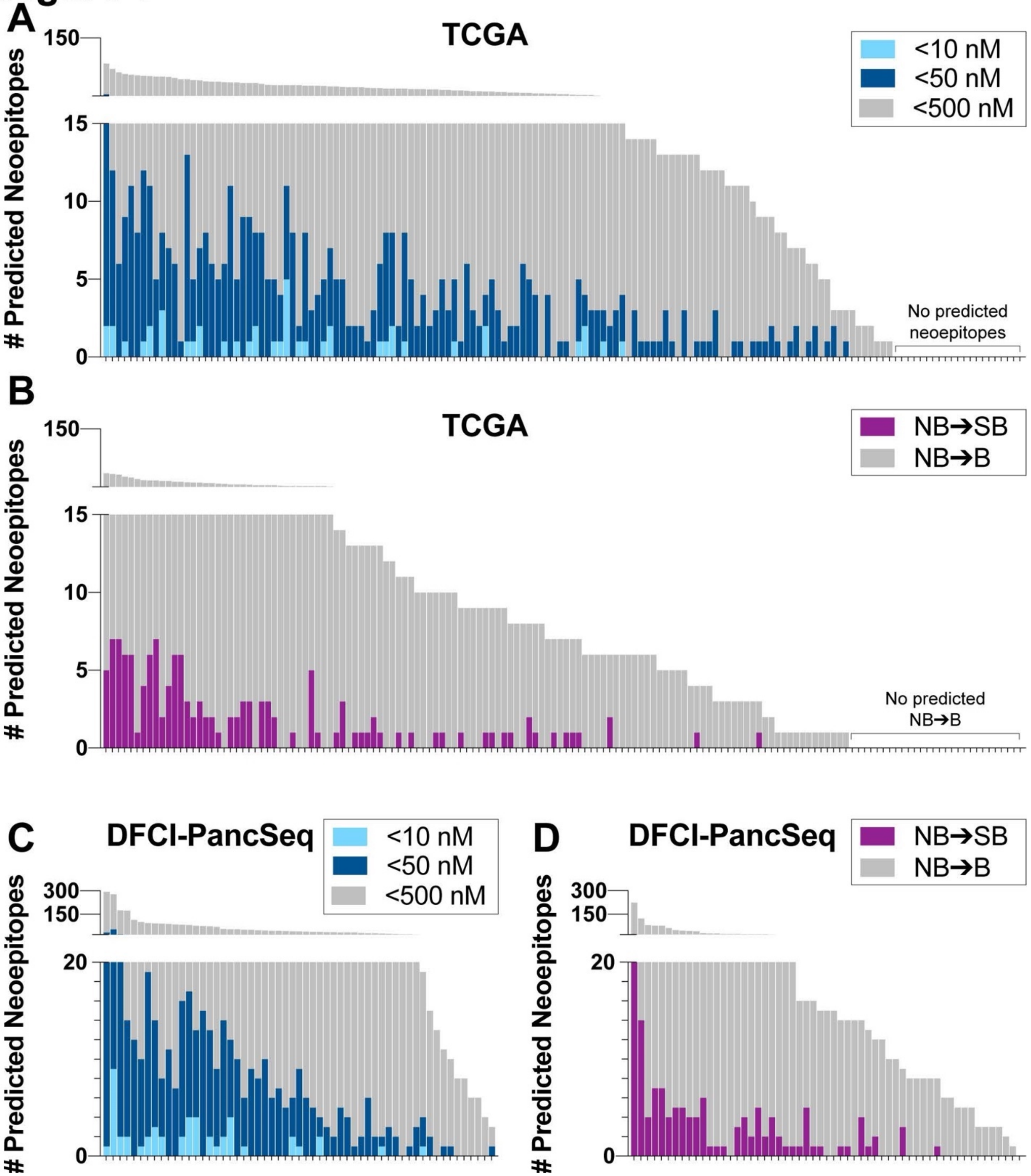


Figure 1. Pancreatic adenocarcinoma harbors predicted high affinity MHC class I-restricted neoepitopes

(A) Neoepitope landscape in TCGA PAAD cohort (n=148) color-coded by predicted affinity for MHC class I.

(B) Predicted non-binder (NB) to binder (B) or strong binder (SB; <50 nM) neoepitopes in TCGA.

(C) Neoepitope landscape in DFCI-PancSeq cohort (n=57) color-coded by predicted affinity for MHC class I.

(D) Predicted non-binder (NB) to binder (B) or strong binder (SB; <50 nM) neoepitopes in DFCI-PancSeq.
See also Figure S1 and Table S1.

To extend these results, we performed a parallel analysis on the Dana-Farber Cancer Institute (DFCI) PancSeq cohort, which profiled 71 advanced/metastatic PDAC patients [27], 57 of whom had successful matched whole exome sequencing (WES) and mRNA-sequencing (RNA-seq) available for analysis. All (100%, 57/57) of these advanced/metastatic patients harbored predicted neoepitopes, with a significant proportion (87%; 50/57) possessing predicted high-affinity (<50 nM) neoepitopes (**Figure 1C and Table S1**). By performing a 'non-binding-to-binding' analysis in biospecimens from advanced/metastatic disease, we found that the vast majority (98%; 56/57) harbored one or more of these neoepitopes (**Figure 1D**). In aggregate, we found that the overall PDAC neoepitope landscape is increased by 28.3% with the inclusion of in-frame insertion/deletion and frameshift mutations (**Figure S1 and Table S1**).

Neoantigen-expressing pancreatic organoids model immune clearance and immune evasion

As a significant proportion of both localized and advanced/metastatic PDAC harbors predicted high-affinity MHC class I-restricted neoepitopes, we set out to develop improved preclinical models to delineate the molecular and cellular mechanisms of immune evasion in this subset of patients. Using CRISPR/Cas9-assisted homology-directed repair (HDR), we generated knock-in models using defined neoantigens, expressed from the *Hipp11* safe harbor locus [28]. Specifically, we utilized the high affinity MHC class I-restricted CD8$^+$ T cell antigen (OVA$_{257\text{–}264}$ [SIINFEKL] - flanked by 17 amino acids on its N-terminus and 9 amino acids on its C-terminus to ensure antigen processing), linked on a polycistronic transcript to the bright, monomeric red fluorescent protein mScarlet [29], or two recently described endogenously arising MHC class I-restricted neoantigens: a single missense mutation in the laminin 4 alpha subunit ("LAMA4-G1254V") or a single missense mutation in alpha-1,3-glucosyltransferase ("ALG8-A506T"), which have been extensively characterized in murine sarcoma [30,31], fused to the C-terminus of mScarlet (**Figure 2A and S2A**). We derived 'genetically-defined' pancreatic organoids from healthy pancreata of genetically-engineered *Kras*$^{LSL\text{-}G12D/WT}$*; Trp53*$^{flox/flox}$*; H11*^{neoantigen} ('KP;SIIN'; 'KP;mLAMA4'; 'KP;mALG8') animals. Following *ex vivo* delivery of adenoviral Cre recombinase, pancreatic organoids express oncogenic Kras with loss of the p53 tumor suppressor gene (sufficient for tumorigenesis upon transplantation), in addition to stable and uniform neoantigen expression (**Figure 2B and S2A**). Orthotopic transplantation of neoantigen-expressing pancreatic organoids into immune-deficient recipients, using either a depleting antibody to specifically deplete the CD8 T cell compartment or *Rag2*$^{-/-}$ animals lacking an adaptive immune system, resulted in 100% penetrance of mScarlet-positive (mScarlet^{+}) tumor formation (**Figure 2C**). In contrast, when we orthotopically transplanted neoantigen-expressing organoids into immune-competent recipients, we observed two predominant outcomes: 1) immune-mediated clearance of all neoantigen-expressing tumor cells (no discernable tumor on necropsy and histologically normal pancreas that was negative for mScarlet expression [a surrogate for neoantigen expression], termed 'non-progressor') or 2) immune evasion (macroscopic tumor that retained mScarlet expression, histologically confirmed adenocarcinoma; termed 'progressor') (**Figure 2D-E**).

Figure 2

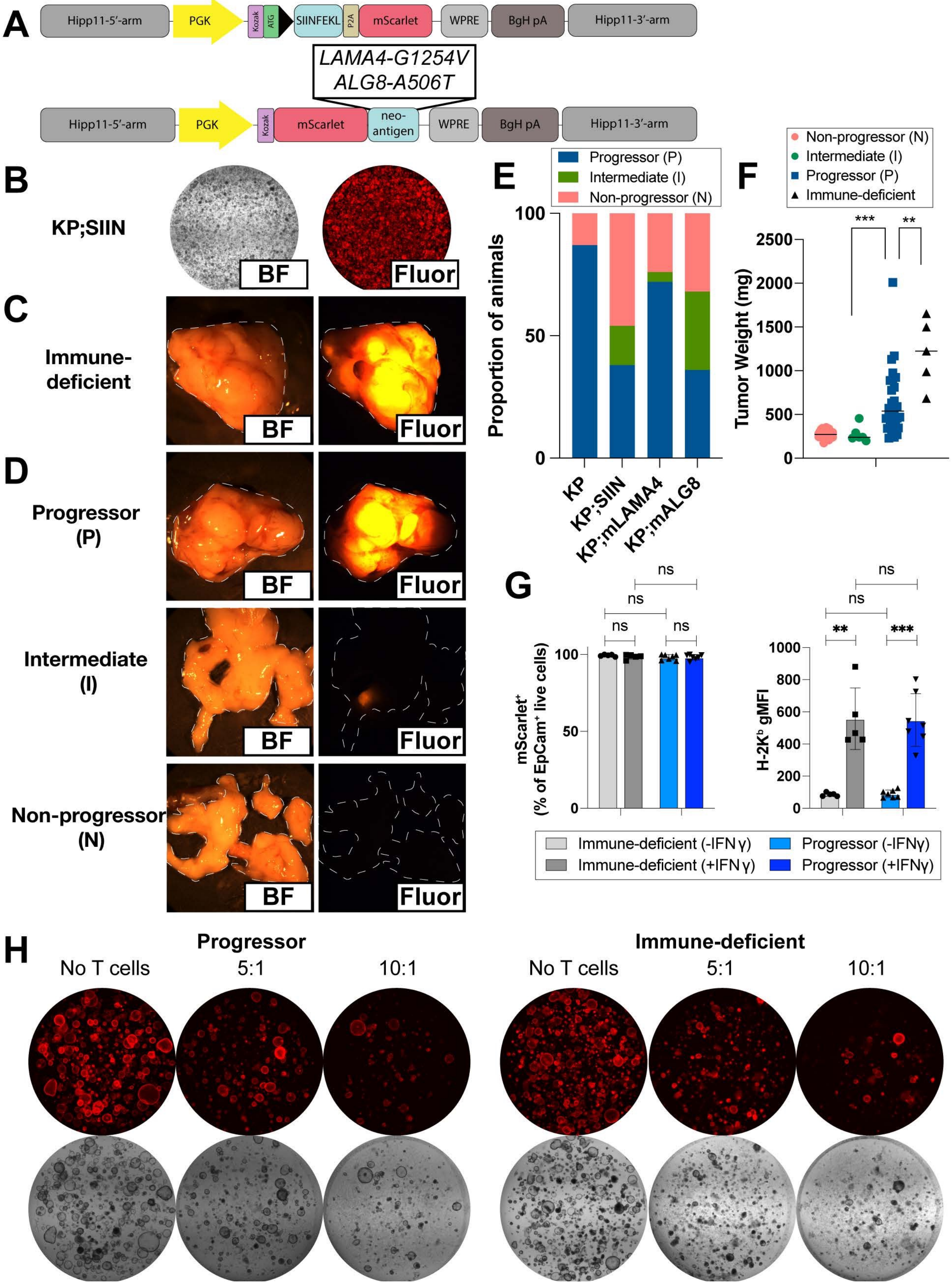

A
Hipp11-5′-arm
PGK
Kozak
ATG
SIINFEKL
P2A
mScarlet
WPRE
BgH pA
Hipp11-3′-arm
LAMA4-G1254V
ALG8-A506T
Hipp11-5′-arm
PGK
Kozak
mScarlet
neo-antigen
WPRE
BgH pA
Hipp11-3′-arm
B
KP;SIIN
BF
Fluor
C
Immune-deficient
BF
Fluor
D
Progressor (P)
BF
Fluor
Intermediate (I)
BF
Fluor
Non-progressor (N)
BF
Fluor
E
Progressor (P)
Intermediate (I)
Non-progressor (N)
Proportion of animals
100
50
0
KP
KP;SIIN
KP;mLAMA4
KP;mALG8
F
Non-progressor (N)
Intermediate (I)
Progressor (P)
Immune-deficient
***
**
Tumor Weight (mg)
2500
2000
1500
1000
500
0
G
ns
ns
ns
ns
mScarlet⁺ (% of EpCam⁺ live cells)
100
50
0
ns
ns
**
***
H-2Kᵇ gMFI
1000
800
600
400
200
0
Immune-deficient (-IFNγ)
Immune-deficient (+IFNγ)
Progressor (-IFNγ)
Progressor (+IFNγ)
H
Progressor
No T cells
5:1
10:1
Immune-deficient
No T cells
5:1
10:1

Figure 2. Neoantigen-expressing pancreatic organoids model immune clearance and immune evasion in the same tissue and antigenic context

(A) '*Hipp11-mScarletSIIN'* (top) and '*Hipp11-mScarlet-mLAMA4'* or *'Hipp11-mScarlet-mALG8'* (bottom) genomic loci after CRISPR/Cas9-assisted homology-directed repair and Cre recombination.

(B) Brightfield (left) and fluorescent (right) images of *Kras*$^{G12D/+}$*;Trp53*$^{-/-}$; *Hipp11-mScarletSIIN* ('KP;SIIN') pancreatic organoids.

(C-D) Brightfield (left) and fluorescence stereomicroscopic (right) images of representative 8-week tumors following orthotopic transplantation of neoantigen-expressing pancreatic organoids into (C) immune-deficient (CD8a-depleted or Rag2$^{-/-}$) animals or (D) immune-competent animals, depicting the range of tumor and antigenic outcomes in this context ('progressor', 'intermediate', 'non-progressor').

(E) Proportion of outcomes 5 weeks post-orthotopic transplantation (KP [n=15]; KP;SIIN [n=45]; KP;mLAMA4 [n-25]; KP;mALG8 [n=25]).

(F) Tumor/pancreas weights 8-10 weeks post-orthotopic transplantation of KP;SIIN pancreatic organoids (n=5 'immune-deficient'; n=24 'non-progressor'; n=6 'intermediate', n=30 'progressor'; horizontal bars represent median).

In addition, we observed a reproducible subset of immune-competent recipients that retained small areas of mScarlet positivity in the absence of macroscopic tumor formation (termed 'intermediate'), potentially reflective of a state of immune equilibrium (**Figure 2D-E**). In line with this hypothesis, immune-evasive tumors were on average significantly smaller than tumors that arose in the absence of an immune selective pressure ($p<0.01$), suggestive of a prior state of immune equilibrium before ultimate immune escape (**Figure 2F**). While all three neoantigens evaluated exhibit high affinity for MHC class I, we were particularly struck by the observation that a substantial portion of pancreatic tumors harboring the highly immunogenic neoantigen (SIINFEKL) were able to escape immune control while maintaining antigen expression, suggesting that further study of this 'progressor' subset could offer insights into the range of mechanisms of immune evasion employed by human PDAC.

Flow cytometric profiling demonstrated a range of CD8^{+} T cell infiltration in immune-evasive tumors, reminiscent of previous profiling efforts in human pancreatic cancer [15]. Likewise, histopathologic analysis of immune-evasive tumors revealed both inter-tumoral and intra-tumoral heterogeneity, with some areas displaying T cell exclusion, a well-documented phenomenon in PDAC [15,32], but with other areas displaying robust CD8$^+$ T cell infiltration into tumor nests (**Figure S2C-D**). Importantly, this inter- and intra-tumoral heterogeneity of T cell infiltration has been recently described in large-scale profiling efforts of the immune contexture in human pancreatic cancer [16,33]. These data suggest that immune-evasive neoantigen-expressing tumors with CD8 infiltration must either acquire defects in antigen processing/presentation and/or tumor-reactive T cells must be rendered dysfunctional over time.

To evaluate potential mechanisms of immune escape, we re-isolated PDAC organoids from both immune-evasive KP;SIIN tumors and KP;SIIN tumors that arose in immune-deficient animals for *ex vivo* characterization. A well-established mechanism of immune escape is loss or downregulation of surface MHC class I, which has recently been explored in pancreatic cancer [34,35]. After purifying the malignant compartment through Nutlin-3a selection [36], we performed flow cytometry to characterize neoantigen expression (assessed via mScarlet expression) and surface expression of MHC class I (H-2K^b, H-2D^b) and MHC class II on tumor-derived organoids (**Figure 2G and S2D-F**). Importantly, we observed no loss of neoantigen expression and equivalent surface expression of H2-K^b, when comparing organoids derived from immune-evasive tumors to organoids derived from tumors that had never been exposed to immune selective pressure (**Figure 2G**). Furthermore, these organoids retained responsiveness to interferon-γ stimulation to further upregulate surface MHC class I, suggesting that loss of neoantigen or MHC class I surface expression was not a driving factor in the observed immune evasion. However, these experiments do not rule out other defects in antigen processing/presentation.

To further establish that immune-evasive tumors retained full capacity to process and present neoantigen on their cell surface, we utilized an organoid-CD8$^+$ T cell co-culture system. Immune-evasive (progressor) or immune-deficient KP;SIIN tumor-derived organoids were co-embedded in a three-dimensional extracellular matrix with antigenically-stimulated OT-I CD8$^+$ T cells (transgenic for a TCR specific for the SIINFEKL neoantigen in the context of H-2K^b; [37]). Both sets of organoids underwent equivalent T cell-dependent killing across multiple effector:target (E:T) ratios (**Figure 2H and Supplemental Movie 1**), definitively demonstrating that organoids derived from immune-evasive tumors retain neoantigen expression and antigen processing/presentation capacity.

Neoantigen-specific CD8$^+$ T cells adopt multiple states of dysfunction in immune-evasive tumors

As we observed evidence of an ongoing CD8$^+$ T cell response in immune-evasive tumors, with retained neoantigen expression, we hypothesized that neoantigen-specific CD8$^+$ TILs had become dysfunctional in these tumors. CD8$^+$ T cell exhaustion, a state of T cell hypofunctionality [38], has been observed in both murine [39] and human PDAC [15,16,33]; however, the (neo)antigen specificity, or even tumor-reactivity, of these CD8$^+$ TILs has not been firmly established. We utilized flow cytometric profiling to assess T cell exhaustion/dysfunction within the intratumoral neoantigen-specific compartment (CD44^{hi}Tetramer$^+$). We observed no significant differences in the abundance of neoantigen-specific CD8$^+$ T cells in progressor versus non-progressor animals (**Figure 3A and S3A-B**), but consistent with the hypothesis that neoantigen-specific CD8$^+$ TILs become hypofunctional within immune-evasive tumors, we observed a decrease in their proliferative capacity (marked by Ki67) **(Figure 3B**). Furthermore, as co-expression of multiple co-inhibitory receptors is thought to distinguish a more dysfunctional phenotype from activation [40,41], we examined co-inhibitory receptor expression (PD-1, TIGIT, TIM-3, LAG-3) on neoantigen-specific TILs and observed a significant accumulation in co-expression of two or more co-inhibitory receptors in immune-evasive tumors (**Figure 3C and S3A-B**). When we compared immune-evasive (progressor) to intermediate or non-progressor animals, in addition to accumulation in progressor animals, we also observed a significant increase in the co-expression of inhibitory receptors: $PD1^+TIGIT^+$, $PD1^+TIM3^+$, and $PD1^+LAG3^+$ in intermediate animals (**Figure S3A-B**), suggesting that the acquisition of a hypofunctional phenotype may precede frank immune escape. To further investigate T cell dysfunction in these tumors, we examined $TIM3^+TCF1^{lo}$ neoantigen-specific TILs, thought to mark a terminally exhausted population [42], which were enriched exclusively in immune-evasive tumors (**Figure S3A-B**). However, we also observed a small, but reproducible, population of $PD1^+TCF1^{hi}$ "progenitor-like" neoantigen-specific TILs (**Figure S3A-B**), suggesting that immune-evasive tumors harbor a population of neoantigen-specific TILs with potential for re-invigoration [42–44].

Figure 3

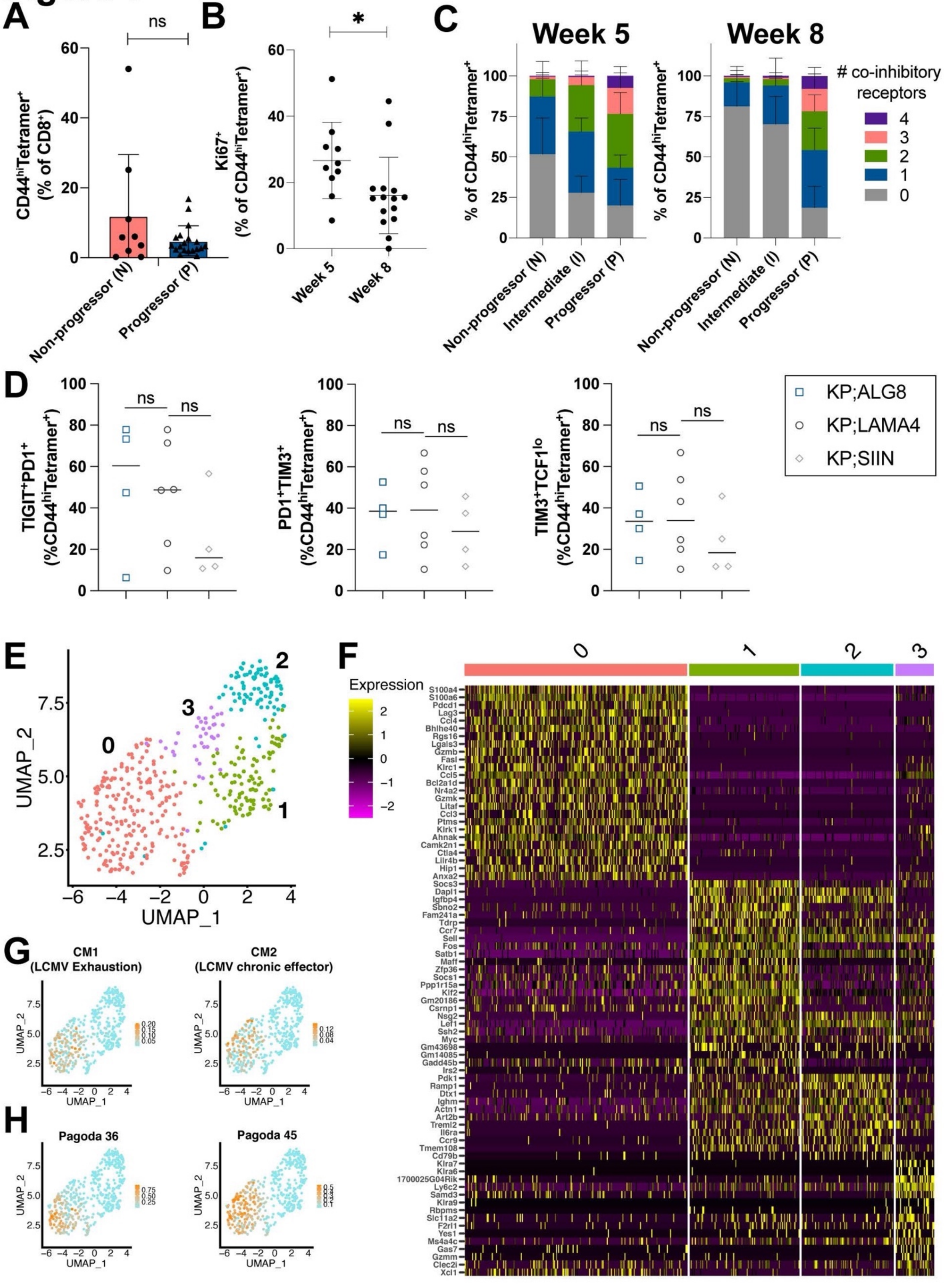

A
ns
60
40
20
0
CD44^{hi}Tetramer$^+$ (% of CD8$^+$)
Non-progressor (N)
Progressor (P)
B
*
Ki67$^+$ (% of CD44^{hi}Tetramer$^+$)
Week 5
Week 8
C
Week 5
Week 8
% of CD44^{hi}Tetramer$^+$
100
75
50
25
0
Non-progressor (N)
Intermediate (I)
Progressor (P)
# co-inhibitory receptors
4
3
2
1
0
D
TIGIT$^+$PD1$^+$ (%CD44^{hi}Tetramer$^+$)
PD1$^+$TIM3$^+$ (%CD44^{hi}Tetramer$^+$)
TIM3$^+$TCF1^{lo} (%CD44^{hi}Tetramer$^+$)
ns
ns
100
80
60
40
20
0
KP;ALG8
KP;LAMA4
KP;SIIN
E
UMAP_2
UMAP_1
7.5
5.0
2.5
−6
−4
−2
0
2
4
0
1
2
3
F
Expression
2
1
0
−1
−2
S100a4
S100a6
Pdcd1
Lag3
Ccl4
Bhlhe40
Rgs16
Lgals3
Gzmb
Fasl
Klrc1
Ccl5
Bcl2a1d
Nr4a2
Gzmk
Litaf
Ccl3
Ptms
Klrk1
Ahnak
Camk2n1
Ctla4
Lilr4b
Hip1
Anxa2
Socs3
Dapl1
Igfbp4
Sbno2
Fam241a
Tdrp
Ccr7
Sell
Fos
Satb1
Maff
Zfp36
Socs1
Ppp1r15a
Klf2
Gm20186
Csrnp1
Nsg2
Lef1
Ssh2
Myc
Gm43698
Gm14085
Gadd45b
Irs2
Pdk1
Ramp1
Dtx1
Ighm
Actn1
Art2b
Treml2
Il6ra
Ccr9
Tmem108
Cd79b
Klra7
Klra6
1700025G04Rik
Ly6c2
Samd3
Klra9
Rbpms
Slc11a2
F2rl1
Yes1
Ms4a4c
Gas7
Gzmm
Clec2i
Xcl1
G
CM1 (LCMV Exhaustion)
CM2 (LCMV chronic effector)
0.20
0.15
0.10
0.05
0.12
0.08
0.04
H
Pagoda 36
Pagoda 45
0.75
0.50
0.25
0.5
0.4
0.3
0.2
0.1

Figure 3. T cell exhaustion typifies the neoantigen-specific TIL response in immune-evasive PDAC

(A) Flow cytometric quantification of neoantigen-specific ($CD44^{hi}Tetramer^{+}$) TILs at 9-10 weeks post-initiation.
(B) $Ki67^{+}$ population of $CD44^{hi}Tetramer^{+}CD8^{+}$ TILs from progressor tumors at indicated times following tumor initiation (scatter plot showing mean +/- SD).
(C) Co-inhibitory receptor (PD-1, TIGIT, TIM-3, LAG-3) co-positivity as indicated by color in $CD44^{hi}Tetramer^{+}CD8^{+}$ TILs from ‘non-progressor’, ‘intermediate’, or ‘progressor’ tumors/pancreata.
(D) Flow cytometric characterization of neoantigen-specific ($CD44^{hi}Tetramer^{+}$) TILs at 5 weeks post-initiation in the respective neoantigen-expressing models.
(E) UMAP projection of scRNA-seq of neoantigen-specific ($CD8^{+}CD44^{hi}Tetramer^{+}$) TILs from immune-evasive tumors.
(F) Heatmap of differentially expressed genes between clusters with selected genes highlighted.
(G) UMAP projections overlaid with gene module expression for “LCMV T cell exhaustion” (CM1) and “LCMV T cell chronic effector” (CM2).
(H) UMAP projections overlaid with PAGODA gene expression programs (Pagoda36, Pagoda45).
Statistical analyses: (A,B,D) two-sided Mann-Whitney U test (n.s. P=non-significant, * $P<0.05$, ** $P<0.01$, *** $P<0.001$).
See also Figure S3 and Table S2.

To confirm that the observed T cell phenotypes were not unique to the SIINFEKL neoantigen, we performed flow cytometric immunophenotyping on neoantigen-specific $CD8^{+}$ TILs from immune-evasive KP;SIIN, KP;mLAMA4, and KP;mALG8 tumors. We observed similar patterns of T cell exhaustion within the neoantigen-specific $CD8^{+}$ TIL compartment in all three neoantigen-expressing models (**Figure 3D**).

To further elucidate potential mechanisms of immune evasion, we performed single-cell RNA-seq [45] on intratumoral neoantigen-specific $CD8^{+}$ T cells sorted from immune-evasive KP;SIIN tumors. After quality control filtering, we retained 447 neoantigen-specific TILs, clustered them into 4 distinct clusters and identified genes that are differentially expressed between cells in the four clusters (**Figure 3E-F and Table S2**). Consistent with our flow cytometric characterization, cells in the largest cluster (cluster 0) had higher expression of numerous genes associated with CD8$^+$ T cell exhaustion (*Pdcd1, Havcr2, Lag3, Tox*) (**Figure 3F**). We then scored the cells for gene modules derived from mouse CD8$^+$ T cells in defined cell states from both acute and chronic lymphocytic choriomeningitis virus (LCMV) [46] and B16 melanoma [47]. In line with our previous observations, cluster 0 was enriched for 'T cell exhaustion' (CM1), but intriguingly was also enriched for a 'chronic effector' signature (CM2) (**Figure 3G**). Cells in two smaller clusters (clusters 1 and 2), showed higher expression of markers of naïve/memory CD8$^+$ T cells (*Sell, Ccr7, Klf2, Tcf7*), potentially reflecting one or more aberrant memory-like cell states, and those in another small cluster (cluster 3) had higher expression of inhibitory Ly49 receptors (*Klra6, Klra7*) (**Figure 3F**), thought to mark a subset of CD8$^+$ T regulatory cells previously described in both autoimmunity [48,49] and cancer [47]. Pathway and Gene Set Overdispersion Analysis (PAGODA) [50] derived three *de novo* gene set signatures from our scRNA-seq data that overlaid clusters 1 and 2 (Pagoda30) and cluster 0 (Pagoda36, Pagoda45) (**Figure 3H and Table S2**), further highlighting the heterogeneity within the neoantigen-specific CD8$^+$ TIL compartment.

We next compared CD44^{hi}Tetramer$^+$ (antigen-experienced, neoantigen-specific) to CD44^{hi}Tetramer^{neg} (antigen-experienced, NOT SIINFEKL-specific) CD8$^+$ TILs by flow cytometry, and observed that, as expected, the CD44^{hi}Tetramer$^+$ subset exhibited a significantly higher proportion of dysfunctional/exhausted TILs (**Figure S3C**), but intriguingly noted that a portion of CD44^{hi}Tetramer^{neg} CD8$^+$ TILs also exhibit marks of T cell dysfunction/exhaustion. This may reflect non-SIINFEKL tumor-reactivity (tumor-associated antigen or non-SIINFEKL neoantigen) or potentially bystander effects. We noted that TIGIT$^+$PD1$^+$ co-positivity best differentiated the neoantigen-specific compartment in these analyses, suggesting that these immune axes may play an outsized role in T cell dysfunction in this disease context.

Human PDAC harbors analogous populations of exhausted intratumoral $CD8^+$ T cells

To investigate if these observations were generalizable to human PDAC, we employed two parallel approaches. First, we obtained pancreatic adenocarcinoma specimens immediately following surgical resection and isolated intratumoral $CD8^+$ T cells for flow cytometric profiling. Of 13 specimens evaluated, 9 had sufficient $CD8^+$ TILs for further immunophenotyping (range 202-17,895 live $CD8^+$ TILs). In line with previous reports [15], the majority (67-99%) of $CD8^+$ TILs were $CD45RO^+$ (**Figure 4A**), reflective of prior antigen-experience, with a substantial portion of intratumoral $CD8^+$ T cells co-expressing multiple co-inhibitory receptors ($PD1^+TIGIT^+$, $PD1^+LAG\text{-}3^+$, $PD1^+TIM\text{-}3^+$, $TIGIT^+TIM\text{-}3^+$) (**Figure 4B**), consistent with T cell exhaustion [51] and in line with our preclinical profiling. Indeed, we observed a population of $TIM3^+TCF1^{lo}$ $CD8^+$ TILs in the majority of tumors (**Figure 4C**), suggestive of terminal exhaustion [42–44]. As $PD1^+TCF1^{hi}$ ("progenitor-like") $CD8^+$ TILs have been demonstrated to underlie the proliferative burst in response to PD-1 blockade and are thought to be a primary driver of clinical responses to PD-1 blockade [42,43], we examined PD-1 and TCF1 co-expression in PDAC TILs. We observed $PD1^+TCF1^{hi}$ $CD8^+$ PDAC TILs in the majority of tumors, but these represented a small subset of $CD8^+$ TILs (**Figure 4D**), potentially in line with the observed lack of clinical benefit using PD-(L)1 blockade as monotherapy in PDAC [5–7]. However, we also observed a population of $HLA\text{-}DR^+Ki67^+CD57^{neg}$ $CD8^+$ TILs in the majority of tumors (**Figure 4E**), suggesting that there are recently-activated, proliferating, and non-senescent intratumoral $CD8^+$ T cells in PDAC with potential for therapeutic reinvigoration.

Figure 4

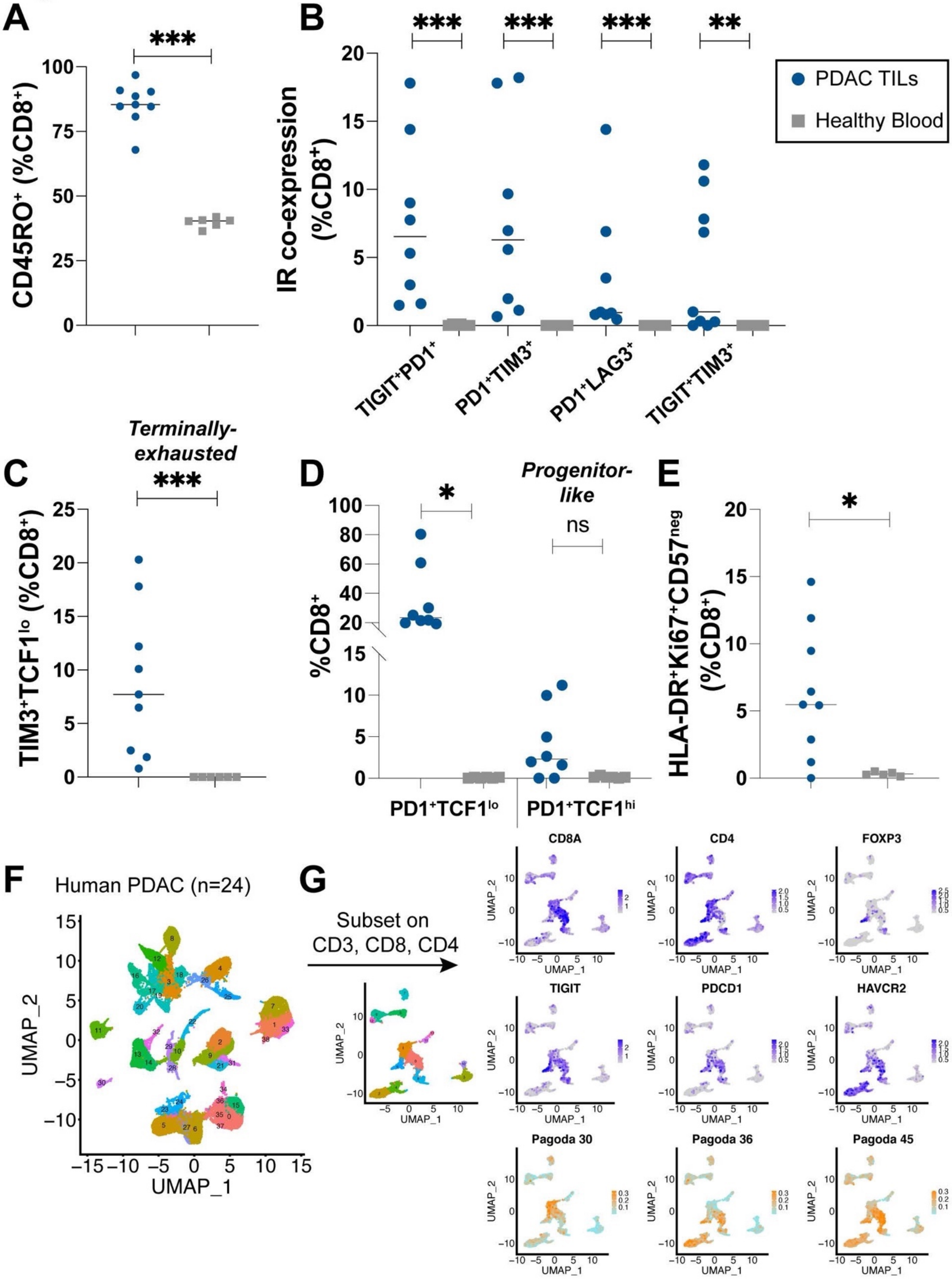

A
***
CD45RO⁺ (%CD8⁺)
100
75
50
25
0
B
***
***
***
**
IR co-expression (%CD8⁺)
20
15
10
5
0
TIGIT⁺PD1⁺
PD1⁺TIM3⁺
PD1⁺LAG3⁺
TIGIT⁺TIM3⁺
PDAC TILs
Healthy Blood
C
Terminally-exhausted
***
TIM3⁺TCF1ˡᵒ (%CD8⁺)
25
20
15
10
5
0
D
*
Progenitor-like
ns
%CD8⁺
100
80
60
40
20
15
10
5
0
PD1⁺TCF1ˡᵒ
PD1⁺TCF1ʰⁱ
E
*
HLA-DR⁺Ki67⁺CD57ⁿᵉᵍ (%CD8⁺)
20
15
10
5
0
F
Human PDAC (n=24)
UMAP_2
UMAP_1
G
Subset on CD3, CD8, CD4
CD8A
CD4
FOXP3
TIGIT
PDCD1
HAVCR2
Pagoda 30
Pagoda 36
Pagoda 45

Figure 4. Human PDAC harbors exhausted CD8$^+$ TILs

(A) Flow cytometric profiling of human PDAC CD8$^+$ TILs and CD8$^+$ T cells from healthy peripheral blood for CD45RO expression.

(B) Quantification of co-inhibitory receptor (TIGIT, PD-1, TIM-3, LAG-3) co-expression on CD8$^+$ TILs from human PDAC resections compared to healthy peripheral blood.

(C) "Terminally exhausted" TIM3$^+$TCF1^{lo} CD8$^+$ TILs in human PDAC compared to healthy peripheral blood.

(D) PD-1$^+$TCF1^{lo} and "progenitor-like" PD-1$^+$TCF1^{hi} CD8$^+$ TILs in human PDAC compared to healthy peripheral blood.

(E) HLA-DR$^+$Ki67$^+$CD57^{neg} CD8$^+$ TILs in human PDAC compared to healthy peripheral blood.

(F) UMAP projection of scRNA-seq data from human PDAC patients (n=24) (Peng et al, 2019).

(G) Computationally sorted cell subsets and UMAP projections overlaid with indicated genes/signatures.

Statistical analyses: (A-E) two-sided Mann-Whitney U test (n.s. P=non-significant, ** P <0.01, *** P <0.001). (A-E: scatter plots showing median +/- SD).

See also Figure S4.

We next investigated T cell immunophenotypes in previously reported single-cell RNA-seq of human PDAC patients (n=24) [52] (**Figure 4F and S4B-C)**. We computationally 'sorted' CD3$^+$CD8$^+$ and CD3$^+$CD4$^+$ cells from this scRNA-seq dataset, and then examined the expression in these selected cells of T cell exhaustion-associated genes (encoding PD-1, TIGIT, TIM-3), as well as expression of *de novo* gene signatures derived from murine neoantigen-specific TILs (**Figure 4F-G**). All three *de novo* gene signatures (Pagoda30, 36, 45) from murine neoantigen-specific TIL profiling were expressed in specific subsets of CD8$^+$ TILs in human PDAC (**Figure 4G**). Collectively, these data support the conclusion that CD8$^+$ TILs in human PDAC are largely characterized by markers of T cell exhaustion, with analogous signatures to neoantigen-specific TILs from murine PDAC, suggesting that these preclinical models accurately recapitulate a subset of human disease.

Elevated expression of PVR/CD155 in murine and human PDAC

Given the clear association with neoantigen-specific T cell exhaustion/dysfunction in murine immune-evasive tumors and the presence of analogous populations in human PDAC, we next examined inhibitory ligand expression. We evaluated protein expression of PD-L1 (the ligand for PD-1), Galectin 9 (the ligand for TIM-3), and CD155 (the ligand for TIGIT) using immunohistochemical analysis of murine immune-evasive PDAC tissue microarrays (TMAs). While we detected occasional PD-L1 and/or Galectin 9 positivity, the vast majority of tumor cells were negative or expressed low levels of these inhibitory ligands (**Figure S5A**). This is consistent with prior reports [15,53] that have shown human PDAC to be largely devoid of PD-L1 positivity within the malignant compartment. In contrast, elevated CD155 expression was observed on the majority of murine immune-evasive tumors (**Figure 5A and S5A**). Likewise, using immunohistochemical staining of human PDAC (n=36), CD155 was found to be elevated in a substantial fraction of cases (**Figure 5B**). We computed H-scores [54] for CD155 in both murine and human PDAC and found that CD155 expression was significantly higher in tumor samples compared to healthy pancreas controls (**Figure 5A-B**), although it is important to point out that approximately 20% of human tumors showed no CD155 expression. While we observed elevated CD155 expression, we cannot rule out the contribution of additional inhibitory ligands such as CD112, Ceacam1, FGL-1, and others, that were not directly assessed in this analysis.

Figure 5

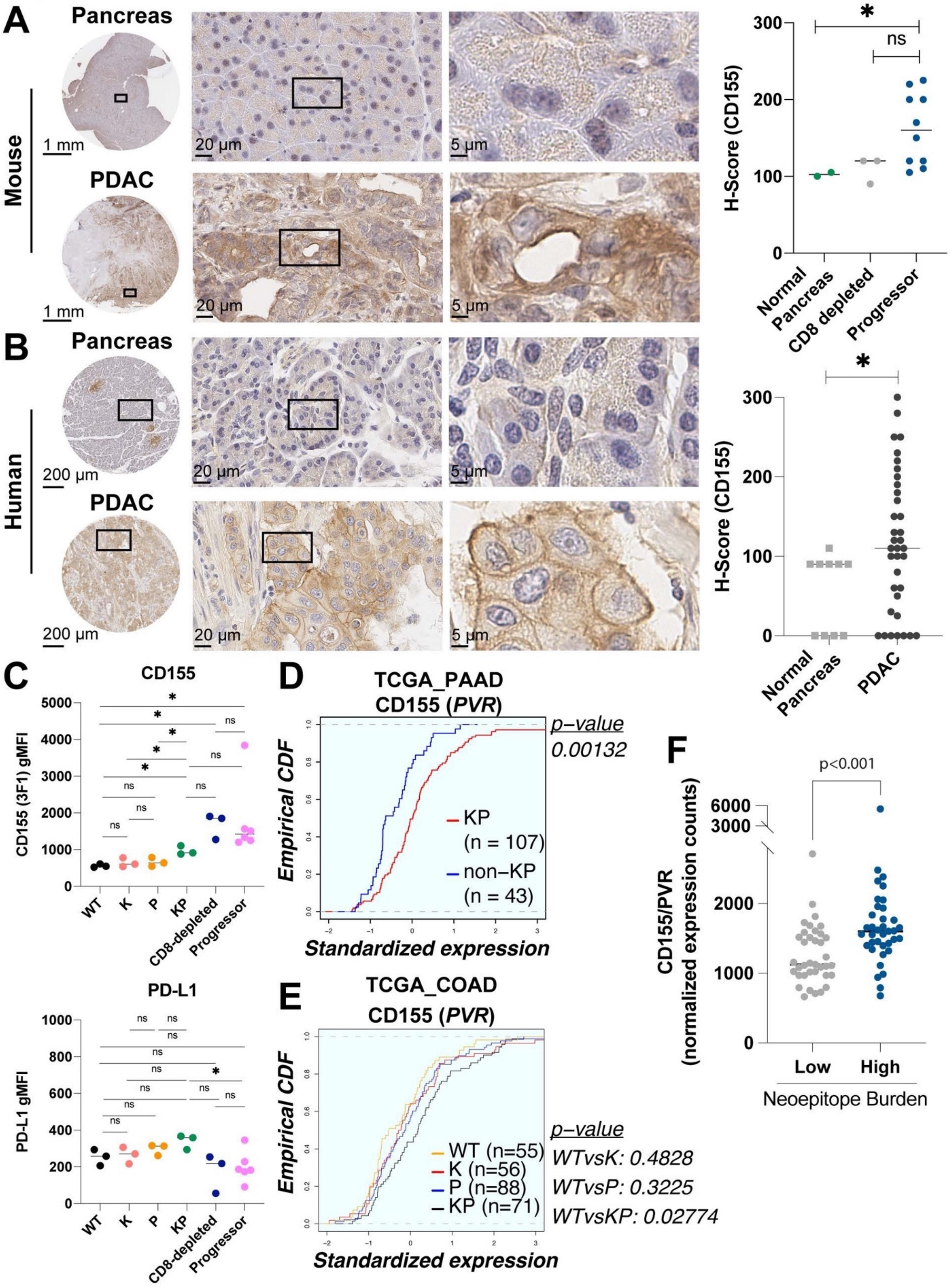

A
Pancreas
PDAC
Mouse
1 mm
20 µm
5 µm
H-Score (CD155)
300
200
100
0
*
ns
Normal Pancreas
CD8 depleted
Progressor
B
Pancreas
PDAC
Human
200 µm
20 µm
5 µm
H-Score (CD155)
*
Normal Pancreas
PDAC
C
CD155
CD155 (3F1) gMFI
5000
4000
3000
2000
1000
0
*
ns
WT
K
P
KP
CD8-depleted
Progressor
PD-L1
PD-L1 gMFI
1000
800
600
400
200
0
D
TCGA_PAAD
CD155 (PVR)
p–value
0.00132
Empirical CDF
KP (n = 107)
non–KP (n = 43)
Standardized expression
E
TCGA_COAD
CD155 (PVR)
Empirical CDF
WT (n=55)
K (n=56)
P (n=88)
KP (n=71)
p–value
WTvsK: 0.4828
WTvsP: 0.3225
WTvsKP: 0.02774
Standardized expression
F
p<0.001
CD155/PVR (normalized expression counts)
6000
3000
2000
1000
0
Low
High
Neoepitope Burden

Figure 5. Elevated CD155 expression within the malignant compartment in murine and human PDAC
(A) Immunohistochemical analysis of CD155 (anti-mouse CD155; LS-C376428) on murine PDAC TMA, quantified by H-score (right).
(B) Immunohistochemical analysis of CD155 (anti-human CD155; EPR22672-151) on human PDAC TMA, quantified by H-score (right).
(C) Flow cytometric assessment of surface expression of CD155 (top) or PD-L1 (bottom) on genetically-defined (WT, K, P, KP) or tumor-derived organoids (from progressor or immune-deficient animals) (scatter plots, horizontal bar represents median).
(D-E) ECDF analysis of *PVR* expression in PAAD (D) or COAD (E) within indicated genetic cohorts.
(F) CD155 (*PVR*) expression in TCGA PAAD patients stratified by total neoepitope burden (high: top 25%, low: bottom 25% from Figure 1A)
Statistical analyses: A-B: two-sided Mann-Whitney U test (n.s. P=non-significant, * $P < 0.05$), C: Welch's t-test (n.s. P=non-significant, * $P < 0.05$), D-F: Kolmogorov-Smirnov (KS).
See also Figure S5.

To assess inhibitory ligand expression specifically within the malignant compartment, we utilized our panel of murine tumor-derived organoids that had been purified using Nutlin-3a selection and human PDAC patient-derived organoids. Consistent with immunohistochemical results, we observed elevated surface expression of CD155 on patient-derived PDAC organoids (**Figure S5C**) and murine tumor-derived organoids using two independent monoclonal antibodies (**Figure 5C**). In contrast, we observed low level basal expression of surface PD-L1 on murine tumor-derived organoids (**Figure 5C**), although PD-L1 surface expression could be induced following treatment with 10 ng/mL interferon-γ (**Figure S5D**). Notably, we observed elevated CD155 expression on organoids isolated from immune-evasive tumors, as well as organoids derived from tumors never exposed to an immune selective pressure, suggesting that CD155 upregulation may be a common feature of pancreatic tumorigenesis that is co-opted for immune evasion rather than an acquired feature during immune escape.

As CD155 expression appeared to be a characteristic of pancreatic tumorigenesis, we further investigated the genetic basis for this upregulation. To assess the impact of major genetic driver events on CD155 expression, we derived an allelic series of isogenic pancreatic organoids from *C57BL/6J* mice: wild-type (WT), *Kras*$^{LSLG12D/+}$ (K), *Trp53*$^{flox/flox}$ (P), *Kras*$^{LSLG12D/+}$*Trp53*$^{flox/flox}$ (KP), which were isolated from normal pancreata of genetically engineered animals and *ex vivo* recombined with adenoviral Cre recombinase. Using these genetically-defined pancreatic organoids (GDO), we observed that CD155 surface expression was unchanged upon expression of either oncogenic Kras or loss of p53 alone, but was significantly increased in the presence of concomitant oncogenic Kras expression and p53 loss (**Figure 5C**), suggesting a possible interplay between these canonical PDAC-associated oncogenic events leading to upregulation of CD155 expression.

To extend this observation, we assessed mRNA expression of *PVR* (encoding CD155), *CD274* (encoding PD-L1), *PVRL2* (encoding CD112), *PDCD1LG2* (encoding PD-L2), and *LGALS9* (encoding Galectin 9) within PDAC patient samples from TCGA stratified based on *KRAS* and/or *TP53* mutational status. In line with our observations in murine PDAC, we observed significantly elevated expression of CD155 (*PVR*) in PDAC harboring both oncogenic *KRAS* and *TP53* mutations/loss ('KP') compared to samples that were wild-type for either of these genes ('non-KP') (**Figure 5D**). We did not observe significant differences in other inhibitory ligand expression between these patient cohorts (**Figure S5E**). Of note, given the near universal presence of *KRAS* mutations in PDAC, we were unable to isolate the effects of *KRAS* mutation from *TP53* mutation/loss in this disease context. To evaluate whether a similar paradigm extends beyond PDAC and to attempt to isolate the effects of each oncogenic event, we stratified TCGA datasets from lung adenocarcinoma (LUAD) and colon (COAD) based on *KRAS* and/or *TP53* mutational status. While in lung adenocarcinoma, mutation of either *KRAS* or *TP53* alone were associated with elevated expression of CD155 (*PVR*) (**Figure S5F**), only the combination of oncogenic *KRAS* and *TP53* mutation/loss (KP) was associated with elevated CD155 (*PVR*) expression in colon cancer (**Figure 5E**), highlighting potential tissue-specific differences in CD155 (*PVR*) regulation.

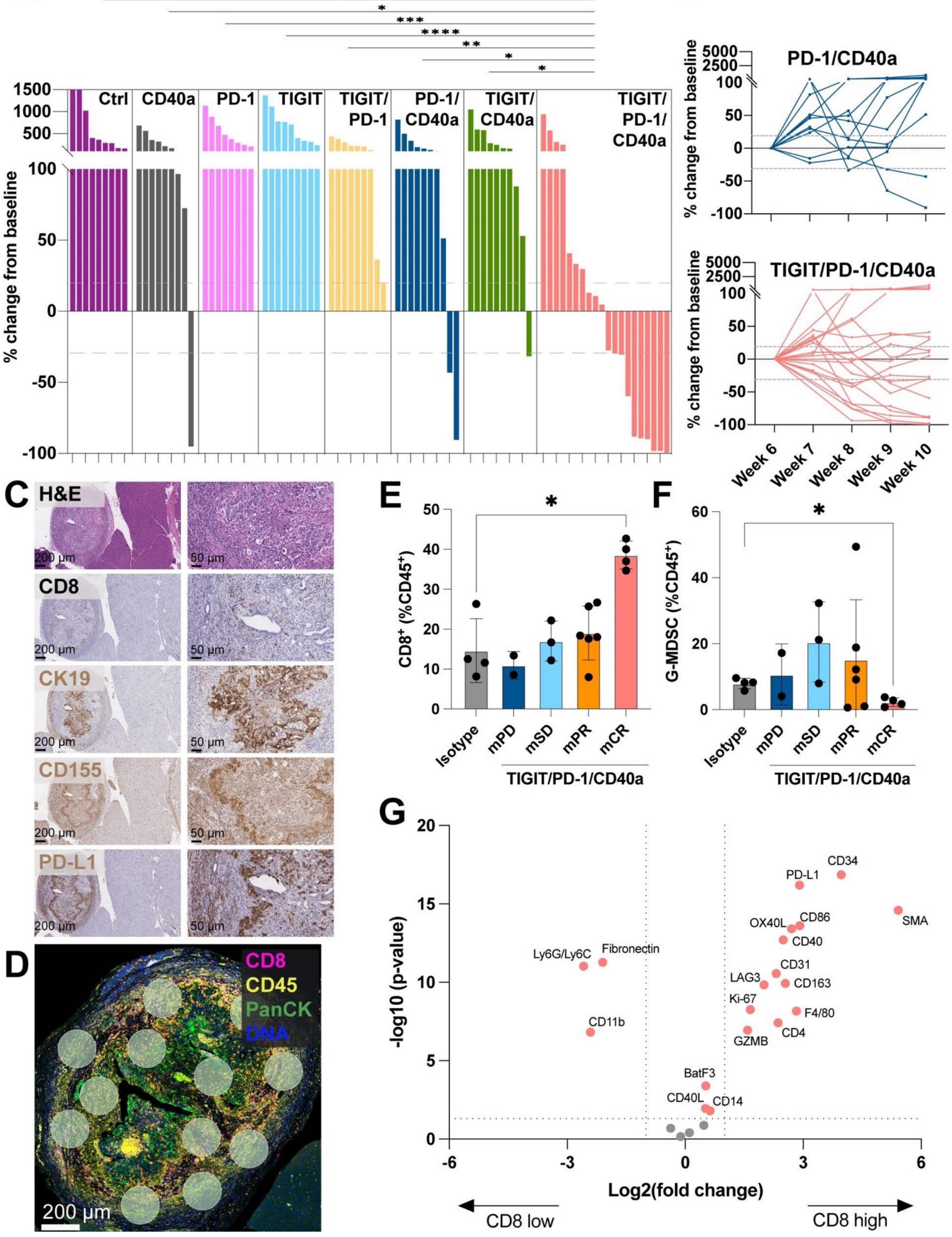
To further investigate the role of the CD155/TIGIT axis in human pancreatic cancer immune evasion, we leveraged our neoepitope prediction pipeline (**Figure 1**) to stratify patients into quartiles by overall neoepitope burden (<500 nM) and then queried CD155 expression. Tumors with a high burden (top quartile) of predicted neoepitopes exhibited on average a significantly elevated CD155 expression (**Figure 5F**). Likewise, as our earlier analyses (**Figure 1**) demonstrated that a substantial portion of human PDAC harbors neoepitopes with high agretopicity (non-binding to binding analysis) or high affinity for MHC class I (<50nM), characteristics that correlate with increased immunogenic potential [55], we queried whether a higher burden of neoepitopes within these classes is also associated with elevated CD155 expression. Analogous to total neoepitope burden, we observed significantly elevated CD155 expression in patients with a high burden of "non-binding to binding" neoepitopes (**Figure S5F**) and in patients with predicted high affinity neoepitopes (**Figure S5G**), further supporting a possible role for the CD155/TIGIT axis in human pancreatic cancer.

Preclinical activity of TIGIT/PD-1/CD40a combination immunotherapy in neoantigen-expressing PDAC

Next, we set out to investigate the relevance of the CD155/TIGIT axis in a therapeutic context. As T cell dysfunction is associated with both chronic antigen stimulation and suboptimal costimulation [51], we reasoned that the combination of a CD40 agonist plus rationally-guided immune checkpoint blockade (ICB) might be able to overcome neoantigen-specific T cell dysfunction in these tumors. CD40 is known to be expressed on the surface of antigen presenting cells (APCs) and is crucial for mediating the crosstalk between APCs and T cells [17,56]. Agonistic CD40 antibodies can substitute for CD154 (the natural ligand for CD40) and bypass the need for $CD4^+$ T cell help [57] and importantly, can enhance anti-tumor responses with immune checkpoint blockade [39,58–60]. Given accumulating evidence that the CD155/TIGIT axis likely plays a central role in PDAC [16], we opted to evaluate combination regimens targeting this axis. Importantly, the PD-1/PD-L1 and TIGIT/CD155 axes coordinately function to dampen a productive $CD8^+$ T cell response and co-blockade has demonstrated synergy in preclinical models and in early-stage clinical trials [61–63].

Following orthotopic transplantation of KP;SIIN organoids into immune-competent animals and confirmation of tumor establishment and progression, animals were randomized by baseline tumor volume to receive mono- or dual-immune checkpoint blockade (ICB), in the presence or absence of CD40 agonism at 6 weeks post-initiation. Specifically, animals were allocated to an isotype control arm or therapeutic arms (CD40 agonist, anti-PD-1, anti-TIGIT, anti-TIGIT/PD-1, anti-PD-1 + CD40 agonist, anti-TIGIT + CD40 agonist, anti-TIGIT + anti-PD-1 + CD40 agonist) for treatment over 4 weeks and tumors were longitudinally evaluated through high-resolution ultrasound imaging. Tumor response was determined according to the more stringent modified RECIST criteria [64], previously validated for volumetric tumor response in preclinical models. As expected, isotype control-treated tumors grew unabated with 0% objective response rate (ORR; mCR+mPR) and 0% disease control rate (DCR; mCR+mPR+mSD) ([n=15], **Figure 6A and S6A**). Consistent with clinical observations, mono- or dual-immune checkpoint blockade (PD-1, TIGIT, or PD-1/TIGIT) exhibited no tumor responses (0% ORR/DCR for monotherapy; 0% ORR and 22% DCR with TIGIT/PD-1 co-blockade; n=9-10 per arm) (**Figure 6A and Figure S6A**). While CD40 agonist (CD40a) monotherapy resulted in an 11% ORR (33% DCR, [n=9]), the majority of animals quickly progressed through monotherapy. When CD40a was combined with either PD-1 blockade or TIGIT blockade, we observed primarily stabilization of disease with few tumor responses (9% ORR; 54% DCR with PD-1/CD40a; 0% ORR; 18% DCR with TIGIT/CD40a; [n=11 per arm], **Figure 6A-B**), consistent with the early clinical promise of CD40a/PD-1 combinations currently being evaluated in clinical trials [65]. In contrast to all other combinations investigated, TIGIT/PD-1 co-blockade plus CD40 agonism produced significant tumor responses (46% ORR, 71% DCR) with 23% complete responses (mCR); [n=48], **Figure 6A-B**). These data support the hypothesis that combinatorial strategies to simultaneously boost and reinvigorate an anti-tumor immune response are needed to overcome the profoundly immunosuppressive PDAC microenvironment and furthermore, that TIGIT blockade may overcome pre-existing or acquired resistance to CD40a/PD-1 therapy. Notably, all therapeutic arms were well tolerated as assessed by body status and animal weight. As anti-TIGIT antibodies have already demonstrated safety/tolerability in human patients, with hints of efficacy in other tumor types [63,66,67], and combination CD40 agonism + PD-1 blockade has shown early-clinical promise in PDAC [58], combination TIGIT/PD-1/CD40a is poised for rapid clinical evaluation.

Figure 6



A
***
*
***
****
**
*
*
1500
1000
500
100
50
0
-50
-100
% change from baseline
Ctrl
CD40a
PD-1
TIGIT
TIGIT/ PD-1
PD-1/ CD40a
TIGIT/ CD40a
TIGIT/ PD-1/ CD40a
B
PD-1/CD40a
TIGIT/PD-1/CD40a
5000
2500
100
50
0
-50
-100
% change from baseline
Week 6
Week 7
Week 8
Week 9
Week 10
C
H&E
CD8
CK19
CD155
PD-L1
200 µm
50 µm
D
CD8
CD45
PanCK
DNA
200 µm
E
*
CD8⁺ (%CD45⁺)
50
40
30
20
10
0
Isotype
mPD
mSD
mPR
mCR
TIGIT/PD-1/CD40a
F
*
G-MDSC (%CD45⁺)
60
40
20
0
Isotype
mPD
mSD
mPR
mCR
TIGIT/PD-1/CD40a
G
20
15
10
5
0
-log10 (p-value)
CD34
PD-L1
SMA
OX40L
CD86
CD40
Ly6G/Ly6C
Fibronectin
CD31
LAG3
CD163
Ki-67
F4/80
CD11b
GZMB
CD4
BatF3
CD40L
CD14
-6
-3
0
3
6
Log2(fold change)
CD8 low
CD8 high

Figure 6. TIGIT/PD-1/CD40a combination immunotherapy elicits anti-tumor responses in immune-evasive PDAC

(A) Waterfall plot of evaluable tumors after 4 weeks of indicated control or experimental treatments.

(B) Spider plots of treatment response to PD-1/CD40a (top) and TIGIT/PD-1/CD40a (bottom) during 4-week dosing period.

(C) Representative immunohistochemical analysis of CD8a, cytokeratin 19 (CK19), CD155, and PD-L1 of responder tumors (mPR/mCR) following 28 days of TIGIT/PD-1/CD40a treatment.

(D) Representative mfIHC with AOIs for Nanostring GeoMx DSP of a responder tumor (mPR) following 28 days of TIGIT/PD-1/CD40a treatment.

(E-F) Flow cytometric analysis of (E) $CD8^+$ T cells, and (F) G-MDSCs [$CD45^+CD11b^+F4\_80^{low}Ly6C^{low}Ly6G^{high}$] following the indicated treatments.

(G) Differential protein expression in 'CD8 high' versus 'CD8 low' AOIs in non-responder tumors following 28 days of TIGIT/PD-1/CD40a treatment. Red: FDR < 0.05.

Statistical analyses: (A) two-sided Mann-Whitney U test of percent change at 4 weeks of therapy, (E-F) two-sided Mann-Whitney U test, (G) linear mixed effect model with Benjamini-Hochberg FDR (n.s. P=non-significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$).

See also Figure S6.

To investigate the molecular mechanisms of effectiveness and resistance to TIGIT/PD-1/CD40a combination therapy, we treated animals harboring immune-evasive tumors with TIGIT/PD-1/CD40a as described above. Tumors were tracked using weekly ultrasound imaging, to facilitate assignment of experimental treatment animals into responder (mPR/mCR), stable (mSD), or non-responder (mPD). Following 28 days of treatment, pancreatic tumors (or remaining pancreatic tissue in the case of complete response) were harvested for flow cytometric profiling, traditional immunohistochemical analysis, or spatially-resolved multiplex protein profiling using Nanostring GeoMx Digital Spatial Profiling (DSP).

We observed abundant intratumoral $CD8^+$ T cells in responder animals (**Figure 6C-D**), with a less pronounced $CD8^+$ infiltrate with clear areas of T cell exclusion in non-responder animals (**Figure S6B-C**). Immunohistochemical analysis also demonstrated elevated expression of PD-L1 and CD155 within the tumor-adjacent stroma of responder tumors (**Figure 6C**), potentially reflecting a mechanism of acquired resistance to CD40 agonism, which may be overcome through TIGIT/PD-1 co-blockade.

To further characterize changes in the immune microenvironment following TIGIT/PD-1/CD40a therapy, we performed flow cytometric immunophenotyping of $CD45^+$ immune cell subsets following 28 days of control (isotype) or experimental (TIGIT/PD-1/CD40a) therapy. In line with histopathologic characterization, we observed an increase in $CD8^+$ T cell infiltration into responder (mCR/mPR) tumors and a concomitant decrease in immunosuppressive myeloid subsets, most strikingly in granulocytic myeloid-derived suppressor cells (G-MDSCs) (**Figure 6E-F**). We also observed a significant decrease in overall myeloid infiltration ($CD11b^+$) and monocytic MDSCs (M-MDSCs), but not in tumor-associated macrophages (TAMs) following TIGIT/PD-1/CD40a (**Figure S6E-H**).

To more deeply profile the tumor immune microenvironment, we employed NanoString GeoMx Digital Spatial Profiling, which utilizes oligonucleotide-tagged antibodies containing a photocleavable linker and UV illumination of defined areas of interest (AOIs) to enable spatially-resolved 40-plex protein labeling of tumors following TIGIT/PD-1/CD40a therapy. Consistent with our prior analyses, we observed abundant intratumoral $CD8^+$ T cells in responder animals, with $CD8^+$ T cells largely restricted to the periphery of non-responder tumors. In both responder and non-responder tumors, we observed markers of T cell activation/co-stimulation (Granzyme B, Ki67) in areas of high $CD8^+$ T cell infiltration, but observed elevated expression of MDSC markers (CD11b, Ly6G/C) in areas of T cell exclusion uniquely in non-responder tumors (**Figure 6G and Figure S6C-D**), suggesting a potential role for MDSC-mediated T cell exclusion in resistance to therapy.

Collectively, these data support CD8$^+$ T cell re-invigoration as a key marker of response and suggest that T cell exclusion, possibly mediated by G-MDSCs, may be a key driver of resistance. However, future studies will be needed to uncover potential biomarkers in tumors prior to treatment that can be used to select patient most likely to benefit from therapy.

The CD155/TIGIT axis is sufficient to promote immune evasion in PDAC

As an orthogonal approach to our organoid-based preclinical model, we adapted retrograde pancreatic duct delivery [68] to generate a genetically-tractable autochthonous mouse model of neoantigen-expressing PDAC. Specifically, we engineered the lentiviral vector used to initiate PDAC in *Kras$^{LSL\text{-}G12D/+}$;Trp53$^{fl/fl}$* (KP) animals to additionally encode expression of a defined neoantigen (‘mScarletSIIN’; OVA$_{257–264}$ [SIINFEKL] and OVA$_{323–339}$) (**Figure 7A-B**). Retrograde ductal instillation of Cre-expressing lentivirus led to histologically confirmed pancreatic intraepithelial neoplasia (PanIN) and/or pancreatic adenocarcinoma (PDAC) formation in ~90% of immune-deficient or immune-competent animals by 9 weeks post-initiation (**Figure S7A**). To examine the effects of neoantigen expression in autochthonous PDAC, we performed parallel surgeries in immune-competent and immune-deficient (using CD8a-depleting antibodies) KP animals. Analogous to Cre expression alone, ~90% of immune-deficient animals transduced with lentivirus expressing mScarletSIIN developed histologically confirmed PanIN/PDAC by 9 weeks post-initiation (**Figure S7A**). Importantly, tumors that developed in the absence of CD8$^+$ T cells retained neoantigen expression (mScarlet positivity) within PanIN/PDAC lesions (**Figure 7C and 7E**). In contrast, approximately 50% of immune-competent animals initiated with mScarletSIIN failed to develop tumors by 9 weeks post-initiation (**Figure 7D-E and S7B**), consistent with observations of immune clearance using orthotopic transplantation of neoantigen-expressing organoids (**Figure 2**). However, unlike our organoid-based model, we observed a subset of animals (~40%) that developed macroscopic tumors that failed to maintain mScarlet expression (assessed by both fluorescence stereomicroscopy and immunohistochemical analysis), suggestive of immune editing (**Figure 7D and S7D**). This difference can likely be attributed site-specific effects of stochastic lentiviral integration in the autochthonous model, compared to expression of a neoantigen from a safe harbor locus in our organoid-based model. Lastly, while less frequent than in our organoid-based model, a reproducible subset (7-20%) of immune-competent animals initiated with mScarletSIIN developed immune-evasive tumors (**Figure 7D-E and S7C**). While we observed a robust neoantigen-specific CD8$^+$ TIL response in both early-stage and late-stage lesions (**Figure S7F-G**), immune-evasive autochthonous tumors harbored intratumoral neoantigen-specific CD8$^+$ T cells with co-expression of multiple co-inhibitory receptors (including PD-1$^+$TIGIT$^+$), suggestive of T cell exhaustion (**Figure S7G**), analogous to observations in our organoid-based model system and in human PDAC.

Figure 7

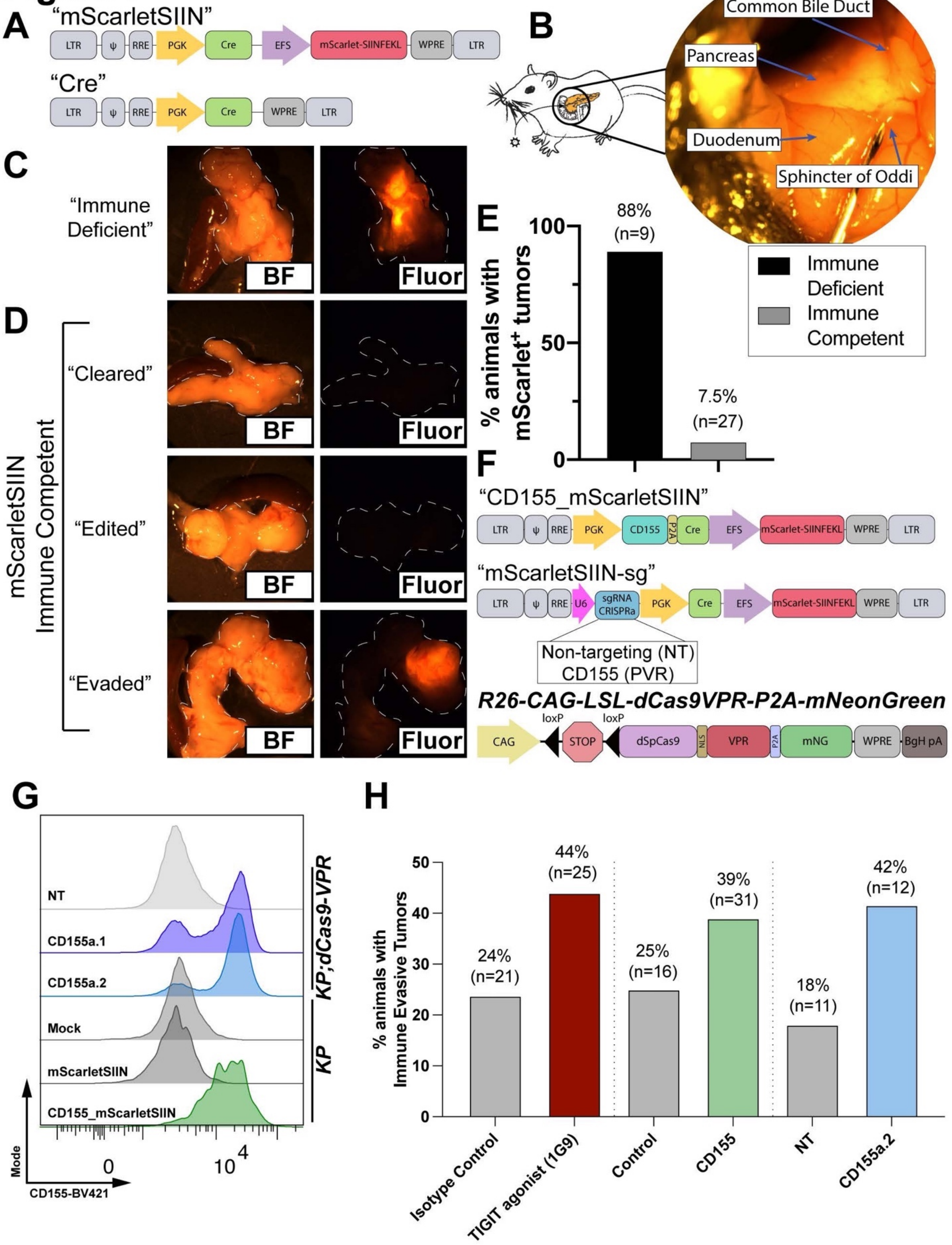

A
"mScarletSIIN"
LTR
ψ
RRE
PGK
Cre
EFS
mScarlet-SIINFEKL
WPRE
LTR
"Cre"
LTR
ψ
RRE
PGK
Cre
WPRE
LTR
B
Common Bile Duct
Pancreas
Duodenum
Sphincter of Oddi
C
"Immune Deficient"
BF
Fluor
D
mScarletSIIN Immune Competent
"Cleared"
"Edited"
"Evaded"
BF
Fluor
E
% animals with mScarlet+ tumors
100
50
0
88% (n=9)
7.5% (n=27)
Immune Deficient
Immune Competent
F
"CD155_mScarletSIIN"
LTR
ψ
RRE
PGK
CD155
P2A
Cre
EFS
mScarlet-SIINFEKL
WPRE
LTR
"mScarletSIIN-sg"
LTR
ψ
RRE
U6
sgRNA CRISPRa
PGK
Cre
EFS
mScarlet-SIINFEKL
WPRE
LTR
Non-targeting (NT)
CD155 (PVR)
R26-CAG-LSL-dCas9VPR-P2A-mNeonGreen
loxP
loxP
CAG
STOP
dSpCas9
NLS
VPR
P2A
mNG
WPRE
BgH pA
G
NT
CD155a.1
CD155a.2
Mock
mScarletSIIN
CD155_mScarletSIIN
KP;dCas9-VPR
KP
Mode
CD155-BV421
0
10^4
H
% animals with Immune Evasive Tumors
50
40
30
20
10
0
24% (n=21)
44% (n=25)
25% (n=16)
39% (n=31)
18% (n=11)
42% (n=12)
Isotype Control
TIGIT agonist (1G9)
Control
CD155
NT
CD155a.2

Figure 7. Elevated CD155/TIGIT signaling is sufficient to promote immune evasion in autochthonous PDAC
(A) Lentiviral vectors used to generate autochthonous neoantigen-expressing PDAC or control.
(B) Retrograde pancreatic ductal instillation of lentivirus.
(C) Brightfield (left) and fluorescence stereomicroscopic (right) images of representative 9-week autochthonous tumors generated using mScarletSIIN in CD8α-depleted animals.
(D) Brightfield (left) and fluorescence stereomicroscopic (right) images of representative outcomes ("cleared", "edited", "evaded") in mScarletSIIN immune-competent animals.
(E) Proportion of animals with mScarlet-positivity as assessed by fluorescence stereomicroscopy at 9 weeks post-initiation.
(F) Lentiviral vectors and R26-dCas9-VPR knock-in allele used to modulate CD155 (*PVR*) expression in autochthonous PDAC.
(G) Flow cytometric assessment of surface CD155 expression on pancreatic organoids following transduction with indicated lentiviruses.
(H) Proportion of animals with mScarlet-positivity as assessed by fluorescence stereomicroscopy at 9-12 weeks post-initiation following indicated genetic or pharmacologic modulation.
See also Figure S7.

We took advantage of both the relative rarity of immune-evasive tumors and the genetic tractability of this model to evaluate whether genetic or pharmacologic modulation of the CD155/TIGIT axis could promote immune evasion. To investigate the effect of tumor-specific inhibitory ligand expression, we re-engineered the mScarletSIIN lentivirus to additionally express the cDNA for *Pvr* (encoding CD155) (**Figure 7F**). Using lentiviral transduction of KP pancreatic organoids, we confirmed expected upregulation of surface CD155 (**Figure 7G**). As an orthogonal approach, we utilized CRISPR-activation (CRISPRa) to overexpress CD155 from the endogenous *Pvr* locus. To facilitate efficient *in vivo* CRISPRa, we generated a knock-in allele at the *Rosa26* safe harbor locus [69] to enable Cre-mediated conditional expression of dCas9-VPR [70] linked to mNeonGreen to track recombined cells (**Figure 7F**). Using lentiviral transduction of KP;dCas9-VPR organoids, we validated two sgRNAs targeting the promoter of *Pvr*, both of which upregulated surface CD155 to similar levels as cDNA expression (**Figure 7G**). One of the two sgRNAs tested (*Pvra.2*) consistently led to more uniform upregulation of surface CD155, and thus we opted to proceed with this sgRNA for *in vivo* evaluation.

To evaluate the effects of tumor-restricted inhibitory ligand overexpression in autochthonous neoantigen-expressing PDAC, we randomized KP animals to receive retrograde ductal instillation of control (mScarletSIIN) or CD155-mScarletSIIN lentiviruses. In parallel, KP;dCas9-VPR animals were instilled with mScarletSIIN lentivirus additionally encoding either a non-targeting (NT) control sgRNA or an sgRNA targeting *Pvr* (Pvra.2). Using either cDNA-based or CRISPRa-based overexpression, tumor-specific CD155 upregulation resulted in an increase in the proportion of immune evasion. Specifically, *Pvr* cDNA elicited 39% [n=31] mScarlet$^+$ tumors (an increase from 25% [n=16] in control animals) and Pvra.2 CRISPRa led to 42% [n=12] immune evasion (an increase from 18% [n=11] in control animals) (**Figure 7H**), suggesting that forced elevation of CD155 promotes immune evasion in PDAC.

Finally, to assay the effect of elevated TIGIT activity, we initiated autochthonous PDAC in KP animals using mScarletSIIN and randomized animals immediately following retrograde pancreatic duct surgery to receive a TIGIT agonistic antibody (clone 1G9) [71] or an isotype control antibody at the same dosing schedule. In line with our observations modulating CD155, 44% [n=25] animals in the TIGIT agonist arm exhibited immune evasion, compared to 24% [n=21] isotype control-treated animals (**Figure 7H**). While no approach was as effective as complete CD8 depletion, collectively these data reinforce the functional importance of the CD155/TIGIT axis in PDAC immune evasion.

Discussion

Neoantigen-specific CD8$^+$ T lymphocytes, which recognize cognate antigen presented in the context of MHC class I, are thought to underlie the success of current immune-based strategies [72]. Our present study bolsters the finding that the majority of patients with mismatch repair-proficient pancreatic cancer harbor predicted high affinity MHC class I-restricted neoepitopes, even in advanced/metastatic disease. Additionally, by expanding the search space for potential neoepitopes to include in-frame insertion/deletion and frameshift mutations, we demonstrate that the landscape of potentially immunogenic neoepitopes in PDAC can be substantially increased. Although our neoepitope predictions do not directly assess immunogenicity, our findings are in line with previous reports that have identified endogenous neoantigen-reactive CD8$^+$ TILs from a subset of human PDAC patients [12].

Taking advantage of insights garnered from our neoepitope profiling, we generated multiple orthogonal preclinical models of neoantigen-expressing pancreatic cancer and demonstrate that PDAC undergoes all three phases of immunosurveillance [73], with a subset of animals successfully evading immune clearance despite continued tumor-specific expression of a high affinity neoantigen. The preclinical models presented herein bear similarities to the recently reported KPC-OG model [20], with a number of key distinctions. While a genetically-encoded autochthonous neoantigen model avoids the potential contribution of wound healing responses associated with both orthotopic transplantation and retrograde pancreatic duct delivery of lentiviruses, the preclinical models we employed avoid widespread antigen expression during either embryonic pancreas development or adulthood, potential confounders in pancreas-specific *Cre* recombinase driven models [20]. Additionally, the preclinical models presented in this study more readily isolate the neoantigen-specific CD8$^+$ T cell contribution to immune evasion, compared to model systems that utilize full length OVA, which contains CD8$^+$ T cell, CD4$^+$ T cell, and B cell epitopes [20]. Lastly, both orthotopic transplantation of pancreatic organoids and retrograde pancreatic duct delivery of lentiviruses offer the flexibility and genetic tractability to interrogate new and diverse neoantigen(s) and leverage CRISPR-mediated gene perturbations to rapidly evaluate specific biological hypotheses. While the incorporation of defined neoantigens facilitates the tracking and immunophenotyping of tumor-reactive CD8$^+$ T cells, it is important to note that these are not the only potential (neo)antigens present in these tumors. In fact, mScarlet, which is used as a surrogate readout of neoantigen expression in these models, is a foreign protein and may also contribute MHC class I- and class II-restricted neoantigens. Furthermore, there may be additional mutations accumulated during pancreatic tumorigenesis. Future studies will be needed to evaluate the functional consequences of varying affinity neoantigens and defined MHC class II neoantigens in pancreatic cancer.

Using single-cell RNA-seq and flow cytometric profiling of neoantigen-expressing immune-evasive murine PDAC, we uncovered multiple classes of CD8$^+$ TILs with markers of dysfunction and identify similar populations of intratumoral CD8$^+$ T cells in human PDAC resection specimens, suggesting that these preclinical models accurately recapitulate a subset of human PDAC. While both murine and human PDAC prominently feature CD8$^+$ TILs with markers of dysfunction, we also observed non-terminally exhausted CD8$^+$ TILs and evidence of an ongoing intratumoral immune response. However, it is important to point out that in the case of human PDAC, we were unable to assess the tumor-reactivity of these populations. We likewise demonstrate the presence of "progenitor-like" [38,42–44] PD1$^+$TCF1^{hi} CD8$^+$ TILs in human and murine PDAC, the latter of which are found within the neoantigen-specific TIL compartment, suggesting the potential for reinvigoration with immune checkpoint blockade. The near complete lack of clinical benefit provided by PD-(L)1-directed immune checkpoint blockade in human PDAC, an observation that is accurately recapitulated in our preclinical models, suggests that PDAC may employ additional molecular and/or cellular mechanisms of immune evasion that serve to limit the anti-tumor immune response.

Our characterization of the neoantigen-specific immune response has functionally implicated the co-inhibitory receptor TIGIT, and its cognate ligand CD155, as a critical axis driving PDAC immune evasion. We demonstrate that CD155 is upregulated on the surface of murine and human PDAC tumor cells, both *in vivo* and *ex vivo*. As *PVR*, the gene encoding CD155, has been reported to be upregulated by oncogenic *KRAS* in cell culture [74,75], it is tempting to speculate that the CD155/TIGIT axis might represent a critical immune checkpoint in additional KRAS-driven tumors. Our data support this and further points to a potential synergy between *KRAS* and *TP53* mutations to upregulate CD155 expression, further refining the complex regulation of this inhibitory ligand. However, the molecular mechanisms underlying this synergy remain to be identified. We also demonstrate that the co-inhibitory receptor TIGIT is expressed on a subset of human and murine TILs, and in the latter case, further delineate that tumor-reactive (i.e., neoantigen-specific) $CD8^+$ TILs express high levels of TIGIT. Tumor-specific overexpression of CD155 in neoantigen-expressing autochthonous pancreatic cancer leads to an increased proportion of immune-evasive tumors, and these results can be recapitulated using an agonistic TIGIT antibody. Thus, increased signaling flux through the CD155/TIGIT axis is sufficient to promote immune evasion in PDAC. In addition to $CD8^+$ T cells, TIGIT is also expressed on regulatory T cells and natural killer (NK) cells [62,76,77] and TIGIT-mediated ligation of CD155 expressed on the surface of antigen presenting cells (APCs) has been demonstrated to impact dendritic cell maturation and consequent T cell priming [76]. Future studies will be needed to carefully dissect the contributions of TIGIT and CD155 expression and signaling within these various cell compartments and their roles in PDAC immune evasion.

A number of scenarios have been proposed to explain how PDAC evades the anti-tumor immune response, and previous reports have implicated almost every step in the cancer-immunity cycle [78]. PDAC may bypass immune surveillance through loss of MHC class I surface expression [34], exclusion of CD8$^+$ T cells [15,32,79], induction of dysfunctional T cell programs [15,16], deficiencies in type I conventional dendritic cells [20,80], and/or recruitment of immune suppressive cell populations [32,81,82]. However, the majority of these prior studies have been unable to isolate the effects of tumor and/or microenvironmental perturbations on the neoantigen-specific immune response. While our data support a crucial role for neoantigen-specific T cell dysfunction and the CD155/TIGIT axis in PDAC immune evasion, it is likely that multiple facets of the tumor-immunity cycle are disrupted in PDAC. Indeed, we observed intratumoral areas of T cell exclusion within immune-evasive tumors in our model systems, suggesting that this may also contribute to immune escape in a subset of animals. Additionally, beyond the CD155/TIGIT axis, we uncovered multiple states of dysfunction within the neoantigen-specific TIL compartment, which can be functionally interrogated in future studies.

Immune modulation has emerged as a promising therapeutic strategy for numerous tumor types. However, it is likely that tissue of origin, histologic subtype and/or genetic alterations might dictate disparate mechanisms of immune evasion [17,83] that call for rationally-guided combinatorial strategies. Prior efforts have reinforced the notion that approaches to both boost the endogenous anti-tumor immune response, whether through vaccination or CD40 agonism, and prevent T cell exhaustion are likely necessary in PDAC [17]. Indeed, both therapeutic vaccination [84] and pharmacologic [85–87] strategies to boost the endogenous anti-tumor immune response are being investigated. Likewise, CD40 agonism, which licenses antigen-presenting cells to promote priming of T lymphocytes [56] has been extensively evaluated in preclinical PDAC models [39,59,60,88] and combination therapy using CD40 agonism, PD-1 blockade, and cytotoxic chemotherapy has demonstrated early-stage clinical promise in human PDAC patients [58]. Using multi-arm, randomized and blinded preclinical trials, we demonstrate that TIGIT/PD-1 co-blockade in combination with CD40 agonism can reinvigorate an effective anti-tumor immune response in a subset of animals with immune-evasive PDAC. Mechanistically, we observed increased expression of inhibitory ligands following CD40 agonist containing combination immunotherapy. While these tumors displayed elevated CD155 expression within the malignant compartment at baseline, we observed increased non-tumor PD-L1 and CD155 expression following treatment. These results suggest that the PD-1/PD-L1 and CD155/TIGIT axes may represent nonredundant mechanisms of acquired resistance to CD40 agonist-based therapies and further support strategies that leverage co-blockade of these inhibitory axes. In addition, our profiling efforts point to a potential role for MDSC-mediated T cell exclusion as a mechanism of resistance to TIGIT/PD-1/CD40a. Future studies in additional preclinical models of pancreatic cancer will evaluate the requirement for high affinity neoantigens in mediating this response and directly address additional combination strategies to overcome these resistance mechanisms. While our profiling of the neoantigen-specific immune response in PDAC nominates additional immune checkpoints for future preclinical evaluation, combinatorial targeting of TIGIT/PD-1/CD40a represents a particularly promising approach for rapid clinical translation.

Acknowledgments

We thank the entire Jacks laboratory with specific thanks to M. Burger, N. Sacks, and D. Canner for helpful discussions and technical assistance. We thank K. Yee, J. Teixeira, K. Anderson, M. Magendantz for administrative support. We thank A. Aguirre and B. Wolpin for providing raw data and mutation calls for the DFCI-PancSeq cohort. We thank E. Miller for helpful discussions regarding NanoString GeoMx profiling. This work was supported by the Howard Hughes Medical Institute, NCI Cancer Center Support Grant P30-CA1405, the Lustgarten Foundation Pancreatic Cancer Research Laboratory at MIT, DFHCC SPORE in Gastrointestinal Cancer Career Enhancement Award (W.F.P.), the Stand Up To Cancer-Lustgarten Foundation Pancreatic Cancer Interception Translational Cancer Research Grant (Grant Number: SU2C-AACR-DT25-17, W.F.P, T.J.) and the Stand Up To Cancer Golden Arrow Early Career Scientist Award (GA-6182, W.F.P.). Stand Up To Cancer is a program of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, the scientific partner of SU2C. We thank the Koch Institute Swanson Biotechnology Center for technical support, specifically the Flow Cytometry, Histology, Preclinical Modeling, Imaging & Testing and Integrative Genomics & Bioinformatics core facilities.

Supplementary material: All supplementary figures, table, and data are available in our publication at *Cancer Cell*.

Methods

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Tyler Jacks (tjacks@mit.edu).

Materials Availability

Plasmids generated in this study will be deposited to Addgene. Mouse lines and organoid lines generated in this study will made available to the broader scientific community upon request to the Lead Contact.

Data and Code Availability

The datasets generated and/or analyzed during the current study are available in the NCBI Gene Expression Omnibus (GEO) under accession number: GSE163059. Computer code for neoepitope predictions and scRNA-seq analysis will be made available upon request. Other software tools (including version numbers) for exome, RNA-seq, and scRNA-seq analyses are listed in the Key Resource Table.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

All animal studies described in this study were approved by the MIT Institutional Animal Care and Use Committee. All animals were maintained on a pure *C57BL/6J* genetic background. Generation of $Kras^{LSL\text{-}G12D/+}$ and $Trp53^{flox/flox}$ (KP) mice has previously been described [89,90]. *OT-I* TCR transgenic mice have been previously described [37]. $KP;Rosa26^{CAG\text{-}LSL\text{-}dCas9\text{-}VPR\text{-}P2A\text{-}mNeonGreen}$ mice were generated as part of this study (described in detail below).

mESC generation and CRISPR-assisted targeting

"KP*1", a *C57BL/6J Kras*$^{LSL\text{-}G12D/+;}$ *Trp53*$^{flox/flox}$ (KP) murine embryonic stem cell (mESC) line, was generated by crossing a hormone-primed *C57BL/6J Trp53*$^{flox/flox}$ female with a *C57BL/6J Kras*$^{LSL\text{-}G12D/+}$*; Trp53*$^{flox/flox}$ male. At 3.5 days post-coitum, blastocysts were flushed from the uterus, isolated, and cultured on a mouse embryonic fibroblast (MEF) feeder layer in 'ESCM+LIF+2i' [Knockout DMEM (Gibco), 15% FBS (Hyclone), 1% NEAA (Sigma), 2 mM Glutamine (Gibco), 0.1 mM β-mercaptoethanol (Sigma-Aldrich) 50 IU Penicillin, 50 IU Streptomycin, 1000 U/ml LIF (Amsbio), 3 μM CHIR99021 (AbMole), 1 μM PD0325901(AbMole)]. After 5-7 days in culture the outgrown inner cell mass was isolated, trypsinized and re-plated on a fresh MEF layer. ES cell lines were genotyped for *Kras*$^{LSL\text{-}G12D/+}$*;Trp53*$^{flox/flox}$, and *Zfy* (Y-chromosome specific). Primer sequences available upon request. ES cell lines were tested for pluripotency by injection into host blastocysts from albino mice to generate chimeric mice.

DNA mixes (1:1 mix of 'U6-sgH11-eCas9-T2A-BlastR' + 'H11-mScarletSIIN targeting vector' or 1:1 mix of 'U6-sgR26-eCas9-T2A-BlastR' + 'R26-dCas9-VPR targeting vector') were ethanol precipitated prior to DNA (1 μg) transfection of approximately $3*10^5$ KP*1 mESCs in a gelatin-coated 24-well plate using Lipofectamine 2000 (Thermo Fisher) according to the manufacturer instructions. mESCs were selected with Blasticidin (6 μg/mL) for 2 days, starting 36 hours post-transfection, prior to low-density re-plating on MEF feeder lines in absence of Blasticidin. Large mESC colonies were manually picked using a stereomicroscope, expanded and evaluated for correct integration using PCR with primers spanning both the 5' and 3' homology arms (primer sequences available on request). Correct clones by PCR evaluation were evaluated using Southern blot analysis. Briefly, genomic DNA was digested overnight with NsiI-HF (for H11-mScaretSIIN targeting) or PacI (for R26-dCas9-VPR targeting). Digestions were electrophoresed on 0.7% agarose gels and blotted to Amersham Hybond XL nylon membranes (GE Healthcare). Samples were probed with 32P-labeled 5' external, 3' external, and internal probes applied in Church buffer [91] (probe sequences available on request).

Correctly targeted clones were injected into albino *C57BL/6J* blastocysts. Chimerism was assessed by coat color. Pancreatic organoids were isolated from chimeric animals and "donor" organoids were purified from the host pancreas using 72 hours of Puromycin (6 μg/mL) selection (leveraging the presence of the Puromycin resistance gene within the LSL cassette upstream of *Kras-G12D*) [89].

Organoid generation and characterization

Pancreatic organoid isolation and propagation has been previously described [22]. Briefly, for genetically-defined pancreatic organoids, pancreata were manually dissected from genetically-engineered mice of the desired genotype. Pancreata were then manually minced with razor blades and dissociated in pancreas digestion buffer [1x PBS, 125 U/mL collagenase IV (Worthington)] for 20 minutes at 37°C. Cell suspensions were filtered through 70 μm filters, washed with 1x PBS and centrifuged with slow deceleration. Cell pellets were resuspended in 100% growth-factor reduced Matrigel (Corning) and solidified at 37°C. Cells were subsequently cultured in organoid complete media (minor modifications from previously described formulations [22] (see details below) and monitored for organoid outgrowth. Organoids were passaged with TrypLE Express (Life Technologies) for at least 4 passages to purify the ductal component prior to Cre recombinase-mediated recombination. For recombination, organoids were spinfected with adenoviral (Ad5-CMV-Cre) at a MOI >100 to ensure 100% recombination. All organoids were authenticated by genotyping at *Kras* and *Trp53* loci both prior to and following Ad-CMV-Cre to ensure proper recombination. See **Table S3** for details on organoid lines.

Murine and human tumor-derived organoids were isolated following the same procedure as above with the exception of 30 minutes in pancreas digestion buffer. Tumor-derived organoids were passaged at least four times prior to experimental manipulation to remove contaminating cell types. P53 deficient organoids were selected via resistance to Nutlin-3a (10 µM, Sigma-Aldrich). Pancreatic organoids were maintained in culture for <20 passages.

Media for pancreatic organoids was formulated based on L-WRN cell conditioned media (L-WRN CM) [92]. Briefly, L-WRN CM was generated by collecting 8 days of supernatant from L-WRN cells, grown in Advanced DMEM/F12 (Gibco) supplemented with 20% fetal bovine serum (Hyclone), 2 mM GlutaMAX, 100 U/mL of penicillin, 100 µg/mL of streptomycin, and 0.25 µg/mL amphotericin. L-WRN CM was diluted 1:1 in Advanced DMEM/F12 (Gibco) and supplemented with additional RSPO-1 conditioned media (10% v/v), generated using Cultrex HA-R-Spondin1-Fc 293T Cells. The following molecules were also added to the growth media: B27 (Gibco), 1 µM N-acetylcysteine (Sigma-Aldrich), 10 µM nicotinamide (Sigma-Aldrich), 50 ng/mL EGF (Novus Biologicals), 500 nM A83-01 (Cayman Chemical), 10 µM SB202190 (Cayman Chemical), and 500 nM PGE2 (Cayman Chemical). Wnt activity of the conditioned media was assessed and normalized between batches via luciferase reporter activity of TCF/LEF activation (Enzo Leading Light Wnt reporter cell line).

T cell culture

OT-I splenocytes were harvested from *C57BL/6J OT-I* transgenic mice, and spleens were mashed through 70 µm filters. Red blood cells were lysed with ACK buffer for 2 min before cell suspension neutralization with PBS and pelleted for plating. Splenocytes were counted and adjusted to $1*10^6$ cells/mL in T cell medium [RPMI 1640 (Corning) supplemented with 10% heat-inactivated FBS, 20 mM HEPES (Gibco), 1 mM Sodium Pyruvate (Thermo Fisher), 2 mM L-

Glutamine (Gibco), 50 μM β-mercaptoethanol (Gibco), 1X Non-Essential Amino Acids (Sigma), 0.5X Pen/Strep (Gibco) with 10 ng/mL hIL-2 (Peprotech) and 1 μM SIINFEKL peptide (Anaspec)]. Splenocytes were activated for 24h at 37°C in a tissue culture incubator, before manual CD8a isolation according to manufacturer instructions (Milteny Biotec). *OT-I* T cells were subsequently expanded 4-6 days in T cell medium with 10 ng/mL hIL-2 prior to organoid co-culture.

Organoid + CD8 T cell co-culture

Pancreatic organoids were dissociated using TrypLE Express (Life Technologies) and single cell suspensions were generated by vigorous resuspension. Activated *OT-I* $CD8^+$ T cells (see above) and organoid cell numbers were determined by manual hemocytometer cell counting, and T cells + organoids were mixed at defined effector:target (E:T) ratios. Matrigel was then added (5 μL per well in black-walled 96-well plates (Corning) for Incucyte live cell imaging; 20-50 uL per well for culture in 24-well plates; final 85% Matrigel) before solidification at 37°C. Cells were cultured in complete organoid medium supplemented with 10 ng/mL hIL-2 (Peprotech). Incucyte images of co-cultures were acquired every 4 hours (Brightfield and RFP channels) for 6-10 days for Incucyte live cell imaging or imaged at Day 5-7 for larger cultures.

Orthotopic transplantation

Orthotopic transplantation of organoids was performed with minor modifications to previously reported protocols for orthotopic transplantation of pancreatic monolayer cell lines [93]. Briefly, animals were anesthetized using Isoflurane, the left subcostal region was depilated (using clippers or Nair) and the surgical area was disinfected with alternating Betadine/Isopropyl alcohol. A small (~2 cm) skin incision was made in the left subcostal area and the spleen was visualized through the peritoneum. A small incision (~2 cm) was made through the peritoneum overlying the spleen and the spleen and pancreas were exteriorized using ring forceps. A 30-gauge needle was inserted into the pancreatic parenchyma parallel to the main pancreatic artery and 100 μL (containing $1.25*10^5$ organoid cells in 50% PBS + 50% Matrigel) was injected into the pancreatic parenchyma. Successful injection was visualized by formation of a fluid-filled region within the pancreatic parenchyma without leakage. The pancreas/spleen were gently internalized and the peritoneal and skin layers were sutured independently using 5-0 vicryl sutures. All mice received pre-operative analgesia with Bup-SR and were followed post-operatively for any signs of discomfort or distress. Organoid/Matrigel mixes were kept on ice throughout the entirety of the procedure to prevent solidification prior to injection. For orthotopic transplantation, syngeneic *C57BL/6J* mice (aged 4-12 weeks) were transplanted. Male pancreatic organoids were only transplanted back into male recipients.

Small rodent ultrasound

Quantification of murine pancreatic tumors by high resolution ultrasound has been previously described [94]. Briefly, animals were anesthetized using Isoflurane and the lateral and ventral abdominal areas were depilated using Nair. Sterile 0.9% saline (1 mL) was administered by i.p. injection prior to imaging to improve visualization of the pancreas. Animals were imaged using the Vevo3100/LAZRX ultrasound and photoacoustic imaging system (Fujifilm-Visualsonics). Animals were placed on the imaging platform in the supine position and a layer of ultrasound gel was applied over the entirety of the abdominal area. The ultrasound transducer (VisualSonics 550S) was placed on the abdomen orthogonal to the plane of the imaging platform. Landmark organs, such as the kidney, spleen, and liver, were identified in order to define the area of the pancreas. The transducer was set at the scanning midpoint of the normal pancreas or pancreatic tumor and a 3D image of 10-20 mm, depending on tumor size, at a Z- slice thickness of 0.04 mm. 3D images were uploaded to the Vevo Lab Software. The volumetric analysis function was used to define the tumor border at various Z-slices through the entirety of the tumor and derive the final calculated tumor volume.

Preclinical trials

Age- and sex-matched recipient *C57BL/6J* mice were purchased from The Jackson Laboratory (JAX) or bred in house. Orthotopic transplantation was performed as described above. Mice were monitored for tumor development at 4, 5, 6 weeks post-initiation using high-resolution ultrasound (as described above) to confirm tumor establishment and interval growth. Animals with established tumors (baseline 10-220 mm^3 by 6 weeks post-initiation; median 68 mm^3) were randomized by tumor burden within 24 hours of baseline imaging to either control or experimental treatment arms. Researchers performing health checks, ultrasound imaging and interpretation were blinded to cohort allocation. Isotype (control) arm consisted of 200 µg/mouse Rat IgG2a (BioXCell) + 100 µg/mouse Mouse IgG1 (BioXCell). Experimental arms consisted of anti-PD-1 [95] (BioXCell; Clone 29F.1A12; Rat IgG2a; 200 µg/mouse, dosed i.p. every 2-3 days), anti-TIGIT [71] (Absolute Antibody; Clone 1B4; Mouse IgG1; 100 µg/mouse, dosed i.p. every 2-3 days), CD40 agonist [96] (BioXCell; Clone FGK4.5/FGK45; Rat IgG2a; 100 µg/mouse, dosed i.p. once every 4 weeks) monotherapy or combination therapy as described in the text. Animals were treated for 4 weeks and weekly weights and ultrasound imaging was performed as described. Tumor response was assessed on all evaluable animals at time points (t) >10 days using modified RECIST (mRECIST) criteria, previously adapted for volumetric imaging and preclinical testing [64]. Briefly, tumor volume (V) at each time point (t) was compared to the baseline tumor volume ($V_{baseline}$) in a given animal: (percent change) $\Delta Vol_t = ((V_t - V_{baseline})/V_{baseline})*100\%$. For each animal, we calculated both a "BestResponse" (defined as the minimum ΔVol_t for t>10 days) and

"BestAvgResponse" (defined as the minimum average for $t > 10$ days, where for each time point (t), the average ΔVol_t reflects that time point's ΔVol_t and all prior ΔVol_t). Modified RECIST (mRECIST) criteria were defined as: mCR = BestResponse <−95% and BestAvgResponse < −40%; mPR = BestResponse < −50% and BestAvgResponse < −20%; mSD = BestResponse < 35% and BestAvgResponse < 30%; mPD = not otherwise specified. We also revised the mPD category to be more stringent and consider animals as mPD if the final tumor measurement showed ≥ 20% growth from baseline. See [64] for full details on mRECIST derivation and validation.

Retrograde pancreatic duct delivery

Retrograde pancreatic duct instillation of lentivirus has been previously described [68]. We adapted this technique in a number of ways. Briefly, the ventral abdomen was depilated (using clippers or Nair) 1-2 days prior to surgery. Animals were anesthetized with Isoflurane and the surgical area was disinfected with alternating Betadine/Isopropyl alcohol. A small skin incision was made in the anterior abdomen (~2-3 cm midline incision extending caudally from the xiphoid process). A subsequent incision was made through the linea alba and incision edges were secured in place with a Colibri retractor. The remainder of the procedure was conducted under a Nikon stereomicroscope. A moistened (with sterile 0.9% saline) sterile cotton swab was used to gently move the left lobe of the liver cranially towards the diaphragm. A second moistened sterile cotton swab was used to gently reposition the colon/small intestine into the right lower abdominal quadrant, until the duodenum was visualized. The duodenum was gently repositioned (still in the abdominal cavity) using moistened cotton swabs until the pancreas, common bile duct and sphincter of Oddi were well visualized. The common bile duct and cystic duct were gently separated from the portal vein and hepatic artery using blunt dissection with Moria forceps. A microclip was placed over the common bile duct (cranial to pancreatic duct branching) to prevent influx of the viral particles into the liver or gallbladder, forcing the viral vector retrograde through the pancreatic duct. To infuse the viral vector, the common bile duct was cannulated with a 30-gauge needle at the level of the sphincter of Oddi and 150 μL of virus was injected over the course of 30 seconds. Gentle pressure was applied at the sphincter of Oddi upon needle exit to prevent leakage into the abdominal cavity. Subsequently, the microclip and Colibri retractor were removed. The peritoneum was closed using running 5-0 Vicryl sutures. The cutis and fascia were closed using simple interrupted 5-0 Vicryl sutures. The entire procedure was conducted on a circulating warm water heating blanket to prevent intra-operative hypothermia. All mice received pre-operative analgesia with sustained-release Buprenorphine (Bup-SR) and were followed post-operatively for any signs of discomfort or distress. For retrograde pancreatic ductal installation, male mice (aged 3-6 weeks) and female mice (aged 3-8 weeks) were transduced with 250,000 TU (transducing units, see viral titering) in serum-free media (Opti-MEM; Gibco).

Consistent with prior reports using retrograde pancreatic duct delivery of Cre-containing lentivirus [68], we observed that 17-24% of animals developed small soft tissue sarcomas (most frequently near the abdominal wall incision site) in addition to development of PanIN/PDAC in the pancreas, but these were easily discernable from pancreatic tumors.

For experiments involving CD8 depletion, animals were dosed with CD8a depleting antibody (BioXCell, Clone 2.43, 200 μg/mouse, dosed intraperitoneally [i.p.] every 3-4 days) beginning one day prior to surgery. For TIGIT agonist experiments, animals were dosed with TIGIT agonistic mAb (BioXCell, Clone 1G9, Mouse IgG1, 100 μg/mouse, dosed intraperitoneally [i.p.] every 2-3 days) or Mouse IgG1 isotype control Ab (BioXCell, 100 μg/mouse, dosed intraperitoneally [i.p.] every 2-3 days) beginning one day after surgery.

METHOD DETAILS

Molecular cloning

H11-mScarletSIIN and R26-dCas9-VPR-mNG targeting vectors were generated using gBlocks (IDT) and Gibson assembly [97,98]. U6-sgfiller-eCas9-T2A-BlastR was generated using Gibson assembly. In order to insert sgRNAs, the vector was digested with FastDigest Esp3I (Thermo Fisher) and ligated with BsmBI-compatible annealed oligonucleotides. sgRNAs targeting *Hipp11* or *Rosa26* were designed using Benchling (www.benchling.com), which was also used to predict potential off-target sites.

Lentiviral vectors (LV-PGK-Cre, LV-PGK-Cre-EFS-mScarletSIIN, LV-PGK-PVR-P2A-Cre-EFS-mScarletSIIN, LV-U6-sgRNAfiller-PGK-Cre-EFS-mScarletSIIN) were generated using Gibson assembly. In order to insert sgRNAs into LV-U6-sgRNAfiller-PGK-Cre-EFS-mScarletSIIN, the vector was digested with FastDigest Esp3I (Thermo Fisher) and ligated with BsmBI-compatible annealed oligonucleotides. CRISPRa-compatible sgRNAs targeting *Pvr* were adapted from [99]. See **Table S5** for sgRNA and oligonucleotide sequences. All vectors with detailed maps and sequences will be deposited into Addgene.

Lentiviral production/titering

Lentiviral plasmids and packaging vectors were prepared using endotoxin-free maxiprep kits (QIAGEN). Lentiviruses were produced by co-transfection of HEK293 cells with lentiviral constructs plus packaging vectors: PsPax2 (psPAX2 was a gift from Didier Trono - Addgene plasmid # 12260 ; http://n2t.net/addgene:12260 ; RRID:Addgene_12260) and Pmd2.G (pMD2.G was a gift from Didier Trono - Addgene plasmid # 12259 ; http://n2t.net/addgene:12259 ; RRID:Addgene_12259). Viral supernatant was harvested 48 and 72 hours post transfection, filtered through a 0.45 µm low-protein binding PVDF filter (EMD Millipore), and concentrated by ultracentrifugation (25,000 rpm for 2 hours at 4°C). Concentrated virus was resuspended in Opti-MEM (Gibco) and lentiviral aliquots were frozen and stored at -80°C. Lentiviral titers were determined using Green-Go cells as previously described [100].

Flow cytometry

Flow cytometry of pancreatic organoids

Pancreatic organoids were grown as described above. Where indicated, organoids were treated with interferon-gamma (10 ng/mL; PeproTech) for 48-72 hours prior to analysis. Organoids were dissociated using TrypLE (15 minutes to minimize cleavage of surface proteins) washed with PBS, and filtered through 70 µm filters. Single cell suspensions were pelleted at 2000 rpm and transferred to 96-well round-bottom plates for flow cytometric staining. Prior to surface staining, cell pellets were resuspended in Live/Dead dye (Ghost Dye Red 780, Tonbo Biosciences) diluted 1:1000 in PBS on ice for 20 minutes in the dark. Surface staining was performed on cells in PBS with 1% heat-inactivated FBS on ice for 30 min in the dark. Antibody information in **Table S4**.

Flow cytometry of murine PDAC

Tumors/pancreata were collected in RPMI 1640 supplemented with 1% heat-inactivated FBS. Tumors were finely minced with scissors in MACS C tubes (Miltenyi Biotec), and digested for 30 minutes at 37°C with gentle agitation in 5 mL digestion buffer [1x HBSS (Gibco), 1 mM HEPES (Gibco), 1% heat-inactivated FBS, 125 U/mL collagenase IV (Worthington), 40 U/mL DNase I, grade II (Roche)]. Pancreas tumors were processed on a gentleMACS Octo Dissociator using the “m_spleen_04” program. Digestion buffer was neutralized with 5 mL heat-inactivated FBS, washed with PBS, and filtered through 70 µm filters. Single cell suspensions were pelleted at 1500 rpm with slow deceleration, and transferred to 96-well round-bottom plates for flow cytometric staining. Spleen samples were mashed through 70 μm filters, collected in RPMI 1640 supplemented with 1% heat-inactivated FBS and pelleted. Red blood cells were lysed with ACK buffer for 2 min before cell suspension neutralization with PBS, pelleted for plating and transferred to 96-well round-bottom plates for flow cytometric staining. Prior to surface staining, cell pellets were resuspended in Live/Dead dye (Ghost Dye Red 780, Tonbo Biosciences or Zombie Aqua Fixable Viability Dye, BioLegend) diluted 1:1000 in PBS on ice for 20 minutes in the dark. Surface staining was performed on cells in PBS with 1% heat-inactivated FBS on ice for 30 min in the dark. Cell pellets were fixed overnight in 1X fixation buffer (eBioscience), prior to permeabilization and intracellular staining for 1 hour in the dark at room temperature. Full antibody information in **Table S4**.

Flow cytometry of human PDAC

All human studies were performed using de-identified human biospecimens and studies were approved by the Massachusetts General Brigham Institutional Review Board and conducted according to the principles expressed in the Declaration of Helsinki. The study was in strict compliance with all institutional ethical regulations. All tumor samples were surgically resected primary pancreatic ductal adenocarcinomas and were de-identified prior to researcher processing. Briefly, freshly resected human PDAC specimens were transferred in RPMI 1640 on ice to the laboratory. Pancreas tumors were finely minced with scissors in MACS C tubes, and processed as described above for murine PDAC. Healthy peripheral blood (human PBMCs) from IRB-consented healthy individuals was purchased from StemCell. Antibody information in **Table S4**. As all biospecimens were de-identified, information about age and sex is unavailable. PD-1 staining was omitted during processing of one PDAC biospecimen, so this sample was not included into co-expression analyses with PD-1, but was included in other analyses.

For all flow cytometry experiments, samples were acquired on BD LSR II or LSR Fortessa machines, cell sorting was performed on a BD Aria IIIu. UltraComp eBeads (eBioscience) or single-fluorophore expressing organoids were used for compensation. For murine *in vivo* experiments, endogenous CD44^{lo}CD8^{+} T cells and healthy spleens were used for negative controls and gating. For human experiments, healthy peripheral blood was used as negative controls and gating. For *in vitro* experiments, unstained controls and fluorescence minus one were used for negative controls and gating. Specimens with fewer than 100 live CD8s (mouse) or 200 live CD8s (human) were not considered for further immunophenotyping. FACS data was analyzed using Flowjo v10 software (BD Biosciences).

Immunohistochemistry and pathology review

Tissues were preserved in zinc formalin fixative for 16-24 hours within 1 hour of necropsy, transferred to 70% EtOH, and processed for paraffin embedding. For immunohistochemical staining, slides were blocked using Endogenous Peroxidase Block (Dako) or Bloxall Endogenous Peroxidase and Alkaline Phosphatase Block (Vector Labs) according to manufacturer instructions, followed by incubation with horse serum (Vector Labs) for 1 hour at room temperature. Slides were incubated with primary antibody overnight at 4°C. Details on epitope retrieval and primary antibodies can be found in **Table S4**. The following day, slides were incubated with the appropriate anti-species HRP-conjugated secondary antibody (Vector Labs) for 30 minutes at room temperature. Slides were developed with DAB Peroxidase Substrate Kit (Vector Labs) unless otherwise indicated.

For CD8 and CD4 co-staining, slides were blocked with Bloxall and normal horse serum as above. Slides were incubated with primary rabbit anti-CD8 antibody (Abcam EPR21769, 1:1000) overnight at 4°C and with secondary Alkaline phosphatase anti-Rabbit IgG for 30 minutes at room temperature. Slides were then developed with Vector Black Alkaline phosphatase substrate (Vector Labs) and blocked again with Bloxall and horse serum. Slides were incubated with primary rabbit anti-CD4 (Abcam EPR19514, 1:400) for 3 hours at room temperature and secondary HRP conjugated anti-Rabbit antibody for 30 minutes. Slides were developed with HRP Vina Green Chromogen (Biocare Medical). All murine histologic diagnoses were confirmed with a pathologist (R.T.B.) specialized in rodent pathology.

For CD155 IHC in human tissues, a pancreatic adenocarcinoma tissue microarray (PA1002b) was purchased from Biomax. Anti-CD155 (EPR22672-151] (ab267788; Abcam) was used at 1:500 dilution (final 1.01 μg/mL) following HIER with Tris-EDTA pH 9 (ab93684; Abcam). Slides were reviewed and scored by a board-certified pathologist (G.E.) with membranous staining on tumor cells scored based on intensity of staining as 0, 1^+, 2^+, 3^+. Cores that were missing from TMA (n=1), lacked tumor epithelium (n=2), or that were found on pathologic review to likely represent pancreatic neuroendocrine tumor (n=1) were excluded from further analysis. H-scores were obtained by the formula: (3*percentage of strongly staining cells [3^+]) + (2* percentage of moderately staining cells [2^+]) + (1*percentage of weakly staining cells [1^+]) as previously described [54]. Histopathologic and immunohistochemical analyses were performed using QuPath [101].

Nanostring GeoMx Digital Spatial Profiling

FFPE tissues were sectioned and processed according to Nanostring GeoMx DSP guidelines. Briefly, 5 μm sections were placed on SuperFrost Plus slides. Details on epitope retrieval and primary antibodies can be found in **Table S4**. Tissue morphology markers in the mouse solid tumor morphology kit included PanCK and CD45, and a custom CD8a stain was additionally included. Protein probe sets included the Immune Cell Profiling Core, IO Drug Target Module, Immune Activation Status Module, and Immune Cell Typing Module. Geometric areas of interest (AOIs) were annotated for therapy response and spatial localization as tumor center, tumor periphery, CD8_high, CD8_low, responder, non-responder. The GeoMx platform was similarly used for mfIHC with the antibodies detailed above.

Data QC, normalization, and feature-based selection

Raw expression data were checked for quality and ERCC-normalized prior to statistical analysis. First, expression data were checked for hybridization quality by calculating hybridization factors. A hybridization factor for a given sample was defined as the mean of all HYB-POS values in the dataset divided by that sample's HYB-POS value. Any samples with a hybridization factor of 10 or more were discarded.

Three isotype control molecules were measured for each sample: Rb IgG, Rt IgG2a, and Rt IgG2b. Rt IgG2b, was removed as it showed a reduced correlation with other IgG controls and a greater root mean squared error with the other two (RMSE = 0.29 and 0.30, respectively). Samples were normalized by calculating a normalization factor based on the geometric mean of each sample's Rb IgG and Rt IgG2a expression values. The mean of these geometric means was divided by a given sample's geometric mean value to generate that sample's specific normalization factor. Normalization was then performed by multiplying all proteins for a given sample by its respective normalization factor.

Forty proteins—including S6, Histone H3, GAPDH housekeeping proteins—were measured. These proteins were filtered based on signal to noise ratio (SNR), as calculated by the ERCC-normalized expression for that feature divided by that sample's geometric mean of Rb IgG and Rt IgG2a. A feature was retained if the median SNR value was greater than one. One protein, CD163, had a median SNR of 0.29, but was kept for downstream analysis given prior reports of M2 macrophage polarization in immune evasion.

Statistical Analyses

Hierarchical clustering was performed on the z-scores of the log2 transformed normalized data using the R package pheatmap [102,103]. Differential expression analysis was performed between CD8- compartments (2 mice; 11 ROIs) and CD8+ compartments (3 mice; 25 ROIs). To account for multiple samples taken within a given mouse, a mixed effect model implemented from the R package lmerTest [104]. For a given protein, its log2 transformed expression was used as the dependent variable, CD8 status (CD8-, CD8+) was used as a fixed effect and mouse ID was used as the random effect (with random intercept). Satterthwaite's approximation [104] was used to estimate the degrees of freedom for p-value calculation. Any protein with a singular fitted model were discarded. To account for multiple hypothesis testing, the Benjamini-Hochberg FDR was used [105].

Single-cell RNA sequencing

Sorted cells were washed three times in 1x PBS (calcium and magnesium free) containing 0.04% w/v BSA, and then quantified and titrated to a final concentration of approximately 300 cells/µL. Using the Chromium Single Cell 3' Solution (v3) according to manufacturer's instructions, approximately 2000-5000 cells were partitioned into Gel Beads in Emulsion (GEMs) with cell lysis and barcoded reverse transcription of mRNA into cDNA, followed by amplification, enzymatic fragmentation and 5' adaptor and sample index attachment. The recovery rate was ~800 cells per sample after filtering for quality control. Sample libraries were sequenced on the HiSeq

X Version 2.5 (Illumina) with the following read configuration: Read1 28 cycles, Read2 96 cycles, Index read 8 cycles.

Single-cell RNA sequencing analysis

Data processing, cell clustering, and differential expression analysis

Raw sequencing data was processed using Cell Ranger, version 3.0.2, and sequencing reads were aligned to the mm10 reference mouse transcriptome (version 3.0.0). After processing, Cell Ranger reported 789 cell-associated barcodes and detected 31,053 genes. These data were loaded into R, version 4.0.3, and further processed with Seurat, version 3.2.2 [106]. Genes not expressed in any cells were filtered out. After this, low-quality cells containing more than 10% of reads matching the mitochondrial genome were excluded. Cells with less than 100 detected genes were then filtered. Finally, cells lacking expression of either Cd8a or Cd3e were removed, and cells exceeding the 97th percentile (4,065) for number of detected genes were excluded to remove probable doublets. The resulting matrix used for downstream analyses was defined by 447 cells and 15,065 genes. Data normalization and scaling, variable feature selection, cell clustering, and differential gene expression analysis was performed using Seurat. Data were normalized by total expression per cell and scaled using a factor of 10,000 and log transformed (natural scale). The top 2,000 variable genes were selected using Seurat's default "vst" method. The expression of these genes was then scaled and centered, and these genes were then used for all downstream analysis. Principal component analysis (PCA) was then performed for dimensionality reduction, and the first 30 principal components were selected with the elbow method as a heuristic.

A k-nearest neighbor graph (KNN, k=20) was constructed in PCA space using the top 30 principal components. Four clusters were detected using the Louvain method of community detection (default parameters and resolution = 0.69) [107]. Data was visualized using the Uniform

Manifold Approximation and Projection (UMAP) algorithm implemented in Seurat [108,109]. Default parameters were used, with the following exceptions: the method parameter (“umap-learn”) and the metric parameter (“correlation”). Differential gene expression (min logfc = 0.4; min pct = 20) between clusters was assessed using the default Wilcoxon Rank Sum test.

Gene Module Analysis

Seurat’s AddModuleScore function (control parameter = 8) was used to calculate gene module scores for all cells. For this analysis, gene sets were derived from previously published gene modules (**Table S2**). For datasets providing human gene modules, a custom R script was generated to retrieve corresponding mouse orthologs from Ensembl with the biomaRt package (version 2.42.0) [110,111].

To derive *de novo* gene modules from our scRNA-seq dataset, the Pathway and Gene Set Overdispersion Analysis (PAGODA) [50] framework from the SCDE package (version 2.14.0) was used. The analysis was performed starting with the raw counts for the same 447 cells that remained after filtering in the previous analysis. The knn error model was fit using min.count.threshold = 2 and k = ncol(cd/4), where “cd” represented the matrix after clean.counts was performed with default parameters. Gene expression magnitudes were then normalized with trim = 3/ncol(cd) and max.adj.var=5. *De novo* gene modules were then determined using trim = 7.1/ncol(varinfo$mat) and n.clusters = 50 and otherwise default parameters for the pagoda.gene.clusters function. The top three *de novo* gene sets (modules 30, 36, and 45) with the highest over-dispersion Z score (adjusted for multiple hypotheses) that best distinguished the cellular subpopulations defined by SCDE were selected, and all cells were scored for these modules in Seurat as described above.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical Analyses

All graphs and statistical analyses were generated with GraphPad Prism 8 or in the R statistical programming language (R-project.org) as described above. The following statistical tests were used in this study: (1) Two-sided Mann-Whitney test, (2) t-test with Welch's correction, (3) Kolmogorov-Smirnov (KS). Figure legends specify the statistical tests used, exact value of n, definition of center, and dispersion and precision measures. Figure legends also specify how significance was defined.

Clinical Data Analysis

RNA-seq gene expression profiles (normalized counts) from primary tumors of lung adenocarcinoma (LUAD) [112], pancreatic ductal adenocarcinoma (PAAD) [26], and colorectal adenocarcinoma (COAD) [113] patients were obtained from The Cancer Genome Atlas (TCGA, gdac.broadinstitute.org). Patients within each cohort were limited to those included in the TCGA Pan-Cancer Atlas study [114] for which mutational profiles were available on cBioPortal (cBioPortal for Cancer Genomics, cbioportal.org) [115]. Patients in the PAAD cohort were further limited to those included in the TCGA PAAD study [26]. Within each cancer type, patients were grouped according to *KRAS* and *TP53* mutational status, as retrieved from cBioPortal (KP = alterations in *KRAS* and in *TP53*; nonKP = the remainder of the cohort). Standardized expression levels of *PVR* were illustrated across KP and nonKP patient groups using Empirical Cumulative Distribution Function (ECDF) plots where significance was assessed using a Kolmogorov-Smirnov test.

Neoepitope Prediction

In the TCGA cohort, 148 PDAC patients were analyzed (of 150). One patient lacking a normal BAM file was excluded, and another patient was also excluded due to hypermutation [26]. In the DFCI-PancSeq cohort, 57 patients with (1) annotated mutations, (2) both WES and RNA-Seq data, and (3) sufficient tumor purity [27] were analyzed. Binary Alignment Map (BAM) files were obtained for PancSeq (aligned to GRCh37) and for TCGA (aligned to GRCh38). Thus, GRCh37 was used as the reference genome for the PancSeq cohort in all downstream analyses, and GRCh38 was used for the TCGA cohort.

HLA typing was performed using two programs and with both RNA-Seq and WES data to assess robustness of HLA allele calls. HLA alleles for classical genes (HLA-A, -B, and C) were called using the HLA genotyping algorithm, OptiType, version 1.3.1 [116], as well as seq2HLA, version 2.3 [117], which was also used to identify alleles for HLA-E. Tumor and normal WES BAM files were used to create inputs to OptiType, which outperforms peer programs in WES-based HLA-typing [118], and RNA-Sequencing BAMs were used to create inputs to seq2HLA. WES BAMs were filtered to retain only reads mapping to the HLA region (6:28477897-33448354 in GRCh37; chr6:28510120-33480577 in GRCh38) with the genomics software suite, Samtools, version 1.10 [119]. The BAMs were then converted to FASTQ format, and then filtered with the genome mapping tool, RazerS 3, version 3.5.8 [120], as recommended in the OptiType documentation. RNA-Seq BAMs were sorted, converted to FASTQ format, and compressed before being used as inputs to seq2HLA. Both programs were run with default parameters.

A custom python script was then employed to evaluate concordance between (1) normal and tumor HLA allele calls from WES and (2) seq2HLA and OptiType calls. Only 4/342 alleles (0.58%) in the PancSeq cohort and only 2/888 (0.23%) alleles in the TCGA cohort were called differently between tumor and normal WES-based calls. Given the consistency of OptiType calls for tumor and normal WES data, the OptiType allele was accepted as the final call to resolve discrepancies between OptiType and Seq2HLA.

Mutation Annotation Format (MAF) files were obtained for patients in both datasets and converted to Variant Call Format (VCF) files. VCF files were filtered to only retain single nucleotide variants (SNVs). Only PASS variants were available in the PancSeq MAF file and were thus not filtered further [27]. Mutations in the TCGA cohort included non-PASS variants, which were all filtered in this cohort with the exception of some non-PASS mutations in known PDAC-associated genes that had been marked as either panel_of_normals, clustered_events, or homologous_mapping_event in the TCGA MAF file. For these cases, genes that had variants marked as non-PASS more than twice by at least one of these filters were reconsidered. The following genes with a known association with PDAC based on a literature search were retained: *KRAS*, *TP53*, *GNAS*, *RNF43*, *PLEC*, *FLG*, *AHNAK*, *APOB*, *CSMD1*, *PLXNA1*, *MCM6, MKI67*, and *SIPA1*. This step was intended to reduce false negatives, and in the case of *KRAS*, this step retrieved 30 variants at residue position 12, a site known to confer oncogenic properties when mutated.

Indel variants were called using the variant callers, Strelka2, version 2.9.2 [121], and Scalpel, version 0.5.4 [122]. The structural variant and indel caller, Manta, version 1.6.0 [123], was run prior to Strelka2 and these results were incorporated into the indelCandidates parameter for Strelka2. Scalpel was run with default parameters, with a bed file derived from the CGHub bitbucket account (https://cghub.ucsc.edu; whole_exome_agilent_1.1_refseq_plus_3_boosters.targetIntervals.bed). For the PancSeq cohort, the unmodified first 3 columns of this file were used. For the TCGA cohort, the coordinates in this file were converted to GRCh38 coordinates using the LiftOver tool from the UCSC genome browser [124]. Scalpel failed to call variants for 13/148 TCGA patients due to excessive read buildup at some loci. To enable variant calling with Scalpel for these patients, the Picard tools' DownsampleSam function (http://broadinstitute.github.io/picard/) was employed to randomly downsample reads in the tumor BAM files of these patients by decrements of 10%, starting at 50%. This was done until Scalpel successfully called variants for each patient. Ten patients succeeded at 50%, two at 40%, and one at 30%.

To reduce the contribution of caller-specific biases and hence the indel false positive rate, only those indels that were called and marked as PASS by both Scalpel and Strelka2 were retained. Variant call format (VCF) files containing the union of PASSed variants from Strelka2 and Scalpel were generated with a custom batch script, and variant allele frequencies were calculated using statistics output by Strelka2. These indel VCF files were then merged with the corresponding SNV VCF files for each patient using the vcf-shuffle-cols and vcf-concat functions from VCFtools, version 0.1.13 [125].

Variant consequence was then annotated using the Ensembl Variant Effect Predictor (VEP), version 99 [126]. The corresponding VEP cache for both GRCh37 and GRCh38 was downloaded and used to run the software offline. VEP was run using the Wildtype and Downstream plugins to annotate the effects of indels. The following parameters were employed: --symbol, --terms=SO, --cache, --offline, --transcript_version, --pick. The --pick parameter was reordered from the default to report the transcript with the most extreme consequence for each variant: rank, canonical, appris, tsl, biotype, ccds, length, mane.

Neoepitopes were predicted with the HLA allele calls and variant effect predictions using the antigen prediction toolkit, pVACtools, version 1.5.7 [127]. For each mutation, mutant peptides were generated for lengths of 8-, 9-, 10-, and 11- amino acids, the spectrum of peptide lengths known to bind to MHC class I. MHC:peptide binding affinity was predicted for all peptide:MHC allele pairs with NetMHC-4.0, NetMHCpan-4.0, SMM (version 1.0), and SMMPMBEC (version

1.0) [119,128–130], and the median value across all affinity predictions was taken as a final, composite measure of binding affinity.

After predictions were made by pVACtools, candidate neoepitopes from all patients were merged into a single matrix and filtered using a custom python script and the following criteria (based on parameters output by pVACtools): median peptide:MHC binding affinity < 500 nM, tumor DNA depth >= 5, tumor DNA variant allele frequency >= 0.07, cysteine_count <=1, and a median wildtype:mutant peptide binding affinity fold-change >= 1. After filtering, the total number of remaining candidates was summed per patient and predicted neoepitopes were classified in the following binding affinity ranges: 50-500 nM, 10-50 nM, and 0-10 nM. Neoepitopes were also classified as nonbinders-to-binders (WT nM > 1000 and MT < 500 nM) and nonbinders-to-strong binders (WT nM > 1000 nM and MT < 50 nM). All frameshift-derived neoepitopes with a binding affinity < 500 nM and no corresponding wildtype peptide sequence were also classified as nonbinders-to-binders.

scRNA-seq analysis of human PDAC

Human PDAC scRNA-Seq data [52] was downloaded from the Genome Sequencing Archive (accession: CRA001160). A count matrix of 41,987 pre-processed cells was prepared from these data and used to create a Seurat object. Genes expressed in less than 10 cells were filtered out. Data normalization, scaling, variable feature selection, and principal components analysis were then carried out as described for the murine scRNA-Seq analysis. The first 15 principal components were used for the construction of the k-nearest neighbor graph and the UMAP plot (metric parameter = Euclidean). Clusters were then assigned using the Louvain method with a resolution of 1.

For the T cell subset analysis, cells were selected out from the larger dataset based on expression of *CD8A* and either *CD3E*, *CD3D*, or *CD3G*. Cells expressing both *CD3* and *CD4* were selected in the same way, using *CD4* expression instead of *CD8A*. Genes detected in less than 5 cells were then excluded from the 2 subset matrices individually. These two subsets were then merged to form a new matrix consisting of 3,409 cells and 18,349 genes. These data were then processed as described for the whole dataset, revealing batch effects that caused cells to separate in UMAP space according to their patient of origin. To construct a batch-corrected UMAP, Seurat's integration workflow was performed [131]. Cells were split into individual matrices according to their patient of origin, and matrices corresponding to patients possessing less than 50 cells were excluded to accommodate a k.filter parameter of 50 for the integration anchor identification step. 3,320 cells remained after this step. Pearson residuals were then utilized for data normalization and scaling, as implemented in the SCTransform function [132]. 3,000 integration features were then selected and incorporated as input in the integration anchors identification step. Principal components analysis, k-nearest neighbor graph and UMAP construction (PC dimensions = 14), and cluster annotation (resolution = 0.8) were then performed as described before. For feature plotting and differential gene expression analysis, the UMI count matrix of these cells was separately normalized and scaled as described for the whole dataset. Differential gene expression between clusters was then assessed with the Wilcoxon Rank Sum test. Genuine T cell clusters (0, 1, 8, and 9) were then distinguished by differentially higher expression of *CD3* and/or *CD4* and *CD8A* and by a differentially reduced or complete lack of expression of antigen-presenting cell markers that defined other clusters. To plot murine TIL-derived PAGODA modules on human data, the custom R script described for module analysis of murine scRNA-Seq was employed to retrieve human orthologs of each gene comprising each PAGODA module. The human orthologs were then used to compute module scores with Seurat’s AddModuleScore function (control parameter = 8).

Supplementary Material: All supplementary figures, tables, and data our available from our publication in *Cancer Cell*.

References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians* **70**, 7–30 (2020).
2. Goldstein, D. *et al.* nab-Paclitaxel plus gemcitabine for metastatic pancreatic cancer: long-term survival from a phase III trial. *Journal of the National Cancer Institute* **107**, (2015).
3. Conroy, T. *et al.* FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer. *New England Journal of Medicine* **364**, 1817–1825 (2011).
4. Eso, Y., Shimizu, T., Takeda, H., Takai, A. & Marusawa, H. Microsatellite instability and immune checkpoint inhibitors: toward precision medicine against gastrointestinal and hepatobiliary cancers. *Journal of Gastroenterology* vol. 55 15–26 Preprint at https://doi.org/10.1007/s00535-019-01620-7 (2020).
5. Royal, R. E. *et al.* Phase 2 trial of single agent Ipilimumab (anti-CTLA-4) for locally advanced or metastatic pancreatic adenocarcinoma. *Journal of Immunotherapy (Hagerstown, Md.: 1997)* **33**, 828–833 (2010).
6. Brahmer, J. R. *et al.* Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *The New England Journal of Medicine* **366**, 2455–2465 (2012).
7. O'Reilly, E. M. *et al.* Durvalumab with or Without Tremelimumab for Patients with Metastatic Pancreatic Ductal Adenocarcinoma: A Phase 2 Randomized Clinical Trial. *JAMA Oncology* **5**, 1431–1438 (2019).
8. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
9. Balachandran, V. P. *et al.* Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* **551**, 512–516 (2017).
10. Bailey, P. *et al.* Exploiting the neoantigen landscape for immunotherapy of pancreatic ductal adenocarcinoma. *Scientific Reports* **6**, 1–8 (2016).
11. Tran, E. *et al.* Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science* **350**, 1387–1390 (2015).
12. Parkhurst, M. R. *et al.* Unique neoantigens arise from somatic mutations in patients with gastrointestinal cancers. *Cancer Discovery* **9**, 1022–1035 (2019).
13. Gros, A. *et al.* Recognition of human gastrointestinal cancer neoantigens by circulating PD-1+ lymphocytes. *Journal of Clinical Investigation* **129**, 4992–5004 (2019).
14. Sakellariou-Thompson, D. *et al.* 4-1BB agonist focuses CD8+ tumor-infiltrating T-Cell growth into a distinct repertoire capable of tumor recognition in pancreatic cancer. *Clinical Cancer Research* **23**, 7263–7275 (2017).
15. Stromnes, I. M., Hulbert, A., Pierce, R. H., Greenberg, P. D. & Hingorani, S. R. T-cell Localization, Activation, and Clonal Expansion in Human Pancreatic Ductal Adenocarcinoma. *Cancer Immunology Research* **5**, 978–991 (2017).
16. Steele, N. G. *et al.* Multimodal mapping of the tumor and peripheral blood immune landscape in human pancreatic cancer. *Nature Cancer* 1–16 (2020) doi:10.1038/s43018-020-00121-4.
17. Johnson, B. A., Yarchoan, M., Lee, V., Laheru, D. A. & Jaffee, E. M. Strategies for increasing pancreatic tumor immunogenicity. *Clinical Cancer Research* vol. 23 1656–1669 Preprint at https://doi.org/10.1158/1078-0432.CCR-16-2318 (2017).
18. Evans, R. A. *et al.* Lack of immunoediting in murine pancreatic cancer reversed with neoantigen. *JCI insight* **1**, (2016).

19. Burrack, A. L. *et al.* Combination PD-1 and PD-L1 Blockade Promotes Durable Neoantigen-Specific T Cell-Mediated Immunity in Pancreatic Ductal Adenocarcinoma. (2019) doi:10.1016/j.celrep.2019.07.059.
20. Hegde, S. *et al.* Dendritic Cell Paucity Leads to Dysfunctional Immune Surveillance in Pancreatic Cancer. *Cancer Cell* **37**, 289-307.e9 (2020).
21. Hwang, C.-I., Boj, S. F., Clevers, H. & Tuveson, D. A. Preclinical models of pancreatic ductal adenocarcinoma. *The Journal of Pathology* **238**, 197–204 (2016).
22. Boj, S. F. *et al.* Organoid models of human and mouse ductal pancreatic cancer. *Cell* **160**, 324–338 (2015).
23. Huch, M. *et al.* Unlimited in vitro expansion of adult bi-potent pancreas progenitors through the Lgr5/R-spondin axis. *The EMBO journal* **32**, 2708–2721 (2013).
24. Li, X. *et al.* Oncogenic transformation of diverse gastrointestinal tissues in primary organoid culture. *Nature Medicine* **20**, 769–777 (2014).
25. Ryan, D. P., Hong, T. S. & Bardeesy, N. Pancreatic adenocarcinoma. *The New England Journal of Medicine* **371**, 1039–1049 (2014).
26. Cancer Genome Atlas Research Network. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **32**, 185-203.e13 (2017).
27. Aguirre, A. J. *et al.* Real-time Genomic Characterization of Advanced Pancreatic Cancer to Enable Precision Medicine. *Cancer Discovery* **8**, 1096–1111 (2018).
28. Hippenmeyer, S. *et al.* Genetic mosaic dissection of Lis1 and Ndel1 in neuronal migration. *Neuron* **68**, 695–709 (2010).
29. Bindels, D. S. *et al.* MScarlet: A bright monomeric red fluorescent protein for cellular imaging. *Nature Methods* **14**, 53–56 (2016).
30. Alspach, E. *et al.* MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* **574**, 696–701 (2019).
31. Gubin, M. M. *et al.* Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577–581 (2014).
32. Joyce, J. A. & Fearon, D. T. T cell exclusion, immune privilege, and the tumor microenvironment. *Science (New York, N.Y.)* **348**, 74–80 (2015).
33. Liudahl, S. M. *et al.* Leukocyte Heterogeneity in Pancreatic Ductal Adenocarcinoma: Phenotypic and Spatial Features Associated with Clinical Outcome. *Cancer Discovery* candisc.0841.2020 (2021) doi:10.1158/2159-8290.CD-20-0841.
34. Yamamoto, K. *et al.* Autophagy promotes immune evasion of pancreatic cancer by degrading MHC-I. *Nature* 1–6 (2020) doi:10.1038/s41586-020-2229-5.
35. Michelakos, T. *et al.* Tumor Microenvironment Immune Response in Pancreatic Ductal Adenocarcinoma Patients Treated With Neoadjuvant Therapy. *JNCI: Journal of the National Cancer Institute* (2020) doi:10.1093/jnci/djaa073.
36. Fumagalli, A. *et al.* Genetic dissection of colorectal cancer progression by orthotopic transplantation of engineered cancer organoids. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E2357–E2364 (2017).
37. Hogquist, K. A. *et al.* T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17–27 (1994).
38. Blank, C. U. *et al.* Defining ‘T cell exhaustion’. *Nature Reviews Immunology* **19**, 665–674 (2019).

39. Winograd, R. *et al.* Induction of T-cell Immunity Overcomes Complete Resistance to PD-1 and CTLA-4 Blockade and Improves Survival in Pancreatic Carcinoma. *Cancer Immunology Research* **3**, 399–411 (2015).
40. Chihara, N. *et al.* Induction and transcriptional regulation of the co-inhibitory gene module in T cells. *Nature* **558**, 454–459 (2018).
41. Schietinger, A. *et al.* Tumor-Specific T Cell Dysfunction Is a Dynamic Antigen-Driven Differentiation Program Initiated Early during Tumorigenesis. *Immunity* **45**, 389–401 (2016).
42. Miller, B. C. *et al.* Subsets of exhausted CD8+ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nature Immunology* **20**, 326–336 (2019).
43. Siddiqui, I. *et al.* Intratumoral Tcf1 + PD-1 + CD8 + T Cells with Stem-like Properties Promote Tumor Control in Response to Vaccination and Checkpoint Blockade Immunotherapy Article Intratumoral Tcf1 + PD-1 + CD8 + T Cells with Stem-like Properties Promote Tumor Control in Response to Vaccination and Checkpoint Blockade Immunotherapy. (2019) doi:10.1016/j.immuni.2018.12.021.
44. Sade-Feldman, M. *et al.* Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. *Cell* **175**, 998-1013.e20 (2018).
45. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **8**, 1–12 (2017).
46. Doering, T. A. *et al.* Network Analysis Reveals Centrally Connected Genes and Pathways Involved in CD8+ T Cell Exhaustion versus Memory. *Immunity* **37**, 1130–1144 (2012).
47. Singer, M. *et al.* A Distinct Gene Module for Dysfunction Uncoupled from Activation in Tumor-Infiltrating T Cells. *Cell* **166**, 1500-1511.e9 (2016).
48. Kim, H. J. *et al.* CD8+ T regulatory cells express the Ly49 class I MHC receptor and are defective in autoimmune prone B6-Yaa mice. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 2010–2015 (2011).
49. Saligrama, N. *et al.* Opposing T cell responses in experimental autoimmune encephalomyelitis. *Nature* **572**, 481–487 (2019).
50. Fan, J. *et al.* Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nature Methods* **13**, 241–244 (2016).
51. Wherry, E. J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nature Reviews Immunology* vol. 15 486–499 Preprint at https://doi.org/10.1038/nri3862 (2015).
52. Peng, J. *et al.* Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Research* **29**, 725–738 (2019).
53. Yarchoan, M. *et al.* PD-L1 expression and tumor mutational burden are independent biomarkers in most cancers. *JCI Insight* **4**, (2019).
54. Hirsch, F. R. *et al.* Epidermal growth factor receptor in non-small-cell lung carcinomas: Correlation between gene copy number and protein expression and impact on prognosis. *Journal of Clinical Oncology* **21**, 3798–3807 (2003).
55. Wells, D. K. *et al.* Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. *Cell* **183**, 818-834.e13 (2020).
56. Vonderheide, R. H. The Immune Revolution: A Case for Priming, Not Checkpoint. *Cancer Cell* vol. 33 563–569 Preprint at https://doi.org/10.1016/j.ccell.2018.03.008 (2018).
57. Ribas, A. *et al.* CD40 cross-linking bypasses the absolute requirement for CD4 T cells during immunization with melanoma antigen gene-modified dendritic cells. *Cancer research* **61**, 8787–93 (2001).

58. Vonderheide, R. H. CD40 Agonist Antibodies in Cancer Immunotherapy. *Annual Review of Medicine* **71**, 47–58 (2020).
59. Byrne, K. T. & Vonderheide, R. H. CD40 Stimulation Obviates Innate Sensors and Drives T Cell Immunity in Cancer. *Cell Reports* **15**, 2719–2732 (2016).
60. Ma, H. S. *et al.* A CD40 Agonist and PD-1 Antagonist Antibody Reprogram the Microenvironment of Nonimmunogenic Tumors to Allow T-cell–Mediated Anticancer Activity. *Cancer Immunology Research* **7**, 428–442 (2019).
61. Hung, A. L. *et al.* TIGIT and PD-1 dual checkpoint blockade enhances antitumor immunity and survival in GBM. *OncoImmunology* **7**, e1466769 (2018).
62. Anderson, A. C., Joller, N. & Kuchroo, V. K. Lag-3, Tim-3, and TIGIT: Co-inhibitory Receptors with Specialized Functions in Immune Regulation. *Immunity* vol. 44 989–1004 Preprint at https://doi.org/10.1016/j.immuni.2016.05.001 (2016).
63. Rodriguez-Abreu, D. *et al.* Primary analysis of a randomized, double-blind, phase II study of the anti-TIGIT antibody tiragolumab (tira) plus atezolizumab (atezo) versus placebo plus atezo as first-line (1L) treatment in patients with PD-L1-selected NSCLC (CITYSCAPE). *J Clin Oncol* **38**, (suppl; abstr 9503) (2020).
64. Gao, H. *et al.* High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nature Medicine* **21**, 1318–1325 (2015).
65. O'Hara, M. H. *et al.* CD40 agonistic monoclonal antibody APX005M (sotigalimab) and chemotherapy, with or without nivolumab, for the treatment of metastatic pancreatic adenocarcinoma: an open-label, multicentre, phase 1b study. *The Lancet Oncology* **22**, 118–131 (2021).
66. Schnell, A., Bod, L., Madi, A. & Kuchroo, V. K. The yin and yang of co-inhibitory receptors: toward anti-tumor immunity without autoimmunity. *Cell Research* vol. 30 285–299 Preprint at https://doi.org/10.1038/s41422-020-0277-x (2020).
67. Harjunpää, H. & Guillerey, C. TIGIT as an emerging immune checkpoint. *Clinical and Experimental Immunology* vol. 200 108–119 Preprint at https://doi.org/10.1111/cei.13407 (2020).
68. Chiou, S.-H. *et al.* Pancreatic cancer modeling using retrograde viral vector delivery and in vivo CRISPR/Cas9-mediated somatic genome editing. *Genes & Development* **29**, 1576–1585 (2015).
69. Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain [1]. *Nature Genetics* vol. 21 70–71 Preprint at https://doi.org/10.1038/5007 (1999).
70. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nature Methods* **12**, 326–328 (2015).
71. Dixon, K. O. *et al.* Functional Anti-TIGIT Antibodies Regulate Development of Autoimmunity and Antitumor Immunity. *Journal of Immunology (Baltimore, Md.: 1950)* **200**, 3000–3007 (2018).
72. Snyder, A. *et al.* Genetic basis for clinical response to CTLA-4 blockade in melanoma. *New England Journal of Medicine* **371**, 2189–2199 (2014).
73. Dunn, G. P., Old, L. J. & Schreiber, R. D. The Three Es of Cancer Immunoediting. *Annual Review of Immunology* **22**, 329–360 (2004).
74. Ikeda, W. *et al.* Tage4/nectin-like molecule-5 heterophilically trans-interacts with cell adhesion molecule nectin-3 and enhances cell migration. *Journal of Biological Chemistry* **278**, 28167–28172 (2003).

75. Nishi, K. *et al.* Mutant KRAS promotes NKG2D+ T cell infiltration and CD155 dependent immune evasion. *Anticancer Research* **40**, 4663–4674 (2020).
76. Yu, X. *et al.* The surface protein TIGIT suppresses T cell activation by promoting the generation of mature immunoregulatory dendritic cells. *Nature Immunology* **10**, 48–57 (2009).
77. Kurtulus, S. *et al.* TIGIT predominantly regulates the immune response via regulatory T cells. *Journal of Clinical Investigation* **125**, 4053–4062 (2015).
78. Chen, D. S. & Mellman, I. Oncology meets immunology: The cancer-immunity cycle. *Immunity* vol. 39 1–10 Preprint at https://doi.org/10.1016/j.immuni.2013.07.012 (2013).
79. Li, J. *et al.* Tumor Cell-Intrinsic Factors Underlie Heterogeneity of Immune Cell Infiltration and Response to Immunotherapy. *Immunity* **49**, 178-193.e7 (2018).
80. Lin, J. H. *et al.* Type 1 conventional dendritic cells are systemically dysregulated early in pancreatic carcinogenesis. *Journal of Experimental Medicine* **217**, (2020).
81. Bayne, L. J. *et al.* Tumor-Derived Granulocyte-Macrophage Colony-Stimulating Factor Regulates Myeloid Inflammation and T Cell Immunity in Pancreatic Cancer. *Cancer Cell* **21**, 822–835 (2012).
82. Stromnes, I. M. *et al.* Targeted depletion of an MDSC subset unmasks pancreatic ductal adenocarcinoma to adaptive immunity. *Gut* **63**, 1769–1781 (2014).
83. Wellenstein, M. D. & de Visser, K. E. Cancer-Cell-Intrinsic Mechanisms Shaping the Tumor Immune Landscape. *Immunity* vol. 48 399–416 Preprint at https://doi.org/10.1016/j.immuni.2018.03.004 (2018).
84. Tsujikawa, T. *et al.* Evaluation of Cyclophosphamide/GVAX Pancreas Followed by Listeria-mesothelin (CRS-207) With or Without Nivolumab in Patients with Pancreatic Cancer. *Clinical Cancer Research* clincanres.3978.2019 (2020) doi:10.1158/1078-0432.CCR-19-3978.
85. Feig, C. *et al.* Targeting CXCL12 from FAP-expressing carcinoma-associated fibroblasts synergizes with anti-PD-L1 immunotherapy in pancreatic cancer. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 20212–20217 (2013).
86. Ruscetti, M. *et al.* Senescence-Induced Vascular Remodeling Creates Therapeutic Vulnerabilities in Pancreas Cancer. *Cell* **181**, 424-441.e21 (2020).
87. Biasci, D. *et al.* CXCR4 inhibition in human pancreatic and colorectal cancers induces an integrated immune response. *Proceedings of the National Academy of Sciences* **117**, 202013644 (2020).
88. Morrison, A. H., Diamond, M. S., Hay, C. A., Byrne, K. T. & Vonderheide, R. H. Sufficiency of CD40 activation and immune checkpoint blockade for T cell priming and tumor immunity. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 8022–8031 (2020).
89. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes and Development* **15**, 3243–3248 (2001).
90. Marino, S., Vooijs, M., Van Der Gulden, H., Jonkers, J. & Berns, A. Induction of medulloblastomas in p53-null mutant mice by somatic inactivation of Rb in the external granular layer cells of the cerebellum. *Genes and Development* **14**, 994–1004 (2000).
91. Church, G. M. & Gilbert, W. Genomic sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 1991–1995 (1984).

92. VanDussen, K. L., Sonnek, N. M. & Stappenbeck, T. S. L-WRN conditioned medium for gastrointestinal epithelial stem cell culture shows replicable batch-to-batch activity levels across multiple research teams. *Stem Cell Research* **37**, 101430 (2019).
93. Kim, M. P. *et al.* Generation of orthotopic and heterotopic human pancreatic cancer xenografts in immunodeficient mice. *Nature Protocols* **4**, 1670–1680 (2009).
94. Sastra, S. A. & Olive, K. P. Quantification of murine pancreatic tumors by high-resolution ultrasound. *Methods in Molecular Biology* **980**, 249–266 (2013).
95. Liang, S. C. *et al.* Regulation of PD-1, PD-L1, and PD-L2 expression during normal and autoimmune responses. *European Journal of Immunology* **33**, 2706–2716 (2003).
96. Rolink, A., Melchers, F. & Andersson, J. The SCID but not the RAG-2 gene product is required for Sμ-Sε heavy chain class switching. *Immunity* **5**, 319–330 (1996).
97. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343–345 (2009).
98. Akama-Garren, E. H. *et al.* A Modular Assembly Platform for Rapid Generation of DNA Constructs. *Scientific Reports* **6**, (2016).
99. Horlbeck, M. A. *et al.* Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* **5**, (2016).
100. Sánchez-Rivera, F. J. *et al.* Rapid modelling of cooperating genetic events in cancer through somatic genome editing. *Nature* **516**, 428–431 (2014).
101. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Scientific Reports* **7**, (2017).
102. Ihaka, R. & Gentleman, R. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* **5**, 299–314 (1996).
103. Kolde, R. pheatmap: Pretty Heatmaps. Preprint at (2019).
104. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, Articles* **82**, 1–26 (2017).
105. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995).
106. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411–420 (2018).
107. De Meo, P., Ferrara, E., Fiumara, G. & Provetti, A. Generalized Louvain method for community detection in large networks. in *International Conference on Intelligent Systems Design and Applications, ISDA* 88–93 (2011). doi:10.1109/ISDA.2011.6121636.
108. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology* **37**, 38–47 (2019).
109. Mcinnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection Software • Review • Repository • Archive. (2018) doi:10.21105/joss.00861.
110. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics (Oxford, England)* **21**, 3439–40 (2005).
111. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nature Protocols* **4**, 1184–1191 (2009).

112. Collisson, E. A. *et al.* Comprehensive molecular profiling of lung adenocarcinoma: The cancer genome atlas research network. *Nature* **511**, 543–550 (2014).
113. Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
114. Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304.e6 (2018).
115. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling* **6**, (2013).
116. Szolek, A. *et al.* OptiType: Precision HLA typing from next-generation sequencing data. *Bioinformatics* **30**, 3310–3316 (2014).
117. Boegel, S. *et al.* HLA typing from RNA-Seq sequence reads. *Genome Medicine* **4**, 102 (2012).
118. Bauer, D. C., Zadoorian, A., Wilson, L. O. W. & Thorne, N. P. Evaluation of computational programs to predict HLA genotypes from genomic sequencing data. *Briefings in Bioinformatics* **19**, 179–187 (2018).
119. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
120. Weese, D., Holtgrewe, M. & Reinert, K. RazerS 3: Faster, fully sensitive read mapping. *Bioinformatics* **28**, 2592–2599 (2012).
121. Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nature Methods* **15**, 591–594 (2018).
122. Narzisi, G. *et al.* Accurate de novo and transmitted indel detection in exome-capture data using microassembly. *Nature Methods* **11**, 1033–1036 (2014).
123. Chen, X. *et al.* Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
124. Haeussler, M. *et al.* The UCSC Genome Browser database: 2019 update. *Nucleic Acids Research* **47**, D853–D858 (2019).
125. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
126. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122 (2016).
127. Hundal, J. *et al.* PVACtools: A computational toolkit to identify and visualize cancer neoantigens. *Cancer Immunology Research* **8**, 409–420 (2020).
128. Jurtz, V. *et al.* NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *The Journal of Immunology* **199**, 3360–3368 (2017).
129. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: Application to the MHC class i system. *Bioinformatics* **32**, 511–517 (2016).
130. Peters, B. & Sette, A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* **6**, (2005).
131. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
132. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology* **20**, 296 (2019).

APPENDIX II: LOW NEOANTIGEN EXPRESSION AND POOR T CELL PRIMING UNDERLIE EARLY IMMUNE ESCAPE IN CANCER

Authors: Peter M K Westcott[1], Nathan Sacks[1], Jason M Schenkel[1,2], Zackery A Ely[1], Olivia Smith[1], Alex Jaeger[1], Haley Hauck[1], Daniel Zhang[1], Coralie M. Backlund[1], Antonius Koller, Mary C Beytagh[1], JJ Patten[1], George Eng[1,3], Darrell J. Irvine[1,4,5,6,7], Omer H Yilmaz[1,3,8], Tyler Jacks[1,7,8*]

Affiliations:

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

[2]Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA

[3]Department of Pathology, Massachusetts General Hospital, Boston, MA 02114, USA

[4]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[5]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02139, USA

[6]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[7]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[8]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[*]Corresponding author

Contributions: P.M.K.W. and T.J. conceived and directed the study. P.M.K.W., N.S., O.S., H.H. and A.J. carried out all aspects of the research, animal care and experimentation. J.M.S. provided essential conceptual and technical guidance in the design and execution of flow cytometry-based experiments. Z.E. designed and executed the pipeline to generate predicted neoantigens from TCGA COADREAD and analyze RNA variant allele frequencies. A.J. designed and carried out MHC-I pull-down and elution for MS. N.S. developed the triple IHC and an automated CNN for quantification in collaboration with Aiforia. N.S., O.S., D.Z., J.J.P., M.C.B. and R.E. generated lentiviral constructs and primary organoid lines used in the study. C.M.B. and D.J.I. provided guidance, reagents and technical assistance with therapeutic vaccinations. G.E. and O.Y. provided important guidance and reagents for organoid culture and colonoscopy-guided injections. All data analysis was carried out by P.M.K.W. The manuscript was written by P.M.K.W. and T.J. with feedback from all authors.

Abstract

Immune evasion is a hallmark of cancer, and therapies that restore immune surveillance have proven highly effective in cancers with high tumor mutation burden (TMB) (e.g. microsatellite instable (MSI) colorectal cancer (CRC)). Whether low TMB cancers, which are largely refractory to immunotherapy, harbor T cell neoantigens capable of engaging adaptive immunity remains unclear. Here, we show that tumors from all patients with microsatellite stable (MSS) CRC harbor predicted high-affinity clonal neoantigens despite low TMB. Unexpectedly, these neoantigens are broadly expressed at lower levels relative to those in MSI CRC, suggesting a potential role of antigen expression in tumor immune surveillance. To test this, we developed a versatile platform for functional interrogation of neoantigens with variable expression and applied it to novel preclinical colonoscopy-guided mouse models of CRC. While high expression of multiple high-affinity Major Histocompatibility Complex I (MHC-I)-restricted neoantigens universally resulted in tumor rejection, low expression resulted in poor T cell cross priming and tumor progression. Strikingly, experimental or therapeutic rescue of priming rendered T cells fully capable of controlling tumors with low neoantigen expression. These findings underscore a critical role of neoantigen expression levels in immune evasion and suggest that poor expression or presentation may be a general feature of neoantigens acquired early in tumorigenesis. Finally, poorly expressed neoantigens, commonly excluded in tumor vaccine pipelines, may hold untapped therapeutic potential.

Introduction

Approximately 12% of CRC has defects in DNA mismatch repair (MMR) resulting in MSI[1], with a high burden of mutation-derived tumor-specific antigens (neoantigens) that underlies favorable response to ICB[2–4]. The remaining majority of CRC is MSS with a lower

TMB. However, MSS CRC has more mutations, on average, than some cancers that respond favorably to ICB[1,5], and cell surface presentation of neoantigen-derived epitopes (neoepitopes) on Human Leukocyte Antigen class I (HLA-I) has been observed in a small study of this tumor subtype[6]. This suggests that other factors, both tumor intrinsic and microenvironmental, likely contribute to the poor immunogenicity of MSS CRC and other immune "cold" cancers. Indeed, The intestinal microenvironment is tolerant to commensal bacteria and food-derived antigens, and these mechanisms may be co-opted by tumors to undermine immune responses[7,8]. In addition, the vast majority of CRC is associated with aberrant WNT/β-catenin signaling[1], which can promote exclusion of dendritic cells and failure to prime productive T cell responses in melanoma and liver cancer[9,10]. To rigorously study the processes underlying T cell dysfunction and immunotherapy resistance, it is critical that preclinical models faithfully recapitulate the tissue microenvironment and genetics of the human disease. In addition, models should enable isolation of defined antigen-specific T cell responses. To our knowledge, no single model of CRC meets all of these criteria. Therefore, we adapted a technique employing endoscope-guided submucosal injection[11] to induce genetically-defined tumors in the mouse colon harboring model $CD8^+$ T cell antigens.

An additional feature we sought to model is the role of neoantigen expression level in modulating the anti-tumor immune response. While mutations in cancer are generally enriched in poorly expressed genes due to reduced transcription-coupled repair[12], this pattern is absent in MMR-deficient tumors[13]. This raises the possibility that poor immunogenicity of MSS CRC and other MMR-proficient cancers is not only due to lower burden but also lower expression of neoantigens. It is appreciated that antigen expression is a central determinant of the magnitude of T cell response in viral infection[14], and low neoantigen expression or affinity for MHC-I results in tumor immune evasion in flank transplant mouse models[15,16]. It is also clinically appreciated that tumors frequently evade immune recognition via loss of heterozygosity of HLA alleles and dysregulation of antigen processing and presentation[17]. Despite these results, there remains some controversy surrounding the role of initial T cell receptor (TCR) signaling strength in shaping T cell fate and function[18]. In cancer specifically, the impact of low neoantigen expression on T cell dysfunction is poorly characterized, and it is unclear if such T cell responses persist in the tumor and lymphoid tissues and are amenable to reinvigoration by ICB. As a proof-of-principle of our novel model system, we undertook deep mechanistic characterization of the role of neoantigen expression in tuning the anti-tumor T cell response, and performed preclinical trials of promising immunotherapeutic combinations and neoantigen vaccination.

Results

Microsatellite stable CRC is defined by both lower burden and expression of neoantigens.

To guide development of a mouse model enabling tracking of tumor-specific T cell responses, we first developed a neoantigen prediction pipeline integrating multiple gold-standard HLA haplotype callers and affinity prediction algorithms (see Methods) and applied it to The Cancer Genome Atlas (TCGA) colorectal adenocarcinoma (COADREAD) sequencing dataset. This revealed that tumors from all MSS CRC patients (excluding rare cases with hypermutation) expressed at least 14 (median of 121) single nucleotide variant (SNV)- or insertion/deletion (indel)-derived neoantigens with strong predicted binding to their respective HLA-I ($IC_{50} \leq 500$ nM), despite a lower TMB compared to MSI-H (MSI high) CRC patients (Fig. 1a-b, Extended Data Fig. 1a). Interestingly, average neoantigen expression was also significantly lower in MSS versus MSI-H tumors (Fig 1b, Extended Data Fig. 1b), consistent with the enrichment of mutations in highly expressed genes observed in MMR-deficient tumors[13].

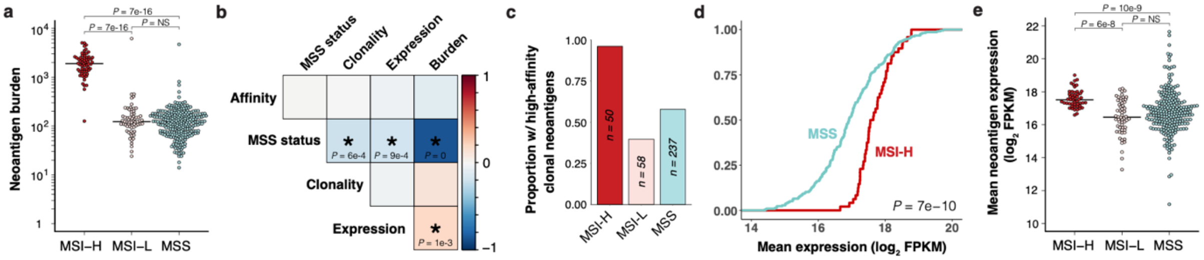


Figure 1. Lower burden and expression of predicted neoantigens in MSS versus MSI-H CRC. Analysis of predicted neoantigens in human CRC (TCGA COADREAD) with high MSI (MSI-H), low MSI (MSI-L), and MSS. **(a)** Total expressed neoantigens with strong predicted HLA-I binding (IC50 ≤ 500 nM) by patient (excluding hypermutant MSS cases). **(b)** Spearman rank correlation matrix of MSS status (MSS versus MSI-H) and mean neoantigen expression, predicted affinity, burden, and clonality (adjVAF) by patient. Strength of correlation is represented by color scale (red = positive, blue = negative), and significance is indicated by asterisk with P- values displayed. **(c)** Proportion of patients expressing at least one clonal (adjVAF ≥ 0.5) neoantigen with very strong predicted binding affinity (IC50 ≤ 10 nM). **(d-e)** Analysis of expressed clonal neoantigens only. **(d)** Empirical cumulative distribution function of mean neoantigen expression by patient, showing enrichment of lower expression in MSS patients. Significance was assessed by two-sided Kolmogorov-Smirnov test. **(e)** Mean expression of clonal neoantigens by patient (FPKM, upper quartile-normalized).

We also observed significantly lower average neoantigen clonality (tumor purity-adjusted variant allele frequency (adjVAF) ≥ 0.5) in MSS tumors (Fig. 1b, Extended Data Fig. 1c), which may contribute to lower expression. Given this and the fact that clonal neoantigens are more likely to elicit productive immune responses following immunotherapy[19,20], we repeated our analysis focusing only on clonal neoantigens. Surprisingly, all tumors from MSS patients still expressed at least two clonal neoantigens (median of 42) (Extended Data Fig. 1d), and 58% of these tumors expressed at least one clonal neoantigen with predicted affinity as high or higher than the commonly used model $CD8^+$ T cell antigen SIINFEKL (H-$2K^b$ IC_{50} ≤ 10 nM)[21] (Fig. 1c). Additionally, average clonal neoantigen expression was still significantly lower in tumors from MSS versus MSI-H patients (Fig. 1d-e). Allele-specific expression of all SNV-derived clonal neoantigens, while limited by sparse coverage, recapitulated these results (Extended Data Fig. 1e).

Finally, a published immunopeptidomics study of human MSS CRC identified three clonal HLA-I neoepitopes in two of five patient-derived organoids (PDOs) analyzed by mass spectrometry (MS)[6]. While this study is small and the number of neoepitopes validated is substantially lower than predicted, this suggests that ~40% of MSS CRC may present *bona fide* neoepitopes. Our analysis of these data showed that these neoepitopes fell in the 6th, 31st, and 45th percentiles of abundance of all detected HLA-I epitopes (Extended Data Fig. 1f-g). While peptide-specific properties can influence the efficiency of separation and ionization in MS, this qualitative analysis suggests lower surface abundance than the majority of self-epitopes. Given that the detection limit of MS in these types of experiments is poorly defined, it is also possible that some *bona fide* neoepitopes were not recovered. Altogether, our integrated analysis of the TCGA and literature argue that many MSS CRC patients may harbor therapeutically actionable neoantigens. However, it remains poorly understood how low expression of neoantigens shapes resulting T cell responses in MSS CRC.

Neoantigen expression level is a critical determinant of immune outcome in a novel orthotopic mouse model of colon cancer. We first developed an autochthonous model in *Apc*$^{flox/flox}$ mice initiated by lentivirus expressing Cre-recombinase and the ovalbumin antigen linked to luciferase (LucOS) (Extended Data Fig. 2a), as our group has previously done in models of lung cancer[22] and soft-tissue sarcoma[23]. Injection with LucOS dramatically reduced tumor incidence in a T cell-mediated manner, and tumors that did arise invariably lost antigen expression (Extended Data Fig. 2b-e). To assess effects of antigen expression in established tumors, T cells were continuously depleted for 5 weeks, at which point tumors retained antigen expression. However, 7 weeks after withdrawal of depleting antibodies, tumors had grown and lost antigen expression (Extended Data Fig. 2f-g). Given this potent immune editing and variability of antigen expression (Extended Data Fig. 2h), we developed an organoid model that maintains distinct levels of antigen expression throughout tumorigenesis. Organoid engineering also enables the use of a *Kras* mutant allele to model metastatic adenocarcinoma, which is confounded in the autochthonous model by concomitant *Kras*-driven fibrosarcoma formation (unpublished observations).

To enforce stable and continuous expression of antigen, we generated CRC organoids with SIINFEKL directly linked to *Apc* knockdown, an essential event in transformation. Specifically, we transformed normal colon organoids from C57Bl/6 *Kras*$^{LSL\text{-}G12D}$; *Trp53*$^{flox/flox}$ (KP) mice with adenoviral Cre, followed by lentivirus expressing miR-30 shRNA against *Apc* (shApc)[24] and SIINFEKL fused to the fluorophore mScarlet (mScarlet$^{\mathrm{SIIN}}$) (Fig. 2a). Given that relief from *Apc* knockdown in shApc-transformed tumors results in regression[24], dependence on shApc provides powerful selection against antigen loss (Extended Data Fig. 2i). Finally, deletion of *Smad4*, commonly mutated in CRC[25], was achieved by CRISPR-Cas9 editing. Selection of organoids harboring complete mutation of all genes was performed following published protocols[26,27]. This resulted in isogenic quadruple-mutant (shAKPS) organoids modeling some of the most common genetic mutations in MSS CRC[25] (Extended Data Fig. 2j), which are co-mutated with high frequency in metastatic disease and associated with poor prognosis[28]. To investigate the importance of neoantigen expression level, we generated organoids with 400-fold range of mScarlet$^{\mathrm{SIIN}}$ fusion protein expression via modifications to the shApc-expressing lentivirus, including placement in reverse orientation to the promoter (EF1α initiates bidirectional transcription), removal of the Kozak consensus sequence, and replacing 25% of codons with rare variants (Fig. 2a-b, Extended Data Fig. 2k). Importantly, this flexible system is broadly applicable to other cancers via linkage to relevant essential events (e.g. knockdown of *Trp53*) and is easily adapted to the study of other immune epitopes.

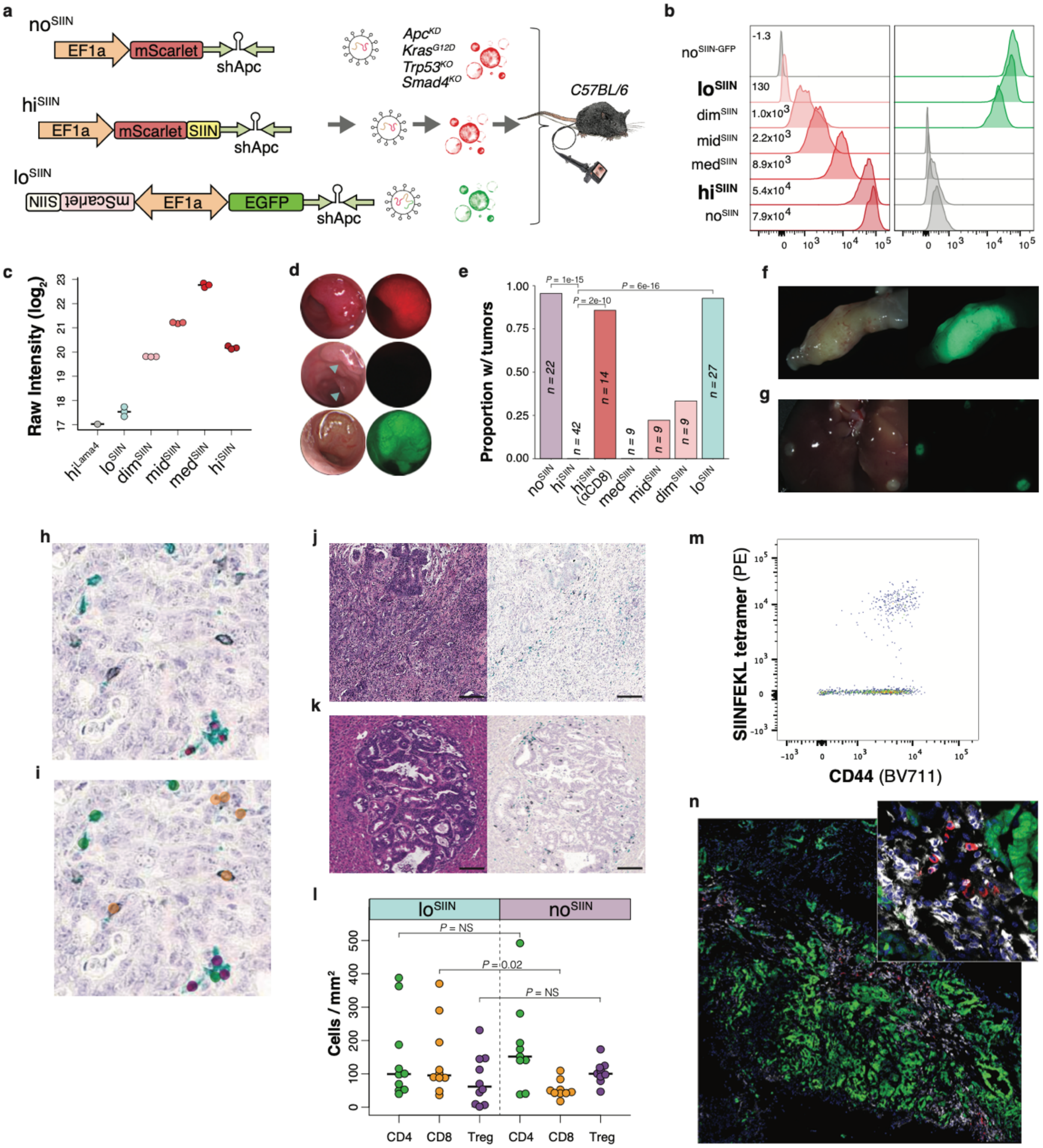


Figure 2 A novel model of CRC reveals a critical role of neoantigen expression level in tumor immune evasion. (a) shApc-expressing lentiviruses used to transform KP organoids, with no (noSIIN), high (hiSIIN) and low (loSIIN) expression of SIINFEKL. Resulting shAKPS organoids were orthotopically-transplanted into the colons of syngeneic mice. **(b)** Expression of

mScarlet/mScarletSIIN and EGFP in expression variant organoids by flow cytometry. **(c)** TMT-MS quantification of MHC-I bound SIINFEKL across expression variants. **(d)** Colonoscopy images of noSIIN tumor (RFP channel), hiSIIN scars (RFP channel, blue arrows indicate injection sites), and loSIIN tumor (GFP channel) 28 days post-transplant. **(e)** Efficiency of tumor formation 42 days post-transplant with noSIIN, hiSIIN, hiSIIN with antibody depletion of CD8+ T cells, and loSIIN organoids. **(f-g)** Stereoscopic brightfield and fluorescent images of loSIIN colon tumor **(f)** and liver metastases **(g)** 42 days post-transplant. **(h-i)** Example three color IHC (black = CD8, green = CD4, red = FOXP3) **(h)** and automated annotation by convolutional neural network **(i)**. **(j-k)** Representative H&E and three color IHC of loSIIN primary colon tumor **(j)** and liver metastasis **(k)** 42 days post-transplant. Scale bar = 100 μM. **(l)** Quantification of CD8, CD4, and regulatory T cells infiltrating loSIIN and noSIIN tumors by convoluted neural network analysis. Each point represents at least one tumor from a single animal. **(m-n)** Identification of SIINFEKL tetramer-specific CD8+ T cells infiltrating 42 day loSIIN tumors by flow cytometry **(m)** and immunofluorescence **(n)** with *in situ* tetramer staining (green = tumor, white = CD8, red = SIINFEKL tetramer, blue = DAPI).

To quantitatively compare surface MHC-I presentation of SIINFEKL across the expression series of organoids, we performed H-2K^b immunoprecipitation and multiplexed Tandem Mass Tag Mass Spectrometry (TMT-MS) on eluted peptides. While lo^{SIIN} through med^{SIIN} showed near perfect linear correlation between SIINFEKL abundance and mScarlet^{SIIN} protein expression, hi^{SIIN} unexpectedly showed an intermediate abundance (Fig. 2c, Extended Data Fig. 2l). Given that antigen expression in mid^{SIIN} and med^{SIIN} was adjusted by altering translation efficiency, it is possible that ribosomal stalling and proteosomal degradation of mScarlet^{SIIN} are enhanced in these lines, resulting in greater surface presentation. Alternatively, high expression of mScarlet^{SIIN} in hi^{SIIN} may lead to insoluble protein aggregates that are sequestered from antigen processing machinery. Regardless, these results validate the flexibility of our system to modulate surface presentation levels of neoantigen. While the multiplexed format of TMT-MS and high SIINFEKL signal intensity in some lines precludes meaningful intrasample comparisons of the abundance of SIINFEKL versus other epitopes, SIINFEKL abundance in lo^{SIIN} was only 0.4-fold above background signal in hi^{VGF}, consistent with a low level of surface presentation.

Endoscope-guided transplant of shAKPS organoids without antigen (no^{SIIN}) efficiently induced tumors (Fig. 2d-e) and spontaneous metastases to liver and lung, with histology remarkably similar to human CRC (Extended Data Fig. 2m). In contrast, transplant of the highest expression variant (hi^{SIIN}) resulted in CD8^{+} T cell-mediated rejection in all animals (Fig. 1h-i). While transplant of med^{SIIN} also resulted in complete rejection, mid^{SIIN} and dim^{SIIN} formed tumors with intermediate efficiency (Fig. 2e, Extended Data Fig. 2n-o), suggesting a non-discrete effect of neoantigen expression level in tumorigenesis. We also generated shAKPS organoids harboring different epitopes with high affinity for MHC-I, including SIYRYYGL (hi^{SIY}), VGFNFRTL (hi^{VGF}), and ITYTWTRL (hi^{ITY}) (Extended Data Fig. 2p). The latter two are mutant epitopes of *Lama4* (G1254V) and *Alg8* (A506T) that arose in a methylcholanthrene-induced mouse sarcoma, which were reported to be insufficient for tumor rejection but critical for ICB response in a syngeneic flank transplant model[29]. Here, all three epitopes resulted in tumor rejection (Extended Data Fig. 2q), demonstrating that immunogenicity is not idiosyncratic to SIINFEKL but a general feature of high expression of high-affinity epitopes. This also argues that the major genetic features of MSS CRC do not confer cell-autonomous resistance to T cell-mediated killing.

Strikingly, transplant of the lowest expression variant (lo^{SIIN}) induced tumors and metastases with similar efficiency, histology and infiltration as no^{SIIN} organoids (Fig. 2d-l). CD8^{+} T cells were sparse and only modestly increased in lo^{SIIN} tumors, while helper and regulatory T cell infiltration was not significantly different (Fig. 2h-l, Extended Data Fig. 2m). This is characteristic of the immune "cold" landscape of MSS CRC in humans[30]. Importantly, immune escape in lo^{SIIN} tumors did not result from neoantigen ignorance, as advanced tumors were infiltrated by antigen-experienced (CD44^{+}) and specific (SIINFEKL H-2K^{b} tetramer^{+}) CD8^{+} T cells (Fig. 2m-n). Altogether, these results demonstrate that MSS tumors can harbor high-affinity neoantigens despite poor T cell infiltration, and low neoantigen expression is an important mechanism of tumor immune evasion.

Low neoantigen expression limits T cell effector commitment and drives early dysfunction. To investigate why lo^{SIIN} tumors escaped immune rejection, we first compared the kinetics of the antigen-specific T cell response in lo^{SIIN} versus hi^{SIIN} lesions. Low neoantigen expression resulted in both delayed and lower magnitude response (Fig. 3a). Interestingly, this difference was far less pronounced in the caudal and iliac draining lymph nodes (DLNs) (Extended Data Fig. 3a), suggesting that early lo^{SIIN} T cells are either impaired in their ability to traffic to or proliferate within the tumor. The latter is unlikely, however, as lo^{SIIN} and hi^{SIIN} T cells within tumors and DLNs showed no difference in proliferation (Extended Data Fig. 3b). Alternatively, T cells arriving at the tumor may have undergone deletional tolerance[31], resulting in lower numbers. A critical step in the early maturation of functional T cell responses is effector differentiation, characterized by production of cytokines and cytolytic granzymes, such as Granzyme B (GZMB), and loss of progenitor potential. TCF1 is a marker of progenitor potential and is expressed in naïve, memory precursor and memory T cells[32–34]. Consistent with impaired effector differentiation, significantly more antigen-specific T cells from lo^{SIIN} tumors and DLNs were TCF1^{+}/GZMB^{-}, and significantly fewer were TCF1^{-}/GZMB^{+}, at 8 days (Fig. 3b,d-e, Extended Data Fig. 3c). This is unlikely a result of delayed kinetics, as the percentage of TCF1^{-}/GZMB^{+} T cells in lo^{SIIN} tumors at peak response (14 days) remained significantly lower (Fig. 3c,f, Extended Data Fig. 3d). Unexpectedly, the percentage of antigen-specific T cells capable of secreting both TNFα and IFNγ effector cytokines was significantly higher in lo^{SIIN} versus hi^{SIIN}

DLNs and not different in tumors at 8 days (Extended Data Fig. 3e). However, this cytokine-proficient population also showed higher TCF1 and lower GZMB (Extended Data Fig. 3f), suggesting similar lack of effector differentiation. Strikingly, the percentage of double-negative (TCF1$^-$/GZMB$^-$) T cells was greater in lo^{SIIN} versus hi^{SIIN} animals at 8 days (Fig. 3b,g), and became even more pronounced by 14 days (Fig. 3c,h). Absence of TCF1 and GZMB implies lack of progenitor and effector functionality and indicates dysfunction. Indeed, by 14 days TCF1$^-$/GZMB$^-$ antigen-specific T cells in lo^{SIIN} versus hi^{SIIN} tumors showed higher expression of co-inhibitory receptors PD-1, TIM3, LAG3, and 2B4 (Extended Data Fig. 3g), and an increased fraction co-expressing three or all four (Fig. 3i, Extended Data Fig. 3h). To further interrogate functionality of the lo^{SIIN} response, we performed an *in vivo* killing assay[35] by transferring SIINFEKL-loaded "target" splenocytes into tumor-bearing mice. Consistent with rejection of hi^{SIIN} organoids, targets were nearly completely eliminated in the DLNs and spleen 6 hours post-transfer in hi^{SIIN}, but not lo^{SIIN}, animals at 8 days (Fig. 3j-k, Extended Data Fig. 3i). Even at the peak of the lo^{SIIN} response (14 days), killing was incomplete (Fig. 3l-m, Extended Data Fig. 3j), with fewer targets killed per antigen-specific T cell (Fig. 3n). Finally, antigen-specific T cells infiltrating early lo^{SIIN} tumors showed a more clonal TCR repertoire (Extended Data Fig. 3k-l). Altogether, these results demonstrate that low neoantigen expression drives an immediately dysfunctional T cell response with attenuated magnitude, diversity, and per cell functionality.

To determine if these effects are mediated by discrete or continuous levels of neoantigen expression, we characterized the T cell responses to dim^{SIIN}, mid^{SIIN}, and med^{SIIN} organoids at 8 and 14 days. Consistent with the tumor efficiency data (Fig. 2e) and a continuous model, SIINFEKL expression correlated strongly and positively with the magnitude of the initial T cell response (8 days) (Fig. 3o-p). While no correlation was seen at 14 days (Extended Data Fig. 3m-n), likely due to T cell contraction in the intermediate expressing lines, SIINFEKL expression correlated strongly and negatively with the percentage of TCF1$^-$/GZMB$^-$ antigen-specific T cells at this time point (Fig. 3q-r, Extended Data Fig. 3o-p). These results demonstrate a continuous positive effect of neoantigen expression level on the magnitude of the T cell response and effector differentiation.

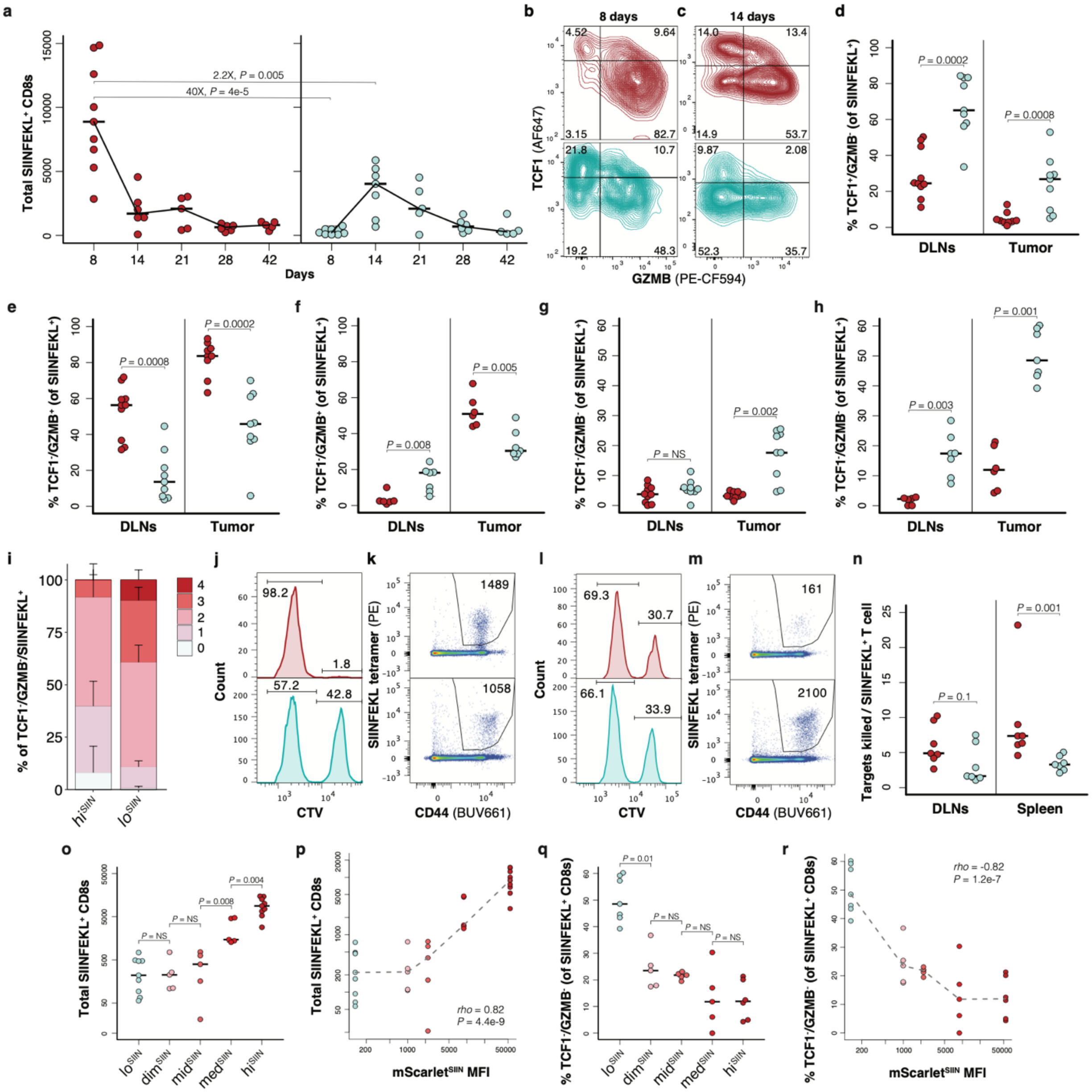


Figure 3 Low neoantigen expression drives impaired T cell effector commitment and dysfunction. (a) Total numbers of CD44+/CD8+ antigen-specific T cells isolated from lesions at 8, 14, 21, 28, and 42 days post- transplant of hiSIIN (red) and loSIIN (blue) organoids by flow cytometry. **(b-c)** Antigen-specific T cell expression of TCF1 versus GZMB in tumors at 8 **(b)** and 14 **(c)** days. **(d-e,g)** Percent of antigen-specific T cells from DLNs and tumors at 8 days positive for TCF1 and negative for GZMB **(d)**, negative for TCF1 and positive for GZMB **(e)**, and double-negative for TCF1 and GZMB **(g)**. **(f,h)** Percent of antigen-specific T cells from DLNs and tumors at 14 days negative for TCF1 and positive for GZMB **(f)**, and double-negative for TCF1 and GZMB **(h)**. **(i)** Percentage of TCF1-/GZMB- antigen-specific T cells from tumors at 14 days expressing 0, 1, 2, 3, and 4 inhibitory receptors (PD-1, TIM3, LAG3, and 2B4). **(j-m)** Representative *in vivo* killing assay histograms of transferred control (weak CTV stain) and SIINFEKL-loaded "target" (strong CTV stain) splenocytes and flow plots of antigen-specific T cells recovered from DLNs at 8 **(j-k)** and 14 **(l-m)** days post-transplant of hiSIIN (red) and loSIIN (blue) organoids. Percent of control and target populations are noted in histograms, and total number of antigen-specific T cells noted in flow plots. **(n)** Target killing normalized to number of antigen-specific T cells recovered in 14 day *in vivo* killing assay. **(o-p)** Total antigen-specific T cells isolated from lesions at 8 days across all expression variant lines **(o)** and versus mScarletSIIN MFI **(p)**. **(q-r)** Percent of antigen-specific T cells from lesions at 14 days double-negative for TCF1 and GZMB across expression variant lines **(q)** and versus mScarletSIIN MFI **(r)**. Dashed lines connect medians. Significance of correlations was assessed by Spearman's rank correlation.

T cells in tumors with low neoantigen expression become progressively dysfunctional. T cell dysfunction in human cancer is often attributed to upregulation of co-inhibitory receptors, terminal differentiation and loss of effector function, or "exhaustion"[36,37]. More recently, it has been shown that terminally-differentiated T cells in cancer are characterized by low TCF1 and high TIM3 expression, and are regenerated by "progenitor exhausted" T cells that have high TCF1 and low TIM3 expression[38–40]. Consistent with progressive dysfunction, terminally-differentiated T cells (TCF1$^-$/TIM3$^+$) made up a greater proportion of the response at 42 compared to 8 days in lo^{SIIN} tumors, but not in hi^{SIIN} rejected lesions (Fig. 4a,b, Extended Data Fig. 4a). In contrast, progenitors (TCF1$^+$/TIM3$^-$) were substantially depleted in lo^{SIIN} tumors by 42 days (Fig. 4a,b, Extended Data Fig. 4b). In addition to TIM3, antigen-specific T cells in lo^{SIIN} tumors at 42 days showed increased expression of PD-1, LAG3, and TIGIT (Fig. 4c). In lo^{SIIN} tumors, T cells negative for TCF1 and triple-positive for TIM3, PD-1, and LAG3 peaked at 42 days (Fig. 4d), consistent with progressive dysfunction. A similar trend was observed in TCF1$^+$/TIM3$^-$ progenitor exhausted T cells double-positive for PD-1 and LAG3 (Fig. 4e). Finally, lo^{SIIN} tumors and DLNs at 42 days showed a reduced percentage of TNFα and IFNγ double-positive antigen-specific T cells (Extended Data Fig. 4c-d), indicating loss of effector functionality. Therefore, in addition to immediate dysfunction, T cells responding to low tumor neoantigen expression undergo progressive exhaustion.

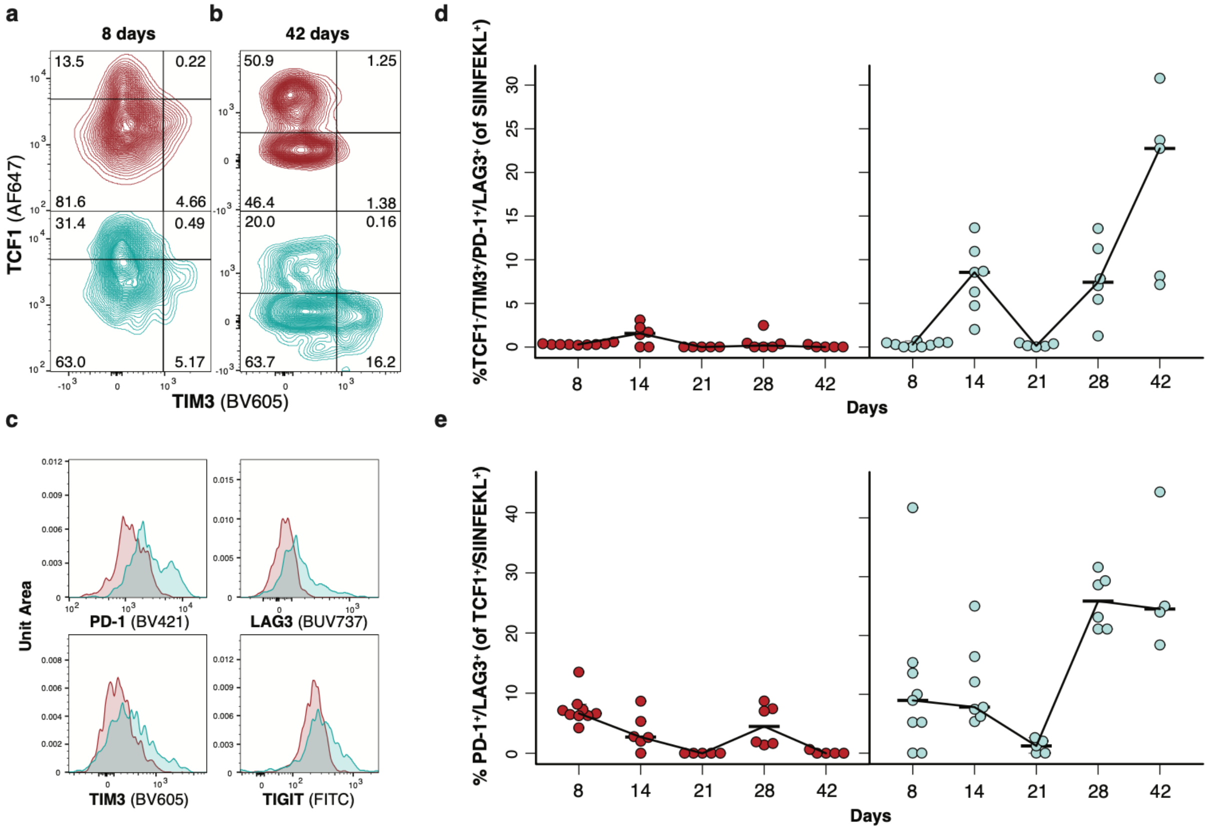


Figure 4 T cells in tumors with low neoantigen expression become progressively dysfunctional.

(a-b) Antigen-specific T cell expression of TCF1 versus TIM3 in tumors at 8 **(a)** and 42 **(b)** days post- transplant. **(c)** Representative histograms of inhibitory receptor expression on antigen-specific T cells from tumors at 42 days post-transplant. **(d-e)** Percent TCF1-/TIM3+/PD-1+/LAG3+ **(d)** and TCF1+/PD-1+/LAG3+ **(e)**
antigen-specific T cells isolated from tumors at 8, 14, 21, 28, and 42 days post-transplant. Red = hiSIIN, blue = loSIIN.

T cell cross priming is limiting in the context of low neoantigen expression. Impaired effector differentiation and early dysfunction are indicative of poor priming, such as occurs in the absence of CD4$^+$ T cell "help"[41]. While lo^{SIIN} and hi^{SIIN} organoids lack a defined MHC-II-restricted model neoantigen, depletion of CD4$^+$ T cells completely rescued formation of hi^{SIIN} tumors (Fig. 5a). Therefore, absence of help is unlikely the mechanism of dysfunction in the lo^{SIIN} model. It is possible that CD4$^+$ T cells are primed against uncharacterized neoantigens in mScarlet, tumor-associated self-antigens, or microbial antigens in the colon microenvironment. Consistent with the importance of neoantigen expression level in priming, hi^{SIIN} tumor formation was partially rescued in *Batf3*$^{-/-}$ mice, which lack conventional cross-presenting dendritic cells (DC1s) (Fig. 5a). To directly test the role of neoantigen expression level in cross presentation, we co-cultured bone marrow-derived dendritic cells (BM-DCs)—comprised of ~13% CD103$^+$ DC1s—with naïve TCR-transgenic T cells specific to SIINFEKL (OT-1) (Extended Data Fig. 5a-b). Compared to BM-DCs loaded with hi^{SIIN} organoids, those loaded with lo^{SIIN} were markedly less capable of priming OT-1 proliferation and effector differentiation (Fig. 5b-e, Extended Data Fig. 5c). On the other hand, *in vitro*-primed OT-1s were equally capable of killing lo^{SIIN} as hi^{SIIN} organoids when co-cultured (Fig. 5f-g). These results argue that neoantigen expression is limiting for T cell cross priming, but not tumor cell recognition by effector T cells.

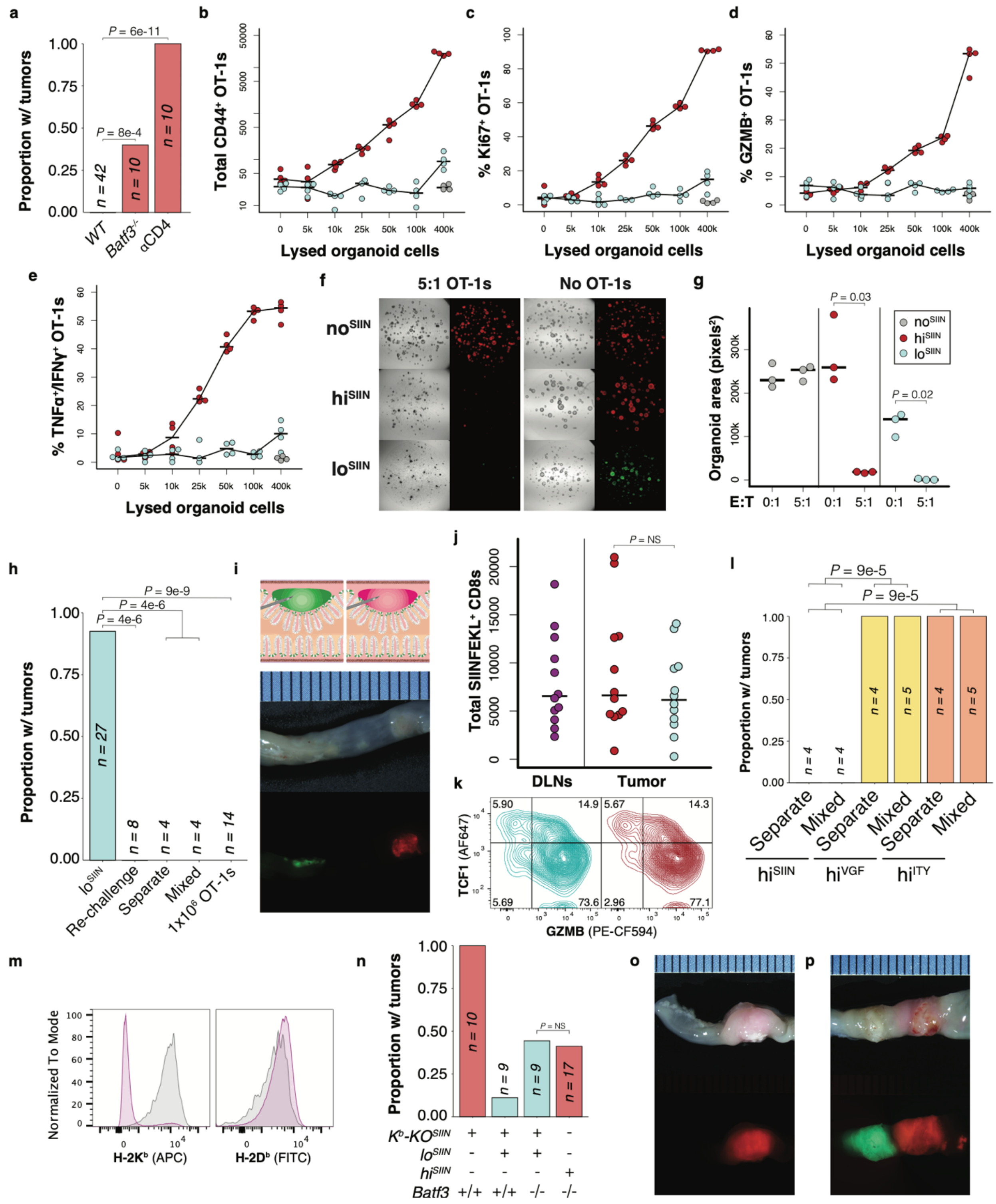


Figure 5 Low neoantigen expression results in poor T cell cross priming. (a) Efficiency of tumor formation 6 weeks post-transplant of hiSIIN organoids into wild-type (*WT*) and *Batf3* knockout mice, and 3 weeks post- transplant into *WT* mice with continuous antibody depletion of CD4+ T cells. **(b-e)** Cross priming assay of 10,000 naïve OT-1s co-cultured with 50,000 activated BM-DCs loaded with a range of water-lysed cells from noSIIN (grey), loSIIN (blue),

and hiSIIN (red) organoids. Flow cytometric quantification of CD44+ **(b)**, Ki67+ **(c)**, GZMB+ **(d)**, and TNFα+/IFNγ+ **(e)** OT-1s. **(f)** Representative images of co-cultures with noSIIN, hiSIIN, and loSIIN organoids and *in vitro*-activated OT-1s at day 4. **(g)** Quantification of co-culture organoid confluence. E:T = effector to target ratio. **(h)** Efficiency of tumor formation 6 weeks post-transplant of loSIIN, loSIIN 4 weeks after transplant of hiSIIN (Re-challenge), loSIIN and hiSIIN at separate sites in the same animal (Separate), loSIIN and hiSIIN at the same sites (Mixed), and loSIIN concurrent with retro-orbital injection of 1x106 *in vitro*-activated OT- 1s. **(i)** Schematic of transplant of loSIIN (green) and hiSIIN (red) organoids at separate sites in the distal colon of the same animal (top), and stereoscopic brightfield (middle) and fluorescent (bottom) images of a co- transplanted colon 8 days post-transplant. **(j-k)** Flow cytometric analysis of antigen-specific T cells isolated from DLNs and lesions 8 days post-co-transplant of hiSIIN (red) and loSIIN (blue) organoids at separate sites in the same animals. **(j)** Total number of antigen-specific T cells. **(k)** Representative expression of GZMB versus TCF1. **(l)** Efficiency of tumor formation 6 weeks post co-transplant of loSIIN with hiSIIN, hiVGF, or hiITY at the same (Mixed) or separate sites in the colon of the same animals. **(m)** Representative histograms of H-2Kb and H-2Db expression on hiSIIN organoids post electroporation with Cas9 and sgRNA targeting the *H2-K1* locus and pre- sorting. Purple = targeted; grey = no targeting control. Organoids were pre-treated with IFNγ. **(n)** Efficiency of tumor formation 4 weeks post-transplant of Kb-KOSIIN, hiSIIN or co-transplant of loSIIN and Kb-KOSIIN in *WT* and *Batf3* knockout mice. **(o-p)** Representative stereoscopic brightfield and merged GFP/RFP fluorescent images of tumors 6 weeks post co-transplant of Kb-KOSIIN and loSIIN organoids in *WT* **(o)** and *Batf3* knockout mice **(p)**.

To further interrogate priming *in vivo*, we performed: 1) re-challenge with lo^{SIIN} organoids 28 days after "vaccination" with hi^{SIIN} organoids, 2) co-injection of lo^{SIIN} and hi^{SIIN} organoids in the same animals, and 3) transfer of *in vitro*-activated OT-1 T cells concurrent with transplant of lo^{SIIN} organoids. All of these approaches resulted in complete rejection of lo^{SIIN} organoids (Fig. 5h), demonstrating unequivocally that efficiently primed T cells are capable of killing tumor cells with low neoantigen expression *in vivo*. Priming in the context of high neoantigen expression also rescued the phenotype of T cells infiltrating lo^{SIIN} lesions. Notably, antigen-specific T cells infiltrating lo^{SIIN} and hi^{SIIN} lesions from the same animals (Fig. 5i) showed similar abundance (Fig. 5j) and overlapping expression of TCF1 and GZMB that is indistinguishable from that of animals transplanted with only hi^{SIIN} organoids (Fig. 5k, Extended Data Fig. 5d-f). To determine if these effects are SIINFEKL specific, we performed co-injection of lo^{SIIN} and hi^{VGF} or hi^{ITY} organoids in the same animals. While hi^{VGF} and hi^{ITY} did not form tumors in any animals, consistent with rejection of these lines when transplanted alone (Extended Data Fig. 1q), 100% of animals developed lo^{SIIN} tumors (Fig. 5l, Extended Data Fig. 5g), demonstrating that rescue of priming by hi^{SIIN} is mediated through increased SIINFEKL expression.

The incomplete penetrance of hi^{SIIN} tumor formation in *Batf3*$^{-/-}$ animals implicates additional mechanisms of priming, such as direct priming by tumor cells or cross priming by non-DC1s. Consistent with this, we detected activated antigen-specific T cells in the colons and DLNs of most of these animals 6 weeks post-transplant, albeit at very low numbers (Extended Data Fig. 5h-j). To assess the role of direct priming, we used CRISPR/Cas9 to generate hi^{SIIN} organoids lacking H-2k^{b} (K^{b}-KO^{SIIN}) (Fig. 5m, Extended Data Fig. 5k). Consistent with failure to present SIINFEKL, these organoids formed tumors with 100% efficiency in immunocompetent hosts. However, co-transplant of lo^{SIIN} with K^{b}-KO^{SIIN} resulted in complete rejection of lo^{SIIN} in all but one animal (Fig. 5n), despite outgrowth of K^{b}-KO^{SIIN} in all animals, demonstrating that SIINFEKL from K^{b}-KO^{SIIN} is efficiently cross presented and this is the dominant mechanism of priming in our model. To further interrogate any potential contribution of direct tumor cell priming, we repeated the co-transplant experiments in *Batf3*$^{-/-}$ animals. In this DC1 deficient context, lo^{SIIN} formed tumors with ~40% efficiency when co-transplanted with K^{b}-KO^{SIIN}, indistinguishable from the efficiency of tumor formation with hi^{SIIN} transplanted alone (Fig. 5n). These results strongly suggest that direct priming by tumor cells is not operative, and that non-DC1s play an important role in anti-tumor T cell cross priming in the colon.

Therapeutic vaccination and agonistic anti-CD40 rescue the poorly primed T cell response.

Given the central role of T cell priming in tumor immune escape in our model, we tested the therapeutic potential of neoantigen vaccination in animals with established tumors. Mice with GFP-expressing tumors, as determined by colonoscopy, were randomly enrolled to receive SIINFEKL containing OVA$_{250-270}$ (CGLEQLESIINFEKLTEWTSS) or non-specific mutant gp100$_{20-39}$ (CAVGALEGPRNQDWLGVPRQL) peptide-based vaccines consisting of a peptide-amphiphile and adjuvant amphiphile-CpG, which were administered at 14 and 21 days post-transplant (Extended Data Fig. 6a). Vaccination with OVA$_{250-270}$, but not non-specific peptide, induced profound expansion of tumor-specific T cells, with an average of 35% of peripheral blood CD8^{+} T cells displaying SIINFEKL specificity (Fig. 6a-b, Extended Data Fig. 6b-c). One week following the second dose, tumors in the OVA$_{250-270}$ vaccine arm showed significantly greater reduction in size relative to the non-specific arm, with 7 of 8 (88%) decreasing in size (Fig. 6c). At termination (6 weeks), tumor burden was significantly reduced in the OVA$_{250-270}$ vaccine arm, with four complete regressions (Fig. 6d). These results suggest that it may be therapeutically tractable to vaccinate against poorly expressed—albeit clonal—neoantigens, and that strict neoantigen expression cutoffs in anti-tumor vaccine pipelines should be re-evaluated.

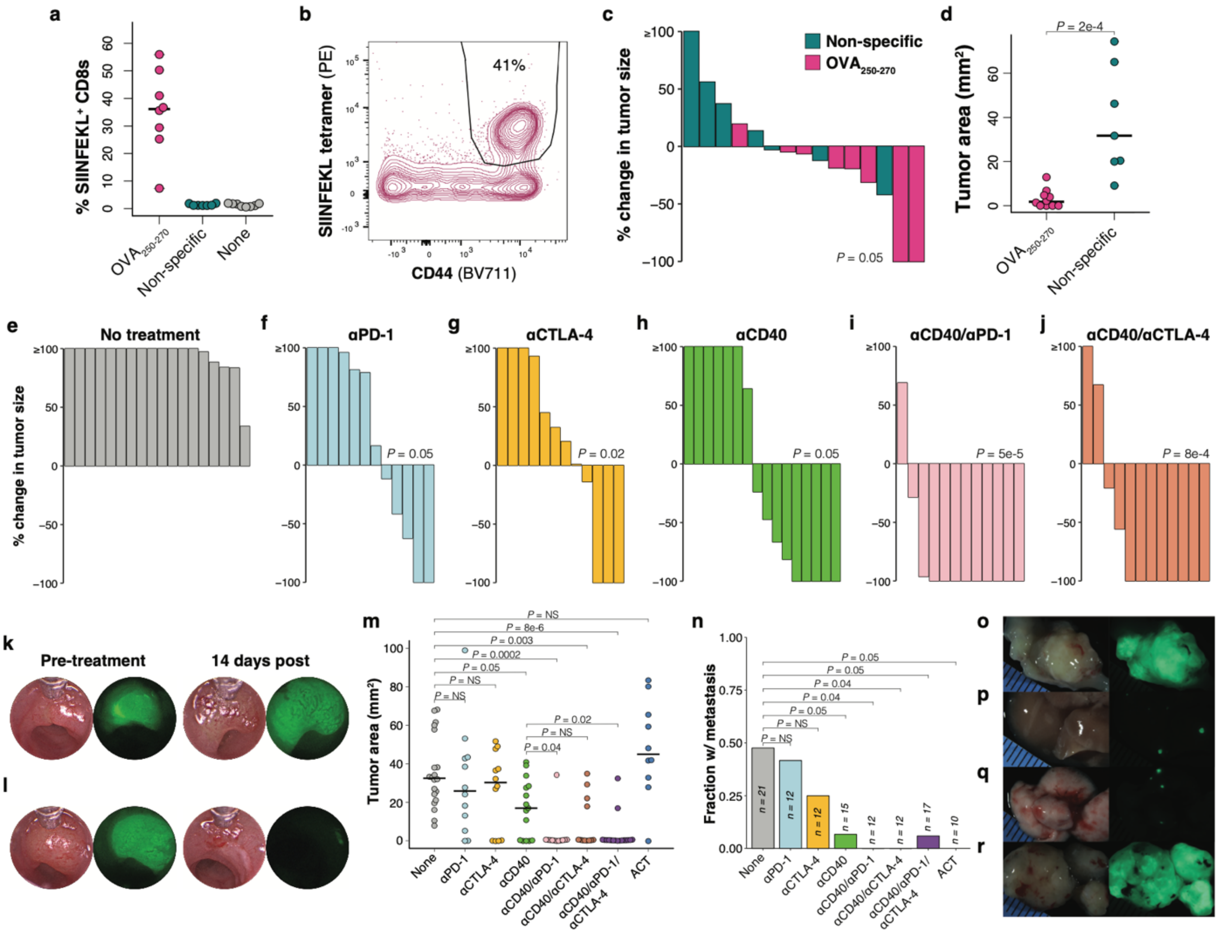


Figure 6 Therapeutic vaccination and agonistic anti-CD40 are efficacious in low neoantigen expressing tumors. (a) Percent of total peripheral blood CD8+ T cells that are antigen specific (CD44+/SIINFEKL tetramer+) following two weeks (two doses) of OVA250-270, non-specific, or no peptide-based vaccination in loSIIN tumor- bearing mice. **(b)** Representative flow plot of peripheral blood antigen-specific CD8+ T cells from OVA250-270 vaccinated mouse. **(c)** Change in loSIIN tumor size as measured by longitudinal colonoscopy following 14 days (two doses) of OVA250-270 or non-specific vaccination. Significance was assessed by Wilcoxon rank-sum test of percent change in tumor size. **(d)** Primary tumor sizes at necropsy 28 days post-vaccine regimen initiation. **(e- j)** Immunotherapy preclinical trial of mice bearing loSIIN tumors. Waterfall plots show change in tumor size after 14 days of treatment, as determined by colonoscopy. **(k-l)** Representative colonoscopy white light and fluorescent images of tumors pre- and post-treatment from mice receiving no treatment **(k)** and αCD40/αPD- 1/αCTLA-4 **(l)**. **(m)** Primary tumor sizes at necropsy 28 days post-treatment initiation. ACT = adoptive cell transfer of OT-1s. **(n)** Fraction of mice with any metastases (liver, lung, or omentum). **(o-r)** Representative brightfield and fluorescent images of primary colon tumor **(o)**, liver **(p)**, lung **(q)**, and omental **(r)** metastases from an αPD-1-treated mouse 28 days post-treatment initiation.

We next asked whether more readily deployable antibody-based immunotherapies are efficacious in our model. Agonistic antibodies against the CD40 receptor (αCD40) enhance priming by potentiating the co-stimulatory function of antigen-presenting cells[42]. αCD40 is efficacious in preclinical mouse models of pancreatic ductal adenocarcinoma (PDAC), particularly when combined with ICB and immunogenic chemotherapy[43]. Recently, a phase Ib clinical trial in PDAC with αCD40 (APX005M), αPD-1 (nivolumab), and gemcitabine/nab-paclitaxel has shown promising early results[44]. This is particularly exciting in light of the low TMB and immunogenicity of PDAC, which, like MSS CRC, is refractory to ICB[45]. Therapeutic combinations with αCD40 may be able to rescue or generate new T cell responses against weak affinity or poorly expressed neoantigens, or against tumor-associated self-antigens that lack high affinity T cell clones due to central tolerance. However, clinical studies in CRC are lacking.

We performed preclinical trials in mice bearing lo$^{\mathrm{SIIN}}$ colon tumors starting 14 days post-transplant (Extended Data Fig. 6a) with single agents αCD40, αPD-1, and αCTLA-4, as well as combinations αCD40/αPD-1, αCD40/αCTLA-4, and αCD40/αPD-1/αCTLA-4. Response was evaluated by colonoscopy at 28 days post-transplant following Response Evaluation Criteria In Solid Tumors (RECIST). All animals in the no treatment arm presented with progressive disease, while 2 of 12 (17%), 3 of 12 (25%), and 5 of 16 (31%) showed complete responses in the αPD-1, αCTLA-4, and αCD40 arms, respectively. Response was notably better in all combination arms, with 9 of 12 (75%), 8 of 12 (67%), and 12 of 17 (71%) complete responses in the αCD40/αPD-1, αCD40/αCTLA-4, and αCD40/αPD-1/αCTLA-4 arms, respectively (Fig. 6e-m, Extended Data Fig. 6d). Comparing all combination arms as a whole against single agent αCD40 showed significantly more objective responses ($P = 0.02$) and complete responses ($P = 0.01$, Fisher's exact test). Interestingly, adoptive cell transfer (ACT) of one million *ex vivo*-activated OT-1 T cells at

14 days post-transplant significantly delayed tumor growth but only resulted in one complete response (Fig. 6m, Extended Data Fig. 6e), suggesting that transferred T cells rapidly become dysfunctional.

Despite initially delayed tumor growth in the single ICB arms, no significant difference in final tumor burden was observed at necropsy, suggesting only transient effect in the majority of tumors (Fig. 6m). In addition, incidence of metastasis was not significantly decreased in single ICB arms (Fig. 6n-r). These results are reminiscent of the poor response to ICB seen in MSS CRC and demonstrate that ICB is only modestly effective at rescuing a poorly primed T cell response. In contrast, single agent αCD40 significantly decreased primary tumor size at endpoint, while combination with ICB significantly reduced tumor size further (Fig. 6m). All treatment arms with αCD40 resulted in significantly reduced rates of metastasis (Fig. 6n), although this could reflect the absence of primary tumors in many of these animals. However, the combined rate of metastasis in animals with progressive disease across all αCD40 arms was still significantly reduced (Extended Data Fig. 6f). Interestingly, while ACT had no effect on reducing primary tumor size at endpoint, it resulted in complete control of metastatic tumor burden (Fig. 6n).

To determine if therapy resistance is mediated by down-regulation of antigen expression or MHC-I, we isolated *ex vivo* lo$^{\mathrm{SIIN}}$ tumor-derived organoids from two αCD40/αPD-1 escapers, three αPD-1 escapers, and three mice that received no treatment. All lines showed comparable sensitivity to IFNγ stimulation, expression of H-2K^{b} and mScarlet$^{\mathrm{SIIN}}$, and were similarly sensitive to killing when co-cultured with activated OT-1s (Fig. 7). These results suggest that low neoantigen expression in MSS CRC obviates the need for dysregulation of antigen presentation, and is consistent with the lower frequency of such events in MSS versus MSI cancers[17]. Critically, MSS tumors may remain sensitive to T cell killing if priming against poorly expressed neoantigens can be rescued (Fig. 8). Why some tumors escape treatment despite retaining intrinsic sensitivity to tumor-specific T cell killing is a question of considerable clinical value for future investigation, and one in which our novel platform is ideally positioned to address.

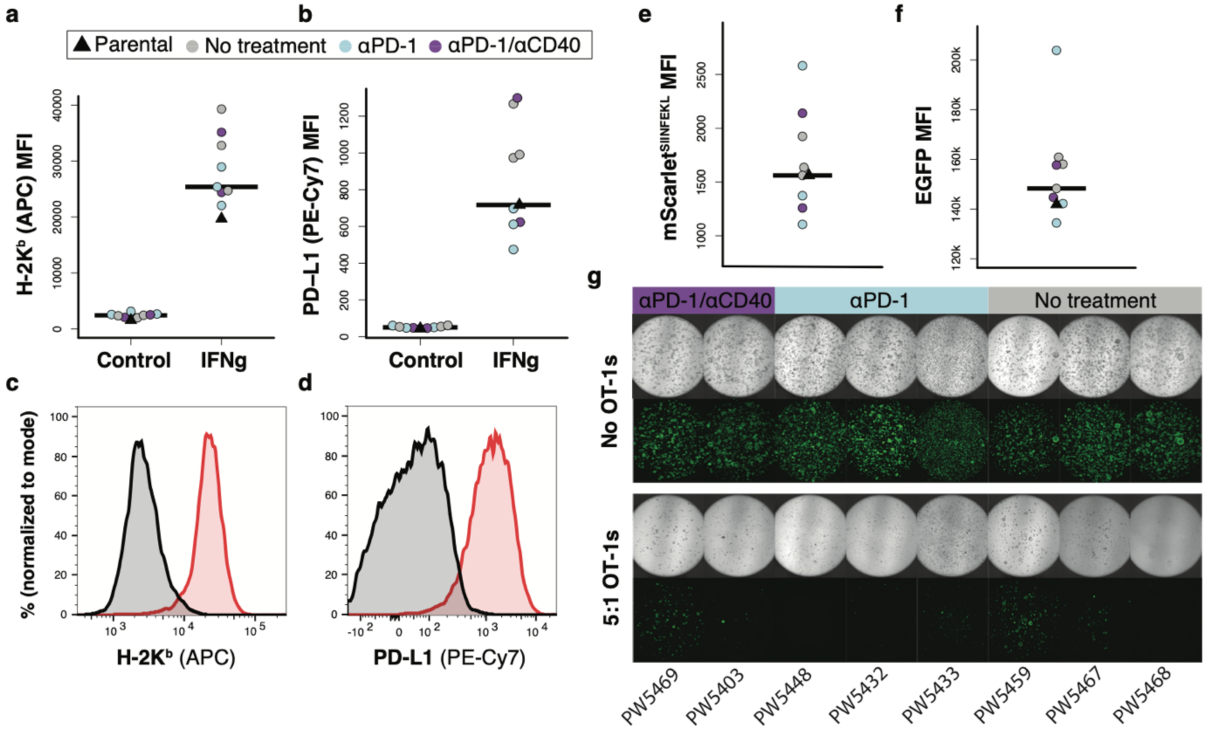


Figure 7 Treatment naïve and refractory low neoantigen expressing tumors retain antigen presentation. **(a-d)** Flow cytometric analysis of H-2Kb and PD-L1 MFI **(a-b)** and representative histograms of expression **(c-d)** following IFNγ stimulation in loSIIN *ex vivo* tumor-derived organoids from mice with and without treatment. **(e-f)** Flow cytometric analysis of mScarletSIIN **(e)** and EGFP **(f)** expression in *ex vivo* tumor-derived organoids. **(g)** Representative images of co-cultures with *ex vivo* tumor-derived organoids and *in vitro*-activated OT-1s at day 4.

Discussion

The poor response of most CRC to immunotherapy represents a major unmet clinical need.

Mouse models have provided invaluable insights into T cell dysfunction in cancer, but none to our knowledge recapitulate essential features of human CRC while facilitating detailed study of antigen-specific T cells. Here, we developed colonoscopy-guided models that enable comparison of functional versus dysfunctional tumor-specific T cell responses in a context highly faithful to the microenvironment, genetics, histopathology, and metastatic progression of the human disease.

We found that tumors from all MSS CRC patients in the TCGA harbored clonal neoantigens with high predicted HLA-I affinity, but that these were broadly expressed at lower levels compared to those from MSI CRC. This raises the intriguing possibility that poor immunogenicity in MSS CRC and other immune cold cancers is driven by both lower burden and lower expression of neoantigens. Consistent with this notion, our low neoantigen-expressing model, like MSS CRC, demonstrated poor T cell infiltration and response to ICB. Leveraging the defined antigen in our model, we showed that neoantigen expression is an analog input that tunes the quality of anti-tumor T cell priming *in vivo*. Limiting neoantigen expression shifts priming towards a tolerogenic response characterized by reduced magnitude, diversity, effector commitment, and per T cell functionality. These findings provide broader context to previous flank transplant studies that found lower levels of epitope expression[15] or MHC-I binding affinity[16] facilitate tumor immune escape, and a study that described early T cell dysfunction in an SV40 large T-antigen-driven model of liver cancer[46]. Indeed, a general feature of early immune evasion in cancer may be that T cell dysfunction begins as a tolerogenic program initiated during priming with insufficient antigen stimulation, in addition to a lack of local inflammatory and/or co-stimulatory cues. By extension, it is likely that immune responses against clonal neoantigens in cancer, at least those acquired early in tumorigenesis, are poorly primed and tolerogenic—axiomatic to their failure to restrain tumor outgrowth. We showed that rescuing early priming is sufficient to prevent tumor initiation with 100% efficiency in our model, consistent with the fact that early neoplasia lack the immunosuppressive mechanisms to evade efficiently primed T cell responses. We also showed that poorly primed T cells in our model undergo progressive exhaustion, in line with prevailing literature and suggesting that T cell dysfunction in cancer is a heterogenous state shaped by multiple processes operative early and late in tumorigenesis. Future studies combining lineage tracing and single cell sequencing technologies should help disentangle the contribution of distinct genetic and epigenetic programs to T cell dysfunction in cancer.

Therapeutically targeting priming via αCD40 was highly efficacious in our model, particularly in combination with ICB, resulting in complete responses in the majority of animals. While ICB alone had no effect on the rate of metastasis, αCD40 and ACT almost completely prevented metastases, even in mice with progressive primary disease. Therefore, targeting priming may be especially efficacious against early metastatic lesions that may not be detected at the time of treatment. These results establish the preclinical utility of our model and highlight the therapeutic promise of combined αCD40 and ICB in the treatment of MSS CRC and other immune cold cancers. Given that no adequately powered clinical trials of αCD40 in CRC have been initiated to date, these results warrant further clinical evaluation.

Our demonstration that anti-tumor immunity against a poorly expressed neoantigen can be rescued by therapeutic vaccination is particularly relevant to analogous efforts ongoing in humans[47,48], justifying exploration of candidate neoantigens that by current practice might be considered too poorly expressed. In a recent consortium study integrating neoantigen prediction pipelines from 28 unique research teams, it was concluded that expression, among other variables, is an important predictor of neoepitope immunogenicity, and a minimum threshold of >33 transcripts per million was imposed[47]. However, neoepitope immunogenicity in this and other studies was validated by experimentally measuring reactivity of existing T cells in patient blood or tumor[47–49]. Given that tumor-specific T cell responses in advanced tumors are by definition dysfunctional, particularly when poorly primed as we have demonstrated in our model, it is likely that the functional readouts of these assays are limited by low sensitivity. Additionally, it is possible that therapeutic priming against neoantigens overlooked by these assays could unleash productive T cell responses from naïve T cells or reservoirs of clonally expanded precursors in lymphoid tissues. Indeed, a recent phase 1b trial (NEO-PV-01) found that while personalized vaccines elicited *de novo* T cell responses in all patients, the vast majority showed no detectable responses in peripheral blood prior to vaccination[48]. Our findings argue that RNA sequencing in these trials should be performed at sufficient depth to distinguish lack of expression from dropout due to poor coverage, and that any detectable allele-specific expression of a clonal neoantigen is sufficient to nominate it for vaccination. Indeed, it has been shown *in vitro* that effector CD8$^+$ T cells can lyse target cells presenting only three cognate epitope-bound complexes of MHC-I[50]. Therefore, tumor cells likely must undergo near absolute loss of neoepitope presentation to render responding T cells truly ignorant.

Altogether, the results of our study describe a model in which tumors harboring a poorly expressed neoantigen elicit tolerogenic T cell responses and evade immune deletion, whilst remaining vulnerable to destruction by those same T cell responses following therapeutic priming (Fig. 8). It will be important to determine if therapies that potentiate priming mediate their effects predominantly through naïve T cells or reservoirs of antigen experienced T cells in the tumor bed or lymphoid tissues. Finally, the flexible organoid-based system developed here should facilitate a broad range of future studies of immune evasion and immunotherapy response in faithful models of cancer.

Supplementary materials: All supplementary figures, tables, and data our available from our publication in *Nature Cancer*.

Methods

TCGA neoantigen prediction analysis

350 colon adenocarcinoma patients and 76 rectum adenocarcinoma patients were analyzed from The Cancer Genome Atlas COAD and READ studies. These represented all samples with tumor (−01A) and matched-normal whole-exome sequencing (WES), RNA-Seq, and mutation annotation format (MAF) files available. Sequencing data were obtained as Binary Alignment Maps (BAMs) files aligned to GRCh38. HLA-A, -B, and -C alleles were called using OptiType, v1.3.1[20]. Tumor/normal WES BAMs were used to create inputs to OptiType. Reads were filtered to those mapping to the HLA region (chr6:28510120–33480577 in GRCh38) with Samtools v1.10[61], converted to FASTQ and filtered with RazerS 3 v3.5.8[62], as recommended in the OptiType documentation. OptiType was run with default parameters.

A custom Python v2.7.13 script was employed to evaluate concordance between normal and tumor HLA allele calls. 1917/2100 alleles (91.3%) in the COAD cohort and 428/456 (93.9%) alleles in the READ cohort were consistent between tumor and normal WES-based calls; the tumor allele was accepted as the final call to resolve discrepancies between calls from tumor and normal sample sequencing data. Patient MAFs were converted to Variant Call Format (VCF) and filtered to SNVs only. All non-PASS variants were removed except for some in *KRAS* and *TP53* that had been marked as either panel_of_normals, clustered_events, or homologous_mapping_event in the TCGA MAF files.

Indels were called using Strelka2 v2.9.2[63] and Scalpel v0.5.4[64]. Scalpel was run with default parameters, with a bed file derived from the CGHub bitbucket account (https://cghub.ucsc.edu; whole_exome_agilent_1.1_refseq_plus_3_boosters.targetIntervals.bed), with coordinates converted to GRCh38. Scalpel failed to call variants for 7/426 patients due to excessive read buildup at some loci; these samples were excluded from downstream analysis. VCF files containing the union of PASS variants from Strelka2/Scalpel were annotated with variant allele frequencies (VAF) from Strelka2 output and merged with corresponding SNV VCFs.

Variant consequence was annotated using Ensembl Variant Effect Predictor (VEP) v99[65] with Wildtype and Downstream plugins and the following parameters: --symbol, --terms=SO, --cache, --offline, --transcript_version, --pick. The --pick parameter was reordered from default to report transcript with most extreme consequence for each variant: rank, canonical, appris, tsl, biotype, ccds, length, mane. Neoepitopes were predicted with HLA allele calls and variant effect predictions using pVACtools v1.5.7[25]. Mutant peptides were generated for lengths 8- through 11-amino acids. MHC:peptide binding affinity was predicted for all peptide:MHC allele pairs with NetMHC-4.0, NetMHCpan-4.0, SMM v1.0, and SMMPMBEC v1.0[21–24], and the median value across all affinity predictions was taken.

Only neoantigens with evidence of expression (RNA-seq FPKM Upper Quartile Normalized (FPKM-UQ) > 0) were included in analyses. Tumor purity estimates (ABSOLUTE algorithm[26]) for TCGA COADREAD were acquired from a previous publication[66]. Neoantigen clonality was estimated by dividing WES level VAF by ABSOLUTE purity (adjVAF), with adjVAF ≥ 0.5 considered clonal. This is an estimate of clonality only, as other factors not considered here can also influence VAF.

Tumor RNA-Seq BAMs were analyzed to detect and quantify SNV expression at the transcriptional level. For all SNVs corresponding to predicted neoepitopes, the corresponding chromosomal coordinates were used as inputs to bam-readcount v0.8.0, which was run with default parameters (-b 20) to obtain the sequencing depth and read counts for reference and alternative alleles at each position. A custom Python v2.7.13 script was constructed to parse the output from bam-readcount and to quantify RNA VAF. Allele-specific expression was calculated as the product of RNA VAF and corresponding gene expression (FPKM-UQ). While neoantigens with no gene level expression were excluded from analysis, those with RNA VAF = 0 were included due to poor RNA-Seq coverage at many SNVs, and the high likelihood of detection failure versus true lack of expression.

Mice

Mice were housed in the animal facility at the Koch Institute for Integrative Cancer Research at MIT with a 12-hour light/12-hour dark cycle with temperatures within 68–72°F and 30–70% humidity. All animal use was approved by the Department of Comparative Medicine (DCM) at MIT and the Institutional Animal Care and Use Committee (IACUC). $Apc^{flox/flox}$[67], $Kras^{LSL-G12D}$[68], $Trp53^{flox/flox}$[69], $Rag2^{-/-}$[70], *OT-1*[71], $R26^{Cas9-2A-EGFP}$[72] and $Batf3^{-/-}$[73] mice were maintained on a pure C57BL/6 background. Approximately equal numbers of male and female mice between 6 to 12 weeks of age were used for all experiments. Organoids were derived from female C57BL/6 mice, allowing transplant into male recipients without minor histocompatibility antigen-driven responses. lo^{SIIN}, dim^{SIIN}, mid^{SIIN}, med^{SIIN}, hi^{SIIN}, lo^{VGF}, and lo^{ITY} organoids were transplanted into $R26^{Cas9-2A-EGFP}$ mice, which are tolerant to EGFP. In some instances, hi^{SIIN} organoids were transplanted into wild-type mice, with no differences in phenotype observed between host genotypes. hi^{VGF}, hi^{ITY}, and K^{b}-KO^{SIIN} organoids were transplanted into wild-type or $R26^{Cas9-2A\text{-}EGFP}$ mice, whereas co-injection experiments with lo$^{\text{SIIN}}$ were performed only in $R26^{Cas9-2A\text{-}EGFP}$ mice.

Organoid isolation and transformation

Normal colon crypts were isolated from wild-type female C57BL/6 mice as previously described[11]. Crypts and organoids were cultured in 60 µL domes comprised of 10 µL conditioned L-WRN (for primary crypts and wild-type organoids) or minimal media (for *Apc* knockdown organoids) and 50 µL of growth-factor reduced phenol-red free Matrigel Matrix (Corning) in 24-well TC-treated Olympus plates (Genesee Scientific). Conditioned L-WRN media was produced as previously described[74]. Minimal media is comprised of 50X B-27 Serum-Free Supplement (Thermo Fisher), Penicillin/Streptomycin (Corning), and 100x GlutaMAX (Thermo Fisher). Lentivirus was produced in HEK-293 cells (ATCC) and concentrated as previously described[75], and functional titers (Cre activity, mScarlet/EGFP fluorescence) measured as previously described[76]. Confluent organoids were dissociated to single cells using TrypLE Express (Thermo Fisher), diluted, and washed with PBS, and resuspended in 1.5 mL of appropriate media with 10 µM of Y-27632 (Sigma-Aldrich). Organoids were divided into three wells of a 24-well plate and transduced with 10k to 100k transduction units (TU) of virus. The plate was then spun at 600g for 1 hour and incubated for 4 hours at 37 °C, after which organoids were plated in Matrigel. Adeno-Cre was used to recombine $Kras^{LSL\text{-}G12D}$ and $Trp53^{flox/flox}$ in normal KP organoids, after which organoids were grown in L-WRN + Y-27632 for 4 days and then selected for 1 week with 10 µM NUTLIN-3A (Sigma-Aldrich). Complete recombination of $Kras^{LSL\text{-}G12D}$ and $Trp53^{floxl/flox}$ was confirmed by PCR using published primers[68,69]. Next, organoids were infected with shApc-expressing lentiviruses. To select for stable integration, organoids were grown in minimal media lacking WNT (described above) one week after infection. Organoids with lowest multiplicity of infection (~0.5), as determined by mScarlet or GFP fluorescence, were chosen. Finally, organoids were infected with integration-deficient lentivirus expressing Cas9 (lenti CRISPR v2)[77] and sgRNA against *Smad4* (5’-GATGTGTCATAGACAAGGT-3’)[37], and selected by addition of TGFβ. Integration deficient lentivirus was generated using a D64V mutant psPax2 packaging vector[78], and absence of integration was confirmed by absence of Cas9 protein by Western blot, and sensitivity to puromycin killing.

MHC-I immunoprecipitation and peptide isolation

MHC-I (H-2K^b) peptide isolation was performed on x210 20 μl plugs per triplicate for each organoid line using a modified immunoprecipitation and protein filtration protocol, as described previously[79]. Organoids were grown to confluence over 3 days before stimulation with 10 ng/mL murine IFNγ (PeproTech) for 18 hours prior to harvest. Organoids were washed with PBS and mechanically liberated by vigorous pipetting in PBS. Cells were washed twice in 50 mL PBS and pellets snap frozen in liquid nitrogen. Pellets were lysed in 2 mL of lysis buffer containing 50 mM Tris pH 8, 100 mM NaCl, 1 mM EDTA, 1% Triton X-100, 60 mM octylglucopyranoside (Sigma), 20 mM iodoacetamide, 10 U DNase, and 1x Halt protease inhibitors (Pierce). Isolations were performed with 40 μL (bed volume) of rProtein A Sepharose beads (GE Healthcare) preloaded with 1 mg anti-H-2K^b antibody (Y3, BioXcell). Peptides were eluted in 500 μL of 10% acetic acid and purified with 10 kDa MWCO spin filters (PALL Life Science).

Tandem mass tag mass spectrometry (TMT-MS)

Dried down MHC-I eluted peptides were resuspended in 100 μL triethylammonium bicarbonate buffer and labeled with TMT16plex (Pierce, Rockford, IL, USA). Samples were then mixed and cleaned with C18 ZipTip (Millipore Sigma). 1/5 sample was used for one LC/MS/MS analysis. Samples were again dried and reconstituted in 2% formic acid (FA) for MS analysis. Peptides were loaded with the autosampler directly onto a 50cm EASY-Spray C18 column (ES803a, Thermo Scientific). Peptides were eluted from the column using a Dionex Ultimate 3000 Nano LC system with a 5 min gradient from 1% buffer B to 5 % buffer B (100 % acetonitrile, 0.1 % formic acid), followed by 84.8 min gradient to 25%, and a 15.2 min gradient to 35%B, followed by a 12min gradient to 60%B, followed by a 4 min gradient to 80%B, and held constant for 4 min. Finally, the gradient was changed from 80% buffer B to 99% buffer A (0.1% formic acid in water) over 0.1 min and held constant at 99% buffer A for 19.9 more minutes. The application of a 2.2 kV distal voltage electrosprayed the eluting peptides directly into the Thermo Exploris480 mass spectrometer equipped with a FAIMS and an EASY-Spray source (Thermo Scientific). Mass spectrometer-scanning functions and HPLC gradients were controlled by the Xcalibur data system (Thermo Scientific). MS1 scans parameters were 60,000 resolution, scan range m/z 390–1500, AGC at 300%, IT at 50ms. MS2 scan parameters were either at 45,000 or 60,000 resolution, isolation width at 0.7, HCD collision energy at 30%, AGC target at 300% and IT set to 300ms. Cycle time for MS2 was 1sec for each MS1 scan. The scan cycle MS1/MS2 was repeated for FAIMS voltages at −40V, −60V and −80V.

Tandem mass spectra were searched with Sequest (Thermo Fisher Scientific, San Jose, CA, USA; version IseNode in Proteome Discoverer 2.5.0.400). Sequest was set up to search a mouse uniprot database (database version July 3, 2020; 55650 entries containing common contaminants and the three proteins mScarlet-SIINFEKL, mScarlet-VGFNFRTL and EGFP) assuming no digestion enzyme (unspecific). Sequest was searched with a fragment ion mass tolerance of 0.02 Da and a parent ion tolerance of 10.0 PPM. TMTpro was added as a fixed modification on K and N-terminus of peptides. Oxidation of methionine was specified in Sequest as a variable modification. Resulting peptides were filtered to exclude peptides with an isolation interference of >30% and ppm error >+/−3ppm of the median ppm error of all PSMs. SIINFEKL intensity across samples was normalized by the overall abundance of all peptides detected in each sample.

Organoid CRISPR/Cas9 RNP electroporation

Confluent hi^{SIIN} organoids were dissociated to single cells as described above and resuspended in 100 μL OPTI-MEM. Ribonucleoprotein (RNP) complexes were formed by mixing 1.64 μL (0.1 nmol) Alt-R Cas9 (IDT) with 3 μL (0.3 nmol) synthetic sgRNA (Synthego) and incubating for 10–20 minutes at room temperature. Cells were then added to the RNP mix, 100 μL transferred to a 2 mm gap cuvette (Bulldog Bio), and electroporated using a NEPA21 electroporator (Bulldog Bio) with the following poring pulse parameters: 175 V, 5 msec length, 50 msec interval, 2 pulses, 10% decay rate, + polarity; and transfer pulse parameters: 20 V, 50 msec length, 50 msec interval, 5 pulses, 40% decay rate, +/− polarity. Electroporated organoids were resuspended gently in pre-warmed minimal media and incubated at 37°C for 15 minutes before plating in Matrigel. sgRNAs used for electroporation are: *H-2K*^{b}: 5’-CAAUGAGCAGAGUUUCCGAG-3’; previously published *B2m* sequence[79]: 5’-UUGAAUUUGAGGGGUUUCUG-3’.

Colonoscopy-guided injections

Orthotopic injection of lentivirus and organoids was performed similarly to previously described methods[11,12]. Intact organoids were always harvested two days post-passaging by washing in PBS and dissociating in Dispase. Matrigel was broken up by gentle scraping and pipetting four times using a 1 mL pipette and incubating at 37°C for 15 minutes. Organoids were washed thoroughly in PBS and resuspended in OPTI-MEM with 10% Matrigel at 50 organoids per μL. Intact organoids and lentivirus (20,000 or 100,000 TU/μL) were injected via Hamilton syringe (Hamilton, 7656–01) and custom injection needle (Hamilton, 33-gauge, small Hub RN NDL, 16 inches, point 4, 45-degree bevel, like 7803–05) fed through the working channel of the colonoscope and inserted into the colonic mucosa at ~30°. ~50 µL was delivered per injection, resulting in large "blebs" within the mucosa.

Tissue preparation and flow cytometry

Colon draining lymph nodes (DLNs, caudal and iliac) were harvested and mechanically dissociated in RPMI1640 (Corning) with 5% heat-inactivated fetal bovine serum (HI-FBS) (harvest media). Tumors were identified using a Dual Fluorescent Protein Flashlight, Model DFP-1 (Nightsea), dissected and placed in a digestion buffer containing 500 Units/mL Collagenase Type 1 (Worthington) and 20 µg/mL DNAse (Sigma-Aldrich) in harvest media, minced using surgical scissors and digested at 37 °C for 40 minutes with gentle agitation. Tumors were then further dissociated with a gentleMACS Octo Dissociator (Miltenyi Biotec) on the tumor_imp1.1 setting and filtered through a 100 µM filter. DLN and tumor preparations were divided for immediate staining or peptide stimulation. Intravenous CD45 staining prior to sacrifice of animals (to differentiate tissue-infiltrating versus circulating T cells) was not routinely performed, as this stained less than 1% of total SIINFEKL-specific T cells.

Live/dead staining (ghost ef780 (Corning), 1:500) was performed in PBS and surface stains in FACS buffer (1 mM EDTA, 25 mM HEPES, 0.5% HI-FBS in PBS). Cells were fixed for 1 hour at room temperature in Fixation/Permeabilization Concentrate (Thermo Scientific) diluted 1:3 in Fixation/Permeabilization diluent (Thermo Scientific) and washed in permeabilization buffer (Thermo Scientific). Intracellular staining was performed in permeabilization buffer overnight at 4 °C. Cells were washed and resuspended in FACS buffer for analysis on a BD LSRFortessa 4 laser, 18 color flow cytometer running BD FACSDiva v8.0 software. Results were analyzed in FlowJo v10.4.2. Single lymphocytes were gated first on FSC-A versus SSC-A and then FSC-A versus FSC-H. Then, live CD8$^+$ T cells were gated on positive CD8α and negative ghost ef780 staining. Antigen-specific CD8$^+$ T cells were further gated on CD44 and tetramer positivity. Expression of additional markers was analyzed specifically in this antigen-specific CD8$^+$ T cell population in all flow cytometric experiments on T cells presented in this manuscript.

Antigen-specific in vivo killing assay

Splenocytes were prepared for *in vivo* transfer as described by Durward et. al, 2010[45]. Briefly, spleens were harvested from female C57Bl/6 mice, red blood cells lysed using ACK Lysing Buffer (Thermo Fisher), and cells resuspended in PBS in a round bottom 96-well plate at 1×10^8 cells per mL. Half of the wells were pulsed with SIINFEKL peptide (Anaspec) at 1 μg/mL, followed by labeling with the membrane dye CellTrace Violet (CTV; Thermo Fisher) at 20 μM. The remaining wells were labeled with 2 μM CTV. Peptide loaded "target" and unloaded control splenocytes were then mixed 1:1 and 200 μL retro-orbitally injected (2×10^7 cells) into experimental animals 8- and 14-days post-transplant of organoids. DLNs and spleens were harvested 6 hours later and processed for flow cytometry as described above. Target and control splenocytes were identified by live/dead staining and CTV labeling intensity, and percent target killing determined relative to the control population. Targets killed per antigen-specific T cell was determined by dividing the total number of targets killed (control minus target splenocytes) by the total number of SIINFEKL tetramer$^+$ CD8$^+$ T cells. This metric was meaningful at 14 days when target killing was incomplete in both lo^{SIIN} and hi^{SIIN} animals but precluded at 8 days by effectively complete target killing in hi^{SIIN} animals.

Peptide stimulation for cytokine staining

Samples were prepared as described above, and prior to surface staining were stained with antibodies for Ly-6G (BioLegend, 1A8, 1:200), EpCAM (BioLegend, G8.8 1:80) and F4/80

(BioLegend, BM8, 1:80) in FACS buffer for 30 minutes at 4°C and depleted using Dynabeads Goat Anti-Mouse IgG kit (Thermo Fisher) following manufacturer recommendations. T cells were then stimulated in T-cell media (RPMI-1640 with 10% HI-FBS, 20 mM HEPES, 1 mM Sodium Pyruvate, 2 mM L-Glutamine, 50 µM β-mercaptoethanol, 1x non-essential amino acids, and 0.5x Penicillin/Streptomycin) with 1:1000 GolgiPlug (BD) and 2 µM Monensin Solution (BioLegend), and 1 µM SIINFEKL peptide (Anaspec) for 3 hours at 37 °C. Cells were washed and stained for surface and intracellular markers as described above.

Bone marrow-derived dendritic cell (BM-DC) isolation and OT-1 co-culture

Bone marrow from C57Bl/6 mouse femurs and tibias was isolated, red blood cells lysed, and cells plated at 1.5×10^6 cells per mL and cultured in T-cell media (described above) plus 600 ng/mL recombinant human Flt-3L-Ig (hum/hum, BioXcell) and 5 ng/mL recombinant mouse GM-CSF (BioLegend). After 1 week, BM-DCs were switched to fresh media and activated with 20 µg/mL of the mouse STING ligand DMXAA (InvivoGen). The following day, BM-DCs were plated in 96-well plates at 10,000 cells per well in fresh media (without DMXAA) and cultured with lysed organoids overnight. Organoids were first dissociated to single cells in TrypLE and counted, then lysed in water at 2×10^6 cells per mL for 10 minutes at 37 °C. Lysed cells were pelleted at 1000 G for 15 minutes and resuspended in BM-DC media at appropriate dilutions. The next day, loaded BM-DCs were washed and cultured with 50,000 naïve OT-1s per well. $CD8^+$ T cells were purified from spleen and LNs of OT-1 mice using the $CD8a^+$ T Cell Isolation Kit, mouse (Miltenyi Biotech) following manufacturer specifications. Cells were harvested at 72 hours for staining and flow cytometric analysis. For cytokine stains, cells were treated with 1:1000 GolgiPlug and 2 µM Monensin Solution for 3 hours prior to harvest.

OT-1 T cell activation and organoid co-culture

Spleen and LNs from OT-1 mice were harvested in PBS, red blood cells lysed, cells resuspended in T cell media (described above) + 10 ng/mL hIL-2 (PeproTech) and 1 μM SIINFEKL peptide (Anaspec), counted, and plated at 1×10^6 cells per mL. Stimulation was performed for 24 hours at 37 °C. $CD8^+$ T cells were then purified using the $CD8a^+$ T Cell Isolation Kit, mouse (Miltenyi Biotech), and expanded in T cell media + hIL-2 with daily splitting. T cells were used for ACT or co-culture assays at day 3 or 4.

Organoids and OT-1s were plated at 0:1 and 5:1 effector to target ratios at 2500 organoid single cells in 10 μl minimal media and 50 μl Matrigel. Co-cultures were plated in triplicate at 20 μl per dome, grown in minimal media, and imaged on day 4. Total fluorescent area of organoids within images was quantified in ImageJ v2.1.0/1.53c by setting left and right thresholds of greyscale images to 22 and 255 on B&W setting, respectively, and analyzing particles with size threshold set to >20 pixel^2 and circularity set to 0.1–1.0 with holes included.

Immunohistochemistry and automated quantification

Tissues were fixed in zinc formalin, washed in 70% ethanol and paraffin embedded. Antigen retrieval was performed in citrate buffer pH 6 in a pressure cooker at 125 °C for five minutes. Blocking was performed with BLOXALL Endogenous Peroxidase and Alkaline Phosphatase Blocking Solution (Vector) followed by Normal Horse Serum (2.5%) (Vector). Slides were stained with CD8α (ab217344, Abcam) 1:1000 overnight, incubated with Alkaline Phosphatase (AP) anti-Rabbit IgG (Vector) and developed with Vector Black substrate (Vector). Sections then underwent a second round of antigen retrieval in a pressure cooker at 110 °C for two minutes, followed by co-incubation with FOXP3 (FJK-16s, eBioscience) 1:125 and CD4 (ab183685, Abcam) 1:400 overnight. Sections were then sequentially incubated with AP anti-Rat IgG (Vector) and HRP anti-Rabbit IgG (Vector) and developed sequentially with Vector Red

(Vector) and Vina Green (Biocare Medical). Slides were counterstained with Harris Acidified Hematoxylin and dehydrated. Aqueous wash steps following counterstain were shortened from 1 minute to 30 seconds to minimize loss of Vina Green stain.

Immune infiltration was calculated by a convolutional neural network (CNN) trained to identify the three cell types stained (black = CD8, green = CD4, green/red = Treg), using Aiforia's cloud-based platform (Aiforia Technologies Oy). Whole slides were scanned with a Leica AT2 (Aperio) using the Rainbow color profile. First, the CNN was trained to identify a tissue layer. Within that layer, the CNN was trained to identify black, green, and green/red staining. Within each of these layers, an object counter was trained to quantify the number of cells with the stain. Training was performed by manual annotation of each layer and counting of objects within training regions across 20 separate slides, with roughly five training regions per layer per slide. Performance was validated against human counting and found to be highly accurate and consistent.

In situ SIINFEKL tetramer staining

Tissue was stained *in situ* with SIINFEKL tetramer as previously described[80]. Tissues were additionally stained with CD8β AF647 (YTS156.7.7, BioLegend) (1:100), and anti-human β2-microglobulin (β2M) PE (2M2, BioLegend) (1:50). Anti-β2M staining is specific to human β2M in the SIINFEKL tetramer and serves to amplify signal. Images were taken at 30X on an Olympus FV1200 Laser Scanning Confocal Microscope and analyzed in ImageJ v2.1.0/1.53c.

In vivo antibody and vaccine dosing

All antibody dosing was performed via intraperitoneal injection in PBS. αCD4 (GK1.5, BioXCell) and αCD8 (2.43, BioXCell) depleting antibodies were administered at 200 ug every 4 days. αPD-1 (29F.1A12, BioXCell) was administered at 200 μg three times a week. αCTLA

(9H10, BioXCell) was administered at an initial dose of 200 μg, with all subsequent doses at 100 μg, three times a week. αCD40 (FGK4.5, BioXCell) was administered once at the beginning of treatment at 100 μg.

The adjuvant amphiphile-CpG (amph-CpG) and antigen amphiphile (amph-peptide) were produced as previously described[53]. Briefly, class B CpG 1826 oligonucleotide with a G_2 spacer (5'-diacyl lipid-GGTCCATGACGTTCCTGACGTT- 3') was conjugated via the 5' end to an 18 carbon diacyl tail. Antigen peptide $OVA_{250–270}$ (CGLEQLESIINFEKLTEWTSS) and non-specific mutant $gp100_{20–39}$ (optimized S27P, EGP long[52], CAVGALEGPRNQDWLGVPRQL) were conjugated via N' cysteine residue to 1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine-*N*-[maleimide(polyethyleneglycol-2000] (Avanti Polar Lipids). Mice were vaccinated subcutaneously at the base of the tail with 1.24 nmol amph-CpG and 25 μg of amph-peptide, with half dose given to each side. Vaccination was performed once weekly starting 14 days post-transplant of lo^{SIIN} organoids.

Colonoscopy imaging

Tumor progression was monitored longitudinally using a Karl Storz colonoscopy system with white light, RFP and GFP fluorescence. This consists of Image 1 H3-Z Spies HD Camera System (part TH100), Image 1 HUB CCU (parts TC200, TC300), 175-Watt D-Light Cold Light Source (part 20133701–1), AIDA HD capture system, and fluorescent filters in the RFP and GFP channels (all from Karl Storz). The endoscope used for imaging was the Hopkins Telescope (Karl Storz, part 64301AA) with operating sheath (Karl Storz, part 64301AA). To consistently measure tumor area, biopsy forceps (Richard Wolf) were fed through the operating sheath and positioned consistently given two landmarks: widthwise grooves that appear as concentric semi-circles in the field of view, and a lengthwise groove at the forceps tip. Images were captured upon gentle contact of forceps with tumor. Tumor area in the field of view and length of the lengthwise forceps groove were calculated using ImageJ v2.1.0/1.53c. Tumor area was normalized to groove length.

TCR sequencing

Using a BD FACSAria flow cytometer, live SIINFEKL tetramer-positive CD8$^+$ T cells were directly sorted into 50 µl lysis buffer with proteinase K, from the Arcturus PicoPure DNA Extraction kit (ThermoFisher), in low binding microcentrifuge tubes (Biotix), and genomic DNA extraction performed following manufacturer instructions. Mouse TCRβ sequencing was performed by Adaptive Biotechnologies. Analysis was performed in R v4.0.2, and Simpson diversity calculated using the 'Vegan' v2.5.7 package. To account for differences in total numbers of T cells surveyed in samples between groups, unique productive TCR sequences were randomly downsampled to match between groups. Down-sampled data is presented in Extended Data Fig. 3k–l, although down-sampling did not impact observed trends.

Statistics and reproducibility

Statistical analyses and figure generation were performed in R v4.0.2 using built in functions and ggplot2 v3.3.3, beeswarm v0.3.1, corrplot v0.88 and RColorBrewer v1.1.2. For statistical assessment of differences in proportionality, Fisher's exact 2×2 test was performed. For continuous data, two-tailed Wilcoxon Rank Sum test was performed, apart from the organoid and OT-1 co-culture results, which were analyzed with two-tailed Student's t-test. Multiple comparison corrections were performed using Holm's method. No statistical method was used to predetermine sample size. Of animals transplanted with lo$^{\mathrm{SIIN}}$ organoids, only those that formed tumors were taken for flow cytometric analysis. No other data were excluded from analyses. Preclinical trials were randomized, and investigators blinded to allocation during dosing, colonoscopy imaging, and tumor quantification. Preclinical studies were performed across three independent cohorts with the aim of validating consistency and reaching 10 or more animals per treatment arm. All *in vivo* and co-culture experiments were repeated at least two times. No experiments presented in this manuscript failed to replicate.

References

1. Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
2. Le, D. T. *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409–413 (2017).
3. Germano, G. *et al.* Inactivation of DNA repair triggers neoantigen generation and impairs tumour growth. *Nature* **552**, 116–120 (2017).
4. Mandal, R. *et al.* Genetic diversity of tumors with mismatch repair deficiency influences anti-PD-1 immunotherapy response. *Science* **364**, 485–491 (2019).
5. Samstein, R. M. *et al.* Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **51**, 202–206 (2019).
6. Newey, A. *et al.* Immunopeptidomics of colorectal cancer organoids reveals a sparse HLA class i neoantigen landscape and no increase in neoantigens with interferon or MEK-inhibitor treatment. *J. Immunother. Cancer* **7**, 309 (2019).
7. Steimle, A. & Frick, J.-S. Molecular Mechanisms of Induction of Tolerant and Tolerogenic Intestinal Dendritic Cells in Mice. *J. Immunol. Res.* **2016**, 1958650 (2016).
8. Kim, M. *et al.* Critical Role for the Microbiota in CX3CR1+ Intestinal Mononuclear Phagocyte Regulation of Intestinal T Cell Responses. *Immunity* **49**, 151-163.e5 (2018).
9. Spranger, S., Bao, R. & Gajewski, T. F. Melanoma-intrinsic β-catenin signalling prevents anti-tumour immunity. *Nature* **523**, 231–235 (2015).
10. Galarreta, M. R. de *et al.* β-catenin activation promotes immune escape and resistance to anti-PD-1 therapy in hepatocellular carcinoma. *Cancer Discov.* (2019) doi:10.1158/2159-8290.CD-19-0074.
11. Roper, J. *et al.* In vivo genome editing and organoid transplantation models of colorectal cancer and metastasis. *Nat. Biotechnol.* **35**, 569–576 (2017).
12. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
13. Supek, F. & Lehner, B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* **521**, 81–84 (2015).
14. Wherry, E. J., Puorro, K. A., Porgador, A. & Eisenlohr, L. C. The induction of virus-specific CTL as a function of increasing epitope expression: responses rise steadily until excessively high levels of epitope are attained. *J. Immunol.* **163**, 3735–45 (1999).
15. Spiotto, M. T. *et al.* Increasing tumor antigen expression overcomes &quot;ignorance&quot; to solid tumors via crosspresentation by bone marrow-derived stromal cells. *Immunity* **17**, 737–47 (2002).
16. Engels, B. *et al.* Relapse or eradication of cancer is predicted by peptide-major histocompatibility complex affinity. *Cancer Cell* **23**, 516–26 (2013).
17. Jhunjhunwala, S., Hammer, C. & Delamarre, L. Antigen presentation in cancer: insights into tumour immunogenicity and immune evasion. *Nat. Rev. Cancer* 1–15 (2021) doi:10.1038/s41568-021-00339-z.
18. Richard, A. C. *et al.* T cell cytolytic capacity is independent of initial stimulation strength. *Nat. Immunol.* **19**, 849–858 (2018).
19. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science (80-.).* **351**, 1463–1469 (2016).
20. Wolf, Y. *et al.* UVB-Induced Tumor Heterogeneity Diminishes Immune Response in Melanoma. *Cell* **179**, 219-235.e21 (2019).

21. Karandikar, S. H. *et al.* Identification of epitopes in ovalbumin that provide insights for cancer neoepitopes. *JCI Insight* **4**, (2019).
22. DuPage, M. *et al.* Endogenous T cell responses to antigens expressed in lung adenocarcinomas delay malignant tumor progression. *Cancer Cell* **19**, 72–85 (2011).
23. DuPage, M., Mazumdar, C., Schmidt, L. M., Cheung, A. F. & Jacks, T. Expression of tumour-specific antigens underlies cancer immunoediting. *Nature* **482**, 405–9 (2012).
24. Dow, L. E. *et al.* Apc Restoration Promotes Cellular Differentiation and Reestablishes Crypt Homeostasis in Colorectal Cancer. *Cell* **161**, 1539–1552 (2015).
25. Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
26. Drost, J. *et al.* Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43–47 (2015).
27. de Sousa e Melo, F. *et al.* A distinct role for Lgr5+ stem cells in primary and metastatic colon cancer. *Nature* **543**, 676–680 (2017).
28. Schell, M. J. *et al.* A multigene mutation classification of 468 colorectal cancers reveals a prognostic role for APC. *Nat. Commun.* **7**, 11743 (2016).
29. Gubin, M. M. *et al.* Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577–81 (2014).
30. Mlecnik, B. *et al.* Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. *Immunity* **44**, 698–711 (2016).
31. Parish, I. A. *et al.* The molecular signature of CD8+ T cells undergoing deletional tolerance. *Blood* **113**, 4575–85 (2009).
32. Lin, W.-H. W. *et al.* CD8+ T Lymphocyte Self-Renewal during Effector Cell Determination. *Cell Rep.* **17**, 1773–1782 (2016).
33. Danilo, M., Chennupati, V., Silva, J. G., Siegert, S. & Held, W. Suppression of Tcf1 by Inflammatory Cytokines Facilitates Effector CD8 T Cell Differentiation. *Cell Rep.* **22**, 2107–2117 (2018).
34. Zhou, X. *et al.* Differentiation and persistence of memory CD8(+) T cells depend on T cell factor 1. *Immunity* **33**, 229–40 (2010).
35. Durward, M., Harms, J. & Splitter, G. Antigen specific in vivo killing assay using CFSE labeled target cells. *J. Vis. Exp.* 45 (2010) doi:10.3791/2250.
36. Wherry, E. J. T cell exhaustion. *Nat. Immunol.* **12**, 492–499 (2011).
37. McLane, L. M., Abdel-Hakeem, M. S. & Wherry, E. J. CD8 T Cell Exhaustion During Chronic Viral Infection and Cancer. *Annu. Rev. Immunol.* **37**, 457–495 (2019).
38. Miller, B. C. *et al.* Subsets of exhausted CD8+ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 326–336 (2019).
39. Siddiqui, I. *et al.* Intratumoral Tcf1+PD-1+CD8+ T Cells with Stem-like Properties Promote Tumor Control in Response to Vaccination and Checkpoint Blockade Immunotherapy. *Immunity* **50**, 195-211.e10 (2019).
40. Kurtulus, S. *et al.* Checkpoint Blockade Immunotherapy Induces Dynamic Changes in PD-1−CD8+ Tumor-Infiltrating T Cells. *Immunity* **50**, 181-194.e6 (2019).
41. Provine, N. M. *et al.* Immediate Dysfunction of Vaccine-Elicited CD8+ T Cells Primed in the Absence of CD4+ T Cells. *J. Immunol.* **197**, 1809–22 (2016).
42. Cella, M. *et al.* Ligation of CD40 on dendritic cells triggers production of high levels of interleukin-12 and enhances T cell stimulatory capacity: T-T help via APC activation. *J.*

Exp. Med. **184**, 747–752 (1996).

43. Byrne, K. T. & Vonderheide, R. H. CD40 Stimulation Obviates Innate Sensors and Drives T Cell Immunity in Cancer. *Cell Rep.* **15**, 2719–2732 (2016).
44. O'Hara, M. H. *et al.* Abstract CT004: A Phase Ib study of CD40 agonistic monoclonal antibody APX005M together with gemcitabine (Gem) and nab-paclitaxel (NP) with or without nivolumab (Nivo) in untreated metastatic ductal pancreatic adenocarcinoma (PDAC) patients. in *Clinical Trials* vol. 79 CT004–CT004 (American Association for Cancer Research, 2019).
45. Kabacaoglu, D., Ciecielski, K. J., Ruess, D. A. & Algül, H. Immune Checkpoint Inhibition for Pancreatic Ductal Adenocarcinoma: Current Limitations and Future Options. *Front. Immunol.* **9**, 1878 (2018).
46. Schietinger, A. *et al.* Tumor-Specific T Cell Dysfunction Is a Dynamic Antigen-Driven Differentiation Program Initiated Early during Tumorigenesis. *Immunity* **45**, 389–401 (2016).
47. Wells, D. K. *et al.* Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction. *Cell* **183**, 818-834.e13 (2020).
48. Ott, P. A. *et al.* A Phase Ib Trial of Personalized Neoantigen Therapy Plus Anti-PD-1 in Patients with Advanced Melanoma, Non-small Cell Lung Cancer, or Bladder Cancer. *Cell* **183**, 347-362.e24 (2020).
49. Simoni, Y. *et al.* Bystander CD8+ T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* **557**, 575–579 (2018).
50. Krogsgaard, M. & Davis, M. M. How T cells 'see' antigen. *Nat. Immunol.* **6**, (2005).