CRISPR-mediated interrogation of small cell lung cancer

by

Sheng Rong Ng

B.A., Natural Sciences
University of Cambridge, 2011

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

Signature of Author ....
Department of Biology
May 25, 2018

Certified by ....…........
Tyler Jacks
Professor of Biology
Thesis Supervisor

Accepted by ....
Amy Keating
Professor of Biology
Chair, Biology Graduate Committee

CRISPR-mediated interrogation of small cell lung cancer

by

Sheng Rong Ng

Submitted to the Department of Biology on May 25, 2018 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biology

ABSTRACT

Small cell lung cancer (SCLC) is a highly aggressive neuroendocrine lung carcinoma that remains among the most lethal of solid tumor malignancies. Despite decades of research, treatment outcomes for SCLC remain very poor, highlighting the need for novel approaches to target the disease. Recent genomic sequencing studies have identified multiple recurrently altered genes in human SCLC tumors, many of which remain to be functionally validated. Genetically engineered mouse models (GEMMs) of SCLC have been developed that recapitulate many key features of human SCLC. These models have been used extensively to investigate various aspects of SCLC biology, including tumor initiation, progression and metastasis.

The development of the CRISPR-Cas9 system has greatly facilitated genome editing in mammalian cells, leading to its widespread adoption for various applications in cancer biology. We have utilized this system in two complementary ways to investigate the molecular mechanisms involved in SCLC initiation, progression and maintenance. Firstly, we have adapted the CRISPR-Cas9 system for use in GEMMs of SCLC, to enable rapid modeling and functional validation of candidate tumor suppressor genes *in vivo*. Using this system, we have demonstrated that p107, a member of the retinoblastoma family that is mutated in a significant fraction of human SCLC tumors, is a functional tumor suppressor in SCLC. Notably, loss of p107 in SCLC tumors resulted in significant phenotypic differences compared with loss of its close relative, p130. We also demonstrated that CRISPR-induced mutations can be used to infer lineage relationships between primary and metastatic tumors in the same animal.

Secondly, we have performed a CRISPR-based genetic screen, utilizing a custom sgRNA library targeting the druggable genome, to identify novel SCLC-specific genetic vulnerabilities. We found that SCLC cells displayed enhanced sensitivity towards disruption of several key metabolic pathways, including the *de novo* pyrimidine biosynthesis pathway. Pharmacological inhibition of Dhodh, a key enzyme in this pathway, reduced the viability of SCLC cells *in vitro* and strongly suppressed SCLC tumor growth *in vivo*, validating this pathway as a promising therapeutic target in SCLC.

Taken together, the work presented here demonstrates the utility of the CRISPR-Cas9 system for performing functional interrogation of SCLC.

Thesis Advisor: Tyler Jacks
Title: Professor of Biology

CURRICULUM VITAE
SHENG RONG NG

Address: Koch Institute for Integrative Cancer Research
Massachusetts Institute of Technology
77 Massachusetts Avenue, 76-453
Cambridge, MA 02139
Email address: srng@mit.edu

EDUCATION

2012 – present   PhD Candidate in Biology
Massachusetts Institute of Technology, USA

2008 – 2011   Bachelor of Arts (First Class Honours) in Natural Sciences (Biochemistry)
University of Cambridge, United Kingdom

RESEARCH EXPERIENCE

2013 – present   PhD Candidate
Advisor: Prof. Tyler Jacks, Koch Institute for Integrative Cancer Research,
Massachusetts Institute of Technology
Investigating tumor progression and metastasis in small cell lung cancer.

2011 – 2012   Research Officer
Advisor: Dr. Leah Vardy, Institute of Medical Biology, A*STAR, Singapore
*Investigated the role of ribosomal proteins in mouse embryonic stem cell
pluripotency and differentiation.*

2011   Undergraduate Research Student
Advisor: Prof. Austin Smith, Centre for Stem Cell Research, University of
Cambridge, UK
*Investigated the role of a transcription factor in mouse embryonic stem cell
commitment to differentiation.*

2010   Summer Research Student
Advisor: Dr. Rick Livesey, Gurdon Institute, University of Cambridge, UK
*Studied the effects of microRNAs on the temporal order of neurogenesis by
neocortical stem cells in mice.*

2009   Research Intern
Advisor: Dr. Lian-Hui Zhang, Institute of Molecular and Cell Biology, A*STAR,
Singapore
*Contributed to the characterization of a protein involved in regulation of
quorum sensing in Pseudomonas aeruginosa.*

2008   Research Intern
Advisor: Dr. Gerald Udolph, Institute of Medical Biology, A*STAR, Singapore
*Contributed to the study of fetal cell migration from fetal to maternal tissues
during pregnancy.*

TEACHING EXPERIENCE

Fall 2015   Teaching Assistant, Introductory Biology (7.012), MIT
Spring 2014   Teaching Assistant, Genetics (7.03), MIT

AWARDS

2017   Koch Institute Marlena Felter Bradford Research Travel Fellowship
2017   MIT School of Science Fellowship in Cancer Research
2012   A*STAR National Science Scholarship (PhD)
2011   A*STAR Roll of Honor
2010, 2011   Simon Wilson Prize, Christ's College, University of Cambridge
2010   J.B. & Millicent Kaye Fund Travel Grant, Christ's College, University of
               Cambridge
2009, 2010   S.W. Greig Prize, Christ's College, University of Cambridge
2009   Valerie Barker Prize, Christ's College, University of Cambridge
2006   A*STAR National Science Scholarship (BS)

PUBLICATIONS

2015   Li CM, Gocheva V, Oudin MJ, Bhutkar A, Wang SY, Date SR, **Ng SR**,
               Whittaker CA, Bronson RT, Snyder EL, Gertler FB, and Jacks T. (2015) Foxa2
               and Cdx2 cooperate with Nkx2-1 to inhibit lung adenocarcinoma metastasis.
               Genes Dev 29 (17), 1850-1862.

2014   Wong QWL, Li J, **Ng SR**, Lim SG, Yang H, and Vardy LA. (2014) RPL39L is
               an example of a recently evolved ribosomal protein paralog that shows highly
               specific tissue expression patterns and is upregulated in ESCs and HCC
               tumors. *RNA Biology* 11:1, 33-41.

PRESENTATIONS

2017   **Ng SR**, Rideout WM 3rd, Wagner BL, and Jacks T. (2017) CRISPR-mediated
               modeling and functional validation of candidate tumor suppressor genes in
               small cell lung cancer. *Frontiers in Cancer Science 2017*. Singapore.
               Selected for poster and oral presentations.

2017   **Ng SR**, Rideout WM 3rd, Wagner BL, and Jacks T. (2017) CRISPR-mediated
               modeling and functional validation of candidate tumor suppressor genes in
               small cell lung cancer. *AACR Special Conference: Advances in Modeling
               Cancer in Mice: Technology, Biology, and Beyond*. Orlando, Florida. **Selected
               for poster presentation.**

2016   **Ng SR**, Rideout WM 3rd, Wagner BL, Yu JK, and Jacks T. (2016)
               CRISPR-mediated interrogation of recurrent genomic alterations in small cell
               lung cancer. *81st Cold Spring Harbor Symposium on Quantitative Biology:
               Targeting Cancer*. Cold Spring Harbor, New York. **Selected for poster
               presentation.**

ACKNOWLEDGEMENTS

unrelated to my own, never failed to offer me useful advice for my work whenever I approached them, as well as after my group meeting presentations. I want to thank all of these people for welcoming me into the lab and helping me to establish my projects.

Among the people who have joined more recently, Britt, Leanne and Caterina, my current SCLC teammates, have contributed to all aspects of my projects. Rodrigo, Santiago and Will, the leading CRISPR experts in the lab, have also contributed significantly to the development of many of my project ideas. It has been very helpful to have people in the lab with similar areas of interest to bounce ideas back and forth with, as well as to plug holes in my knowledge of the literature.

Amy, the current senior graduate student in the lab, started in the lab at around the same time as I did, and has been a good friend throughout this whole time. It has been great to have her in the lab as a fellow graduate student going through the same trials and tribulations together, even if we have been working on largely unrelated projects. It is only appropriate, then, that we are both defending our theses at the same time – certainly a reason for double celebration.

Santiago, Carla, Caterina – these are my daily lunch buddies as well as my closest friends in the lab. At various times, they have provided me not just with scientific and technical help, but also with much-needed emotional support and a listening ear. Their friendship and support have been instrumental in keeping me going and thriving through various stresses and problems, both in and out of lab. It would be difficult to overstate how much this has meant to me.

I have had the chance to work with many collaborators outside of our lab, especially from the Manalis Lab – Bashar, Emily, Chris, Lucy and Kelsey, among others. In particular, Bashar has really been a wonderful collaborator, driving the CTC project through many highs and lows, and it has been a pleasure to work with him, as well as to get to know him as a close friend.

Outside of the lab, many Singaporean friends, both old and new, have provided me with a welcome dose of home from time to time, whether it is during catch-up sessions over Singaporean/Malaysian food at Royal East, or at the various activities organized by the MIT Singapore Students Society. Many of us share similar yet distinct experiences as graduate students in various departments, and it is refreshing to be able to discuss non-work-related subjects from time to time.

Last but definitely not least, I want to thank my family – my father, mother and brother – as well as my girlfriend, Shue Li, for their constant support from home throughout this long sojourn of mine. My brother, who was also a graduate student at MIT, really paved the way for me to pursue a career in research, and also helped me to adapt quickly to life in Boston. My parents have always been supportive of my work and, among other things, often send over local snacks and delicacies to keep me going. My girlfriend has been a constant source of strength for me, and I am forever thankful that she has chosen to stick with me throughout this whole time that we have been apart. I look forward to returning home soon to resume our journey together.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Part I: Functional interrogation of cancer

1. The genetic basis of cancer

1.1 Cancer as a genetic disease

The discovery of the first oncogene some forty years ago laid the foundations for much of modern cancer biology research today. Prior to this, there were already numerous factors associated with the development of various cancers that hinted at a genetic basis for cancer. For example, certain occupations were long known to be associated with elevated incidences of specific cancers, such as chimney sweeps and scrotal cancer, and dye-manufacturing workers and bladder cancer, although the actual causes were not known at that time. Yamagiwa and Ichikawa first demonstrated that chemicals could act as carcinogens by experimentally inducing metastatic tumors in rabbits through the application of coal car to their ears (Yamagiwa and Ichikawa, 1918). Many years later, Ames and colleagues established that many well-known carcinogens, including aflatoxin, polycyclic hydrocarbons such as benzo[*a*]pyrene (found in chimney soot and coal tar), and aromatic amines such as naphthylamine (found in dyes), functioned as mutagens upon activation by mammalian liver homogenate, and therefore were likely to cause cancer by inducing somatic mutations in the genome (Ames et al., 1973b, 1973a).

In apparently unrelated studies, many viruses had been discovered to induce tumors in animals, with the most well-known example being the Rous sarcoma virus in chickens (Rous, 1911). This gave rise to the idea that perhaps cancer was an infectious disease, leading many to begin searching for potential infectious agents in human tumors, in order to establish whether this was relevant to human cancers. Eventually, in

a landmark finding that transformed the field, Bishop, Varmus and colleagues demonstrated that DNA related to the segment of the Rous sarcoma virus genome responsible for its ability to transform cells, the *v-src* oncogene, could be detected in the genomes of normal uninfected avian cells (Stehelin et al., 1976), as well as cells from more distantly related vertebrates (Spector et al., 1978). This showed that retroviral oncogenes were, in fact, modified counterparts of normal cellular genes, or proto-oncogenes, that are likely to have essential cellular functions.

Discovery of cellular oncogenes

Subsequently, multiple groups demonstrated that retroviruses were completely dispensable for the transformation process. Segments of human DNA isolated from tumor cell lines, when transfected into untransformed mouse fibroblast cells, were sufficient to cause their transformation *in vitro* (Goldfarb et al., 1982; Perucho et al., 1981; Pulciani et al., 1982; Shih and Weinberg, 1982; Shih et al., 1981). These segments were later found to correspond to the previously discovered viral *ras* gene (Der et al., 1982; Parada et al., 1982; Santos et al., 1982), thereby validating the relevance of earlier studies of retroviral oncogenes. Furthermore, the version of the *ras* gene isolated from human tumors differed from the normal version by just a single point mutation that modified a single amino acid, demonstrating conclusively that mutations in endogenous cellular genes can transform a cell (Reddy et al., 1982; Tabin et al., 1982; Taparowsky et al., 1982). Collectively, these studies and others showed that cancer was a result of alterations in normal genes present in the genome, generating defective versions of these genes that confer the ability to transform normal cells into neoplastic cells.

Tumor suppressor genes

Despite these groundbreaking findings, there remained observations that could not be explained by the concept of gain-of-function oncogenic mutations in normal cellular genes. In a series of experiments conducted by Harris and colleagues, it was observed that the fusion of a malignant cell with a non-malignant counterpart resulted in daughter cells that were non-malignant (Harris, 1971), which suggested that malignancy was a recessive phenotype, in contrast to the apparent dominant nature of mutated oncogenes. In addition, using elegant statistical analyses of patients who developed unilateral versus bilateral retinoblastoma tumors, Knudson had hypothesized that retinoblastoma formation requires two mutational events (the two-hit hypothesis), with patients who developed bilateral retinoblastomas having already inherited the first mutation (Knudson, 1971). This hypothesis was subsequently validated by experiments demonstrating that patients with hereditary predisposition to retinoblastoma harbored loss-of-function mutations in a specific genomic locus, and that retinoblastoma tumors that developed in these patients frequently lost the second wild-type copy of that locus (Cavenee et al., 1983). The gene at this locus was eventually isolated (Friend et al., 1986) and aptly named the retinoblastoma gene, or *RB1*, proving to be the first example of a tumor suppressor gene. Tumor-predisposing mutations in this second class of cancer-associated genes confer a recessive phenotype, rather than a dominant phenotype observed with mutations in proto-oncogenes.

Ever since the discoveries of the first human oncogene and the first tumor suppressor gene, numerous other examples of both classes of genes have been discovered in various cancers. The advent of whole-genome sequencing technologies,

which made possible the cataloging of all the genes in the human genome as well as the detection of the vast repertoire of mutations in human tumors, has significantly accelerated this endeavor.

1.2   Cancer in the genomics era

One of the major research efforts in the 1990s and early 2000s was the Human Genome Project, which culminated in the unveiling of the sequence of the entire human genome (Collins et al., 2004; Lander et al., 2001; Venter et al., 2001). For the first time, it was possible to begin annotating the complete catalog of human genes, as well as to study their roles in various diseases in a comprehensive fashion. Combined with the subsequent development of next-generation sequencing technologies, which vastly increased the speed of sequencing while hugely decreasing costs, this rapidly ushered in the genomics era in cancer research.

It has long been recognized that cancer is a multistep process, requiring the accumulation of multiple genetic alterations, a paradigm that is best illustrated in colorectal cancer (Vogelstein and Kinzler, 1993). However, recent sequencing studies have shown that human tumors harbor an average of 30-60 non-silent coding mutations each (Lawrence et al., 2013; Vogelstein et al., 2013), which is far higher than the 4-7 genetic events that are thought to be required for tumor progression (Kinzler and Vogelstein, 1996; Renan, 1993). There is considerable heterogeneity in mutational frequencies across different cancer types, with lower frequencies in pediatric cancers such as rhabdoid tumors and Ewing sarcoma, and much higher frequencies in cancers commonly associated with carcinogens, such as melanoma (UV irradiation) and lung

cancer (cigarette smoke) (Lawrence et al., 2013; Vogelstein et al., 2013). To complicate

matters further, there often exists a large range of mutational frequencies even within

each cancer type (Lawrence et al., 2013).

With such high mutation rates in many tumors, it becomes a significant challenge

to distinguish between driver mutations, which are mutations in genes that confer a

selective growth advantage to the cancer cell, and passenger mutations, which are

mutations that have no significant effect on cancer cell growth. This is especially

challenging in cancers with high mutation rates such as melanoma and lung cancer, as

well as subsets of cancers that harbor mutations in key DNA repair genes. For example,

a subset of colorectal tumors harbor mutations in genes involved in DNA mismatch

repair, such as *MSH2* and *MLH1*, which result in tumors that contain thousands of non-

synonymous mutations each (Vogelstein et al., 2013). It is not technically feasible to test

each mutation individually to determine which of these mutations are important for

tumor progression.

1.3   Comprehensive molecular profiling of cancer

Large-scale whole-exome and whole-genome cancer sequencing studies aim to

overcome this problem, based on the idea that functionally important driver events are

likely to occur in a significant fraction of independent tumors, while passenger mutations

that result from random mutational processes should occur at much lower frequencies.

Such studies have grown to include national and international collaborations, such as

The Cancer Genome Atlas (TCGA), which is itself part of the International Cancer

Genome Consortium. To date, the TCGA project has sequenced over 11,000 human

tumor samples across 33 tumor types, with many of these samples also having matched gene expression profiles, as well as clinical data such as overall survival and progression-free interval. This has provided a rich dataset for identifying key cancer driver genes and pathways in various cancer types.

To illustrate the utility of this approach, comprehensive molecular profiling of lung adenocarcinoma revealed that the majority of these tumors harbored known driver mutations within the RTK/RAS/RAF pathway, such as *EGFR*, *KRAS* and *BRAF*, as well as previously identified fusions involving *ALK*, *ROS1* and *RET* (Cancer Genome Atlas Research Network, 2014). However, among the remaining RTK/RAS/RAF pathway-negative tumors, unique focal amplifications in *ERBB2* and *MET*, as well as mutations in *NF1* and *RIT1*, were found to be significantly enriched. In addition to providing potential new targets for the development of novel therapies, the discovery of these genetic events in these tumors opens up the possibility of using existing therapies, such as MET and ERBB2 inhibitors, in lung adenocarcinoma patients.

The success of comprehensive cancer profiling studies has, ironically, resulted in a different problem – how to prioritize candidate genes for subsequent functional validation. Even with rigorous statistical analyses to account for variable mutation rates and heterogeneity in mutational processes in different cancers, there remains a long list of genes that are altered at significant frequencies (Lawrence et al., 2013). Efforts have begun to analyze sequencing data from across different cancer types to identify "pan-cancer" genetic drivers (Bailey et al., 2018), which may aid in the prioritization of important genes. Future efforts will likely require the development of better tools to functionally characterize the consequences of specific genetic alterations, as well as

improved tools to predict the effects of gain or loss of gene function at the global level in tumor cells.

1.4   Genetic screens and functional genomics

An alternative approach to identify novel cancer-specific therapeutic targets involves the use of genetic screens. This approach employs genetic tools that are used to perturb gene function in a large-scale, unbiased fashion, followed by functional assays to assess the resulting phenotypes of the perturbation. Depending on the design of the experiment, phenotypes can include changes in rates of cell proliferation, activation of reporter genes, secretion of protein products, and many more. Such an approach has been termed functional genomics, which underscores the emphasis on characterizing gene function.

Reverse genetics to identify gene function

The concept of perturbing the function of a gene to assess the resulting phenotypes is known as reverse genetics. This approach has long been used in the field of yeast genetics, where a systematic effort to generate yeast deletion strains comprising the entire genome was undertaken, culminating in the *Saccharomyces* Genome Deletion Project (Giaever et al., 2002; Winzeler et al., 1999). The motivation behind this effort was the sequencing of the complete genome of the yeast *Saccharomyces cerevisiae*, which revealed that a significant proportion of the roughly 6,000 open reading frames present in the yeast genome had not been previously studied, and thus were of unknown function (Dujon, 1996). The collection of yeast knockout strains has been used for many different functional studies, including

identification of factors involved in cell growth, mating, sporulation and germination, as well as response to environmental stresses (reviewed in Giaever and Nislow, 2014).

Similar efforts have been carried out to study gene functions in mammalian systems, such as mice. The International Mouse Phenotyping Consortium (IMPC), which builds upon programs such as the Knockout Mouse Project (KOMP), aims to generate knockout mutations for every gene in the mouse genome, so as to discover and ascribe biological functions to each gene (Brown and Moore, 2012). However, it is much more time-consuming and expensive to generate knockout mice than to generate knockout yeast strains. Furthermore, it is simply not feasible to characterize every single knockout mouse strain each time one wishes to interrogate a specific biological question. Therefore, the mainstay of functional genomics in mammalian systems has been genetic screens in cell-based systems, where genetic perturbations can be performed in a large-scale, highly parallel fashion.

Genetic screens in mammalian cells

Performing genetic screens in mammalian cells is complicated by the diploid nature of most mammalian cell lines. As most loss-of-function mutations generate recessive phenotypes, this means that both alleles of a particular gene need to be mutated or deleted before any phenotype becomes apparent. This limits the use of insertional mutagenesis methods, such as retroviruses and transposons, to certain cell lines that are mostly haploid, such as the human chronic myeloid leukemia cell line KBM-7 (Carette et al., 2009; Kotecki et al., 1999).

The development of RNA interference (RNAi)-based methods transformed the way loss-of-function genetic screens could be performed. RNAi, also known as post-

transcriptional gene silencing (PTGS), was first observed in plants and subsequently in animals such as the nematode worm *Caenorhabditis elegans*. Fire and colleagues first determined that RNAi was mediated by double-stranded RNA in cells (Fire et al., 1998). Subsequent work showed that these double-stranded RNAs were processed into 21-nucleotide single-stranded RNAs that mediated gene silencing (Elbashir et al., 2001a; Hamilton and Baulcombe, 1999; Hammond et al., 2000; Zamore et al., 2000). Elbashir and colleagues then showed that 21-nucleotide RNA duplexes could suppress gene expression in a sequence-specific manner in mammalian cells (Elbashir et al., 2001b). At around the same time, multiple groups identified endogenous small RNAs, or microRNAs (miRNAs), that form hairpin structures in the cell and can regulate gene expression in a broad range of organisms (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). Collectively, these discoveries led to the development of short hairpin RNA (shRNA)-based RNAi systems, which can be designed for targeted gene repression (Brummelkamp et al., 2002; Paddison et al., 2002; Paul et al., 2002). Because RNAi represses gene expression post-transcriptionally, this bypasses the need to generate mutations in both copies of a gene in mammalian cells.

Large-scale RNAi-based libraries have been designed and cloned into retroviral and lentiviral vectors for performing genetic screens in mammalian systems, both *in vitro* (Berns et al., 2004; Moffat et al., 2006; Paddison et al., 2004; Silva et al., 2005) as well as *in vivo* (see, for example, Zender et al., 2008). However, there are several issues associated with the use of RNAi in genetic screens. These include incomplete knockdown of gene expression, as well as significant off-target effects (Kaelin, 2012). Many of these shortcomings have been overcome by the use of CRISPR-Cas-based

systems, which have begun to complement, and in some cases supersede, the use of RNAi-based systems for genetic screens. This will be discussed in greater detail later in this chapter.

Gain-of-function genetic screens have traditionally been far more challenging to perform compared with loss-of-function genetic screens. Unlike in the case of RNAi, where 21-nucleotide sequences can be designed to repress expression of essentially any gene in the genome, there has, until recently, been no equivalent method to activate gene expression. Large-scale lentiviral libraries expressing open reading frames (ORFs) have been developed for gain-of-function screens (Yang et al., 2011), but these do not offer the same level of flexibility as RNAi knockdown libraries. As will be discussed, CRISPR-Cas systems have been modified to enable transcriptional activation of genes in a highly flexible fashion, allowing such systems to be adapted for gain-of-function screens as well.

Identification of context-specific genetic vulnerabilities

The power of genetic screens is perhaps best illustrated in its use for identifying context-specific genetic vulnerabilities in cells. Loss-of-function screens have been performed in many different cell lines to identify genes that, when mutated, result in decreased cell proliferation or death (Aguirre et al., 2016; Hart et al., 2015, 2017; Wang et al., 2015). Such screens have identified genes that appear to be essential across all cell lines; as expected, these genes encode proteins that are involved in essential cellular processes such as translation, transcription and DNA replication, and have been termed "core essential genes". However, other genes appear to be essential only in

certain cell lines and do not result in any proliferative defects when lost in other cell lines. These are referred to as context-specific essential genes (Hart et al., 2014).

In the context of cancer, genetic screens can be used to identify potential therapeutic targets that are essential in cancer cells, but non-essential in normal, untransformed cells. This principle underlies the concept of synthetic lethality, in which mutation of a particular gene is lethal to a cell only when it occurs together with mutation of a second gene (Kaelin, 2005). For cancer cells, this could be mutations in oncogenes or tumor suppressor genes, which are not present in normal cells. Loss-of-function screens have been performed to identify such synthetic-lethal interactions. For example, Wang and colleagues performed genome-wide CRISPR-based genetic screens in a panel of acute myeloid leukemia (AML) cell lines, half of which harbored a mutant *RAS* allele, and found that loss of the guanine nucleotide exchange factor *PREX1* was selectively essential in *RAS*-mutant AML lines and not in *RAS*-wild-type AML lines (Wang et al., 2017). More broadly, both RNAi-based and CRISPR-based screens have been used to profile cancer type-specific vulnerabilities across large panels of hundreds of cancer cell lines (McDonald et al., 2017; Meyers et al., 2017; Tsherniak et al., 2017). In combination with cancer genome sequencing studies, such efforts have enabled the discovery and prioritization of genes for downstream functional analyses.

2.   Modeling cancer in mice

Many groundbreaking discoveries in cancer biology have been made using *in vitro* cell culture-based systems, as discussed in the previous section. *In vitro* systems offer many advantages for studying specific mechanisms of oncogene or tumor

suppressor gene function. For example, such systems provide a well-controlled environment where experimental variables can be minimized, allowing the results of a specific perturbation to be interpreted more easily. However, *in vitro* systems are unable to fully replicate certain key features of cancer. To illustrate this, of the six hallmarks of cancer described by Hanahan and Weinberg (Hanahan and Weinberg, 2000), two of these – sustained angiogenesis, and tissue invasion and metastasis – cannot be fully recapitulated in cell culture-based systems. Furthermore, the recent resurgence of cancer immunotherapy has ignited huge interest in understanding the interactions between tumor cells and the immune system, which can only be effectively studied in whole-organism systems.

Mouse models of cancer have long been utilized to bridge the gap between *in vitro* systems and humans. Many aspects of the laboratory mouse make it a suitable model system for studying cancer, including similarities with humans at the genomic and physiological levels. The use of mouse cancer models has been reviewed extensively elsewhere (see, for example, Frese and Tuveson, 2007). In this section, I will focus on genetically engineered mouse models (GEMMs) of cancer. In particular, I will highlight the key technological advances that have enabled the development of advanced GEMMs that allow precise spatial and temporal control of tumorigenesis.

2.1 Transgenic mouse models

The development of methods to generate transgenic mice was the first step towards creating mouse models of cancer. Jaenisch and Mintz first demonstrated the successful transmission of simian virus 40 (SV40) DNA in mice. They microinjected

SV40 DNA into preimplantation mouse blastocysts, allowed the blastocysts to develop into adults, then showed that SV40-specific DNA sequences could be detected in DNA extracted from various adult organs and tissues, providing confirmation that the injected SV40 DNA had been successfully transmitted throughout mouse development (Jaenisch and Mintz, 1974). In a subsequent study, Jaenisch demonstrated successful germline transmission of Moloney murine leukemia virus DNA in mice (Jaenisch, 1976), paving the way for generating heritable changes to the mouse genome *in vivo*.

Many transgenic mouse models of cancer were developed using this approach. For instance, Brinster, Palmiter and colleagues generated mice carrying SV40 large and small T antigens under the control of the SV40 enhancer, and found that these mice developed choroid plexus tumors in the brain (Brinster et al., 1984; Palmiter et al., 1985). Some level of tissue specificity could be conferred by generating hybrid constructs driven by tissue-specific regulatory elements. Mice carrying the *myc* gene under the control of the mouse mammary tumor virus (MMTV) promoter specifically develop mammary adenocarcinomas (Stewart et al., 1984). Likewise, expression of the SV40 large T antigen under the control of the rat insulin promoter in mice results specifically in pancreatic β-cell tumors (Hanahan, 1985).

Although these models allowed tumorigenesis and tumor progression to be studied *in vivo*, there were several problems with transgenic models. In many cases, the inserted gene is expressed at high levels that do not reflect physiological conditions, due to the use of ectopic promoters without endogenous regulatory elements. Furthermore, the random integration of the transgene into the genome can result in

22

variable gene expression. More importantly, transgenic models have limited utility for modeling loss-of-function mutations in tumor suppressor genes.

2.2   Targeted gene modifications in mice

The next generation of genetically engineered mouse models resulted from a confluence of two separate technological developments in the 1980s – the use of homologous recombination to generate targeted genetic modifications in cells, and the derivation of mouse embryonic stem (ES) cells.

Targeted gene modifications by homologous recombination

Homologous recombination is a process that involves the exchange of genetic material between two DNA strands with regions of sequence homology. In eukaryotic cells, this process occurs as part of normal meiosis, and is responsible for the phenomenon of gene conversion in yeast (Holliday, 1964; Meselson and Radding, 1975; Szostak et al., 1983).

The first demonstration that homologous recombination could be used to target an exogenous DNA sequence to a specific locus in the eukaryotic genome was performed in yeast (Hinnen et al., 1978). Hinnen and colleagues introduced a plasmid carrying the yeast *LEU2*+ gene into a *leu2*- yeast strain, and obtained colonies at low frequencies that could grow in medium lacking leucine. Further characterization revealed that some of the transformed colonies had replaced the original *leu2*- locus with the incoming *LEU2*+ sequence. Subsequently, Orr-Weaver and colleagues showed that recombination efficiency could be greatly enhanced by using linear rather than circular plasmids (Orr-Weaver et al., 1981).

Building on the work in yeast, the Smithies and Capecchi groups later reported the successful targeting of genes to specific loci in mammalian cells via homologous recombination, showing that this method was not limited to just yeast cells (Smithies et al., 1985; Thomas et al., 1986). It was now possible not just to replace an endogenous gene with an altered version of the gene to modify its function while retaining locus-specific regulation of the gene, but also to "knock out" an endogenous gene by inserting a sequence within the gene to disrupt its coding sequence.

Derivation of mouse embryonic stem cells

The relatively low efficiency of homologous recombination events compared with random integration events (Thomas et al., 1986) meant that it was not feasible to generate targeted gene modifications directly in blastocysts or zygotes, as had been done with transgenic approaches. Fortunately, Evans and Kaufman had developed a method to isolate and propagate murine embryonic stem (ES) cells from mouse preimplantation blastocysts (Evans and Kaufman, 1981). These cells exhibited pluripotency, or the ability to differentiate into cells from all three germ layers, and gave rise to teratocarcinomas when injected subcutaneously into syngeneic mice. Importantly, unlike the previously derived embryonal carcinoma cell lines that had similar differentiation capabilities (Martin and Evans, 1975; Papaioannou et al., 1975), ES cells were karyotypically normal and stable. Subsequently, ES cells were shown to contribute to chimera formation with high efficiency when injected into donor blastocysts, with ES cell-derived progeny contributing to all tissues in the mouse, including the germ line (Bradley et al., 1984).

Generation of mice with targeted gene modifications

The availability of ES cells meant that low-frequency events such as gene targeting by homologous recombination could be performed *in vitro*, followed by a selection step to isolate correctly targeted cells before blastocyst injections. It was not long before germ line transmission of targeted genetic alterations in mice was successfully achieved (Koller et al., 1989; Thompson et al., 1989). These initial studies involved the correction of a defect in the hypoxanthine guanine phosphoribosyl transferase gene (*HPRT*); successful targeting events could be selected for and enriched by culturing cells in hypoxanthine-aminopterin-thymidine medium, in which only *HPRT*-positive cells survive (Doetschman et al., 1987). Subsequently, Mansour and colleagues developed a generalizable method, known as positive-negative selection, to selectively enrich for cells that have undergone successful homologous recombination-mediated targeting over cells that have undergone random integration (Mansour et al., 1988). This is especially useful in cases where the targeted genes do not have an easily selectable phenotype. With the obvious utility of this approach for studying the effects of loss of tumor suppressor genes *in vivo,* it was no surprise that multiple groups quickly reported the generation of mice with targeted disruptions in key tumor suppressor genes, such as *Rb1* (Jacks et al., 1992; Lee et al., 1992), *p53* (Donehower et al., 1992; Jacks et al., 1994a) and *Nf1* (Jacks et al., 1994b).

Although this approach was useful in uncovering the roles of these tumor suppressor genes in various cancers, some of which were not previously appreciated, one pattern that quickly emerged was that homozygous loss of such genes *in vivo* often resulted in embryonic lethality. This reflected the fact that many these genes play crucial

roles in normal cellular function; however, it also posed a problem for modeling their roles in cancer progression. The use of heterozygous knockout mice partially overcame this problem, but this required sporadic loss-of-heterozygosity events to occur before tumors could develop, resulting in variable penetrance of tumor phenotypes.

Furthermore, in cases where loss of the gene, or a combination of genes, results in multiple tumor types forming in the mouse, these tumors often progress at different rates, causing the mice to succumb to one type of tumor before another type has had time to develop. For example, in mice with combined $Rb1^{+/-};p53^{-/-}$ mutations, approximately 40% of animals developed small hyperplastic foci of neuroendocrine cells in the bronchi and bronchioles of the lung. However, these hyperplastic foci did not develop further due to the animals succumbing to other tumors by two to six months of age (Williams et al., 1994). Given the frequent loss of $RB1$ and $TP53$ in small cell lung cancer (SCLC), which is a subtype of neuroendocrine lung tumors, it is likely that these lesions represented SCLC precursors that did not have sufficient time to develop into frank tumors. Thus, these problems highlighted the need for methods to control when and where genetic modifications occur in the mouse.

2.3   Spatial and temporal control of gene knockout/activation

Further advances in genetic tools led to the development of the next generation of mouse models of cancer. These tools enabled spatial and temporal control of gene activation or disruption, overcoming a number of the problems associated with constitutive gene activation or disruption in earlier transgenic or knockout models.

Site-specific recombinase-based systems

Site-specific recombinases, such as the Cre-*loxP* system derived from the

bacteriophage P1 and the Flp-*FRT* system derived from yeast, were one of the earliest

tools used to enable spatial and temporal control of genetic modifications in the mouse.

These systems involve the use of a recombinase (Cre or Flp) that recognizes and

mediates strand exchange between specific sequences of DNA (*loxP* or *FRT* sites).

Depending on the location and relative orientation of the two recognition sites, site-

specific recombinases can mediate excision of a sequence (two directly repeated sites

on the same DNA strand), inversion of a sequence (two inverted sites on the same

strand), integration of a circular strand into a linear strand (two sites on separate

strands), or exchange of sequences between two linear strands (two sites on separate

strands) (**Fig. 1A**). Such systems are most commonly used for deletion of genes or

portions of genes to disrupt gene function, as well as for gene activation via the deletion

of a transcriptional termination sequence between a gene and its promoter (Lakso et al.,

1992). The recombinase can be placed under the control of tissue-specific or

developmental-specific promoters, or be introduced via a viral vector to specific sites in

the mouse, to allow gene deletion or activation to be spatially and temporally controlled.

The first use of the Cre-*loxP* system for tissue-specific gene modification was

demonstrated by two different groups. Lakso and colleagues generated transgenic mice

carrying the SV40 large T antigen driven by the lens-specific αA-crystallin promoter, but

separated by a transcriptional stop sequence flanked by *loxP* sites, and crossed them to

mice carrying Cre recombinase expressed from either the αA-crystallin promoter or the
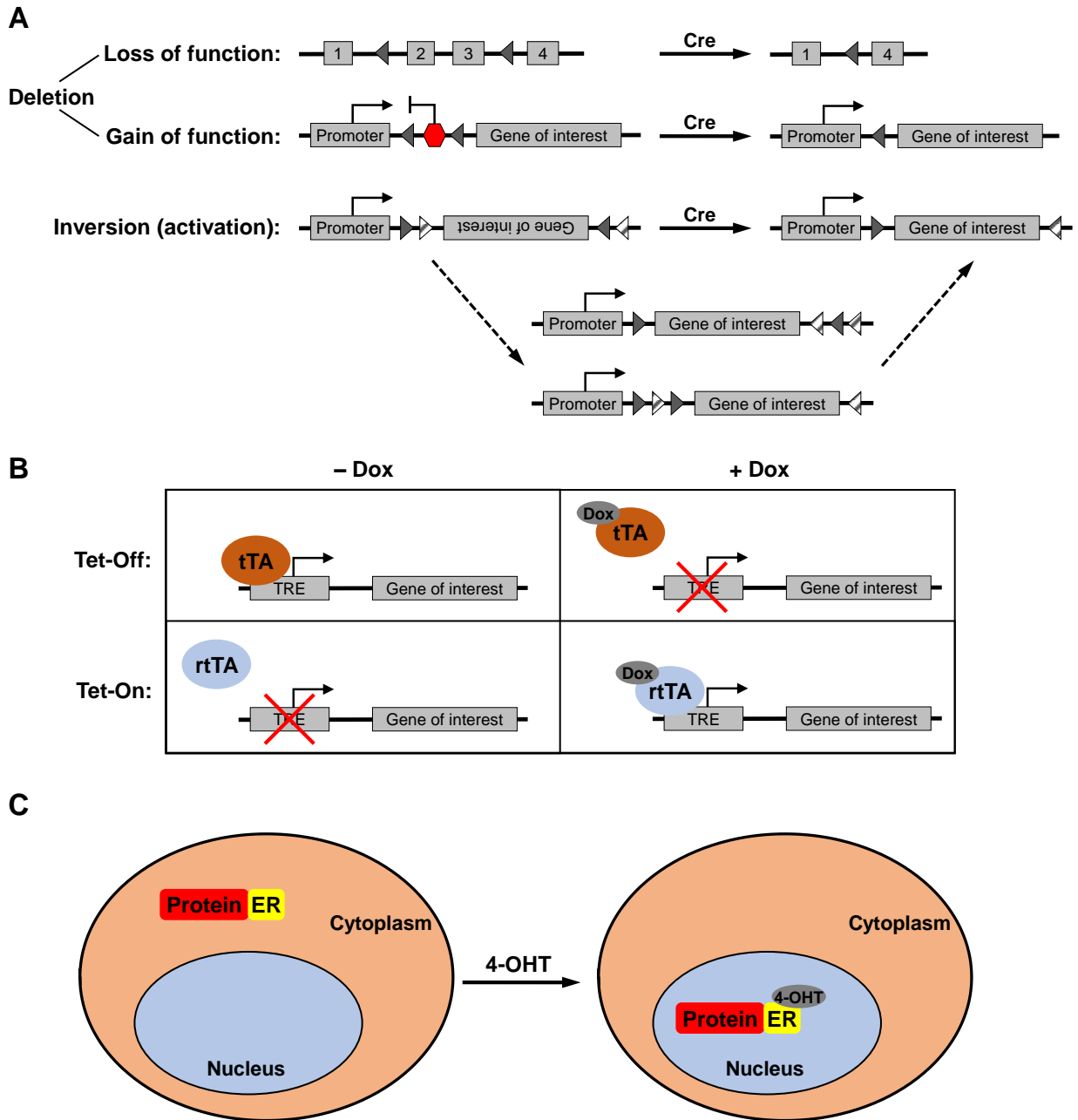
Figure 1: Genetic tools for spatial and temporal control of gene expression.
(A) Use of the Cre-*loxP* system as an example of site-specific recombinase systems. Triangles represent *loxP* sites.
(B) Tetracycline-regulated systems. tTA: tetracycline-controlled transactivator; rtTA: reverse tetracycline-controlled transactivator; TRE: tetracycline-response element; Dox: doxycycline.
(C) Hormone receptor fusion systems. ER: estrogen receptor hormone-binding domain; 4-OHT: 4-hydroxytamoxifen.

human cytomegalovirus promoter (Lakso et al., 1992). These mice went on to develop lens tumors, whereas mice without Cre recombinase did not develop tumors, demonstrating tissue-specific activation of T antigen. Orban and colleagues generated mice expressing a *loxP*-flanked (floxed) β-galactosidase transgene and crossed them to mice expressing Cre recombinase driven by a thymocyte-specific promoter, showing that the β-galactosidase transgene was specifically deleted in thymocytes (Orban et al., 1992).

Subsequently, Gu and colleagues demonstrated the utility of this approach for studying tissue-specific deletion of an endogenous gene (Gu et al., 1994). They generated mice harboring floxed alleles of DNA polymerase β, crossed them to mice harboring a T cell-specific Cre recombinase gene, and demonstrated that they were able to delete DNA polymerase β specifically in T cells with no effects on the rest of the mouse, in contrast to constitutive deletion of DNA polymerase β throughout the mouse, which resulted in embryonic lethality. This study validated the use of Cre-*loxP*-mediated conditional deletion for studying tissue-specific loss of genes that are essential for mouse development.

An alternative method of achieving spatially restricted Cre expression was demonstrated by Shibata and colleagues, who generated an adenoviral vector expressing Cre recombinase, and injected the virus into the colon of mice harboring a modified *Apc* allele that contained *loxP* sites flanking exon 14. This resulted in colon-specific inactivation of *Apc*, leading to the formation of colorectal adenomas and adenocarcinomas (Shibata et al., 1997). In contrast, in an earlier model involving the use of a constitutive loss-of-function *Apc* allele, mice that were heterozygous for the

allele developed tumors predominantly in the small intestine rather than the colon (Fodde et al., 1994).

Cre-*loxP*-based and Flp-*FRT*-based mouse models have now become one of the mainstays of murine models of cancer. Complex models combining multiple conditional alleles have been developed to model multiple genetic events simultaneously, recapitulating what is often observed in human cancers. In addition, because of the ability to restrict Cre expression to specific tissues, the same combination of alleles can be used to model multiple types of cancer. As an example, a mouse model harboring a conditional oncogenic allele of *Kras* and a conditional mutant allele of *Trp53* has been used to model lung adenocarcinoma (Jackson et al., 2005), pancreatic ductal adenocarcinoma (Hingorani et al., 2005), and soft tissue sarcoma (Kirsch et al., 2007).

Tetracycline-regulated systems

In addition to site-specific recombinase-based systems, an alternative system was developed that allowed gene expression to be reversibly toggled on or off. This system was based on the tetracycline resistance operon present in bacteria. In the absence of the antibiotic tetracycline, a repressor protein (TetR) binds to an operator sequence (t*etO*) to repress transcription of the entire operon. When tetracycline is present, it binds to TetR, preventing it from binding to *tetO* and resulting in the transcription of genes that mediate tetracycline resistance.

To adapt this system for use in mammalian cells, Gossen and Bujard fused the TetR protein with the C-terminal domain of the VP16 transactivation domain from the herpes simplex virus (HSV) to form the tetracycline-controlled transactivator (tTA) (Gossen and Bujard, 1992). When combined with a tetracycline-response promoter

30

comprising multiple copies of the *tetO* sequence fused to a minimal promoter, tTA was able to bind to the promoter and activate expression of a transgene in the absence of tetracycline. Expression of the transgene was repressed when tetracycline was added to the cells ("Tet-Off"). A variation of this system was generated by mutation of the TetR gene sequence, generating a version that binds to *tetO* only in the presence of doxycycline, a derivative of tetracycline (Gossen et al., 1995). This version was known as the reverse tetracycline-controlled transactivator (rtTA), and is used in "Tet-On" systems (**Fig. 1B**). Both versions were rapidly adopted for *in vivo* applications (Furth et al., 1994; Kistner et al., 1996). Spatial control of gene expression can be achieved by placing the tTA or rtTA transgene under the control of a tissue-specific promoter (Kistner et al., 1996).

Because of the ability to reversibly switch gene expression on or off, both Tet-Off and Tet-On systems have been used to investigate the requirement for oncogene expression in tumor maintenance. For instance, Felsher and Bishop generated mice expressing human *MYC* in hematopoietic cells using the Tet-Off system, resulting in the development of malignant T cell lymphomas and acute myeloid leukemias. Treatment of tumor-bearing animals with doxycycline to inactivate *MYC* expression caused rapid regression of these tumors (Felsher and Bishop, 1999). Similarly, Chin and colleagues expressed oncogenic H-Ras^{G12V} in melanocytes using the Tet-On system, resulting in melanoma formation in a doxycycline-dependent fashion. Withdrawal of doxycycline resulted in rapid regression of these tumors (Chin et al., 1999). Both of these examples demonstrate that, at least in certain cancers, sustained expression of the initiating oncogene is required for tumor maintenance.

Hormone receptor fusion systems

Another method of achieving reversible control of gene expression involves the fusion of proteins of interest with the hormone-binding domain of nuclear hormone receptors, most commonly the estrogen receptor. Nuclear hormone receptors normally reside in the cytoplasm of the cell in the absence of their cognate hormone. Upon hormone binding, these receptors translocate to the nucleus, where they mediate changes in gene expression. Fusion proteins containing the hormone-binding domain of such receptors can be regulated in a similar manner (**Fig. 1C**). The first demonstration of this method, performed by Picard and colleagues, involved the fusion of the hormone-binding domain of the glucocorticoid receptor to the adenovirus *E1A* gene product, resulting in E1A activity being regulated by dexamethasone (Picard et al., 1988). Subsequently, Eilers and colleagues fused the hormone-binding domain of the estrogen receptor (ER) to *c-myc*, and showed that when introduced into rat fibroblast cells, the fusion protein was able to mediate transformation of the cells in a reversible manner, depending on the presence of β-estradiol (Eilers et al., 1989).

Use of the ER hormone-binding domain for *in vivo* applications was initially complicated by the presence of high levels of circulating estrogens in mice. To overcome this problem, Danielian and colleagues generated a mutant version of the murine ER harboring a G525R mutation, which drastically reduced its affinity for endogenous estrogens but not for the synthetic steroid 4-hydroxytamixofen (4-OHT) (Danielian et al., 1993). The equivalent mutation in the human ER (G521R) similarly abolished binding to endogenous estrogens (Feil et al., 1996). Subsequently, additional mutations to improve the efficiency of activation by 4-OHT resulted in the generation of

the ER^{T2} version of the ER hormone-binding domain, which is currently widely used (Feil et al., 1997; Indra et al., 1999).

The first demonstration of the use of hormone receptor fusion systems *in vivo* was in combination with the Cre-*loxP* system (Feil et al., 1996). Fusion of Cre recombinase with the ER hormone-binding domain enables precise temporal control of Cre activity via the administration of tamoxifen, which is converted to 4-OHT *in vivo*. In addition, by placing expression of the *CreER* fusion gene under the control of a tissue-specific promoter, both spatial and temporal control of Cre activity can be achieved, as first demonstrated in B lymphocytes (Schwenk et al., 1998) and the embryonic neural tube (Danielian et al., 1998).

As with Tet-On and Tet-Off systems, ER domain fusions have also been used to control proto-oncogene and tumor suppressor gene activity *in vivo*. Pelengaris and colleagues generated a transgenic mouse model in which a fusion c-MycER^{TM} protein was placed under the control of the human involucrin promoter to confer epidermal-specific expression. Upon topical application of 4-OHT to the skin, these mice developed premalignant squamous cell neoplasia, which regressed back to normal morphology upon withdrawal of 4-OHT (Pelengaris et al., 1999).

To enable regulation of p53 activity, Christophorou and colleagues generated a knock-in mouse in which the endogenous *Trp53* allele was replaced with one encoding p53ER^{TAM}, rendering p53 activity dependent on tamoxifen administration (Christophorou et al., 2005). In the absence of tamoxifen, the allele functions as a null allele. This allele has been used to study the consequence of p53 restoration in various tumor models,

such as lymphoma and lung adenocarcinoma (Christophorou et al., 2006; Junttila et al., 2010; Martins et al., 2006).

2.4   Speeding up generation of new mouse models

One of the biggest limitations of modeling cancer in mice is the time required to generate new models, as well as the concomitant cost (reviewed in Huijbers et al., 2011). To functionally interrogate a gene of interest in a specific cancer, one needs to generate ES cells with the desired modification (insertion of a transgene or a targeted knockout/knock-in). Following injection of the modified ES cells into blastocysts and implantation into surrogate female mice, chimeric pups are obtained. These are crossed to unmodified mice to determine whether germline transmission has occurred (**Fig. 2**). Mice with successful germline transmission are then crossed into an existing model of interest, which typically requires at least an additional two generations of breeding before a suitable-sized cohort is produced for actual experiments. The whole process takes one to two years and involves significant cost at each step. This has made it a significant challenge to carry out functional validation of the ever-growing lists of candidate genes identified through large-scale sequencing studies or functional genetic screens.

Non-germline mouse models

One way to reduce the time required for generating new mouse models is to bypass the need to introduce genetic modifications into every single cell in the mouse. Traditionally, only chimeric mice in which the modified ES cells have contributed to the germ line are deemed useful, as these mice are subsequently bred to generate mice

Target with desired construct

Derive embryonic stem cells

Inject into wild-type blastocysts

Wild-type mouse or existing GEMM

Chimeric mice

Cross with wild-type mice

Use directly for experiments

X

Non-germline GEMMs

Germline GEMMs
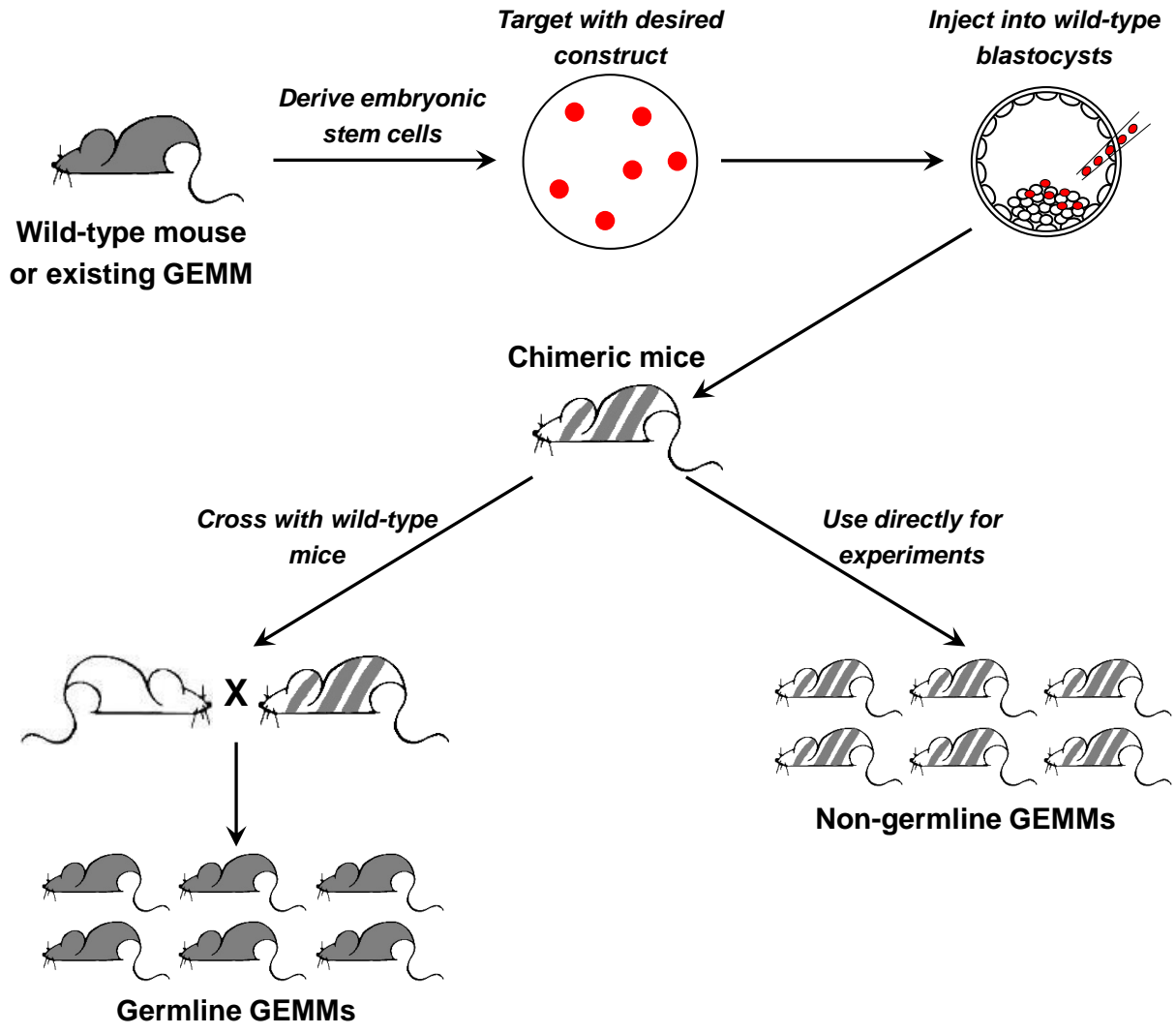
Figure 2: Generation of targeted genetic modifications in mice.

with the desired genetic modification in all the cells in the body. However, several groups have demonstrated the utility of chimeras themselves as models for experimental manipulation, rather than simply as an intermediate step in model generation (Huijbers et al., 2014; Zhou et al., 2010). These have been referred to as non-germline GEMMs (**Fig. 2**).

Instead of starting with ES cells with a single genetic modification, this approach requires the generation of ES cells with all of the desired genetic modifications. Upon injection into wild-type blastocysts and subsequent development into adult mice, each tissue will contain a subset of cells that harbor the full complement of genetic modifications. These chimeric mice are then used for experiments in the same manner as conventional germline models. Two different approaches have been used to generate such ES cells. The first involves the sequential modification of wild-type ES cells, with *in vitro* and *in vivo* validations at various steps in the process. This approach has been used to rapidly generate and compare different models of lung adenocarcinoma driven by distinct oncogenic events, such as $HER2^{V659E}$, $PIK3CA^{myr}$, $EGFR^{L858R}$, and $KRAS^{G12V}$ (Zhou et al., 2010).

The second approach involves the derivation of ES cells from existing well-established GEMMs of cancer (GEMM-ESCs), thereby bypassing the need to re-target these genes. These GEMM-ESCs are then used for introducing new genetic modifications. To further speed up the process of gene targeting, a docking site containing recognition sites for site-specific recombinases can be introduced into the ES cells, which facilitates rapid insertion of DNA constructs expressing genes of interest in a process known as recombinase-mediated cassette exchange (RMCE) (Baer and

Bode, 2001; Bode et al., 2000). Using this technique, Huijbers and colleagues have demonstrated that overexpression of *Mycl1*, *Nfib*, or both *Mycl1* and *Nfib* together, accelerates tumor progression in an established model of small cell lung cancer, thereby validating their oncogenic functions (Huijbers et al., 2014; Semenova et al., 2016).

Generating somatic genetic alterations

Another way to speed up generation of new mouse models of cancer is to bypass the need for ES cell manipulation and chimera generation entirely, by generating genetic alterations directly in the somatic cells of tissues of interest. This can be accomplished, for example, using lentivirus-based vectors. Because lentiviruses stably integrate in the host genome of non-dividing mammalian cells, they can mediate long-term expression of genes of interest in target tissues. This method has been used to model various cancers *in vivo*. For example, a lentiviral vector expressing a Cre-inducible H-RasV12 allele was injected into the brains of mice that expressed Cre under the control of the *GFAP* promoter, which is expressed in neural stem and progenitor cells, as well as terminally differentiated astrocytes. This resulted in the development of glioblastoma multiforme in these mice (Marumoto et al., 2009).

In addition to lentiviral vectors, transposon-based systems have also been used to manipulate somatic cells *in vivo* (reviewed in Tschida et al., 2014). This approach has been successfully used to model hepatocellular carcinoma (Wangensteen et al., 2008) as well as glioblastoma multiforme (Wiesner et al., 2009).

Most recently, the development of the CRISPR-Cas9 system has greatly expanded the ability to precisely target genes of interest in mammalian cells, and has

paved the way for the use of other transient, non-integrating delivery systems, such as adenoviral vectors, for somatic gene editing. This and other aspects of CRISPR-Cas systems will be discussed in greater detail in the next section.

3.　CRISPR-Cas systems in cancer biology

Advances in scientific knowledge are often driven by the development of new technological tools. This has been illustrated in cancer biology by the development of ever-improving models to study the disease. The development of the CRISPR-Cas9 system has proven no different in this regard. However, the CRISPR-Cas9 system has arguably transformed the field more rapidly and broadly than any other preceding technological advance, due to the remarkable simplicity of the system that makes it accessible to practically every aspect of biological research.

In this section, I will discuss key advances in the development of CRISPR-Cas systems that have facilitated their use in cancer biology research.

3.1　Discovery of CRISPR-Cas systems

Clustered regularly interspaced short palindromic repeats, or CRISPR for short, were first observed as an incidental finding in a study by Ishino and colleagues, who were sequencing the *iap* gene in the bacterium *Escherichia coli* (Ishino et al., 1987). They noted that "*Five highly homologous sequences of 29 nucleotides were arranged as direct repeats with 32 nucleotides as spacing*", but the biological significance of these sequences was not investigated. It was only much later that similar arrays of repeats were identified in a wide variety of prokaryotic genomes across many phylogenetic

groups in both Archaea and Bacteria (Mojica et al., 2000). Initially named Short Regularly Spaced Repeats (SRSRs), these arrays were later named CRISPR by Jansen and colleagues in a study that also identified the first CRISPR-associated (*cas*) genes. Such genes were located adjacent to CRISPR loci and were present in CRISPR-positive prokaryotes, but absent from CRISPR-negative prokaryotes (Jansen et al., 2002).

The first clue of the biological function of CRISPR arrays came in 2005, when the spacer sequences between the direct repeats were identified to have homology to foreign genetic elements such as bacteriophages and conjugative plasmids (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005), leading to the proposal that CRISPR systems function as an adaptive immune system in prokaryotic cells (Makarova et al., 2006). In a series of experiments in *Streptococcus thermophilus*, Barrangou and colleagues demonstrated conclusively that this was indeed the case. When wild-type strains of *S. thermophilus* were challenged with two different phages to generate phage-resistant mutants, analysis of the CRISPR loci in the mutant strains showed that they had incorporated new spacer sequences, which corresponded to the regions within the genomes of the respective phages used in the challenge. When these spacer sequences were deleted from the CRISPR locus, the resulting bacterial strain once again became sensitive to infection by that specific phage. Likewise, when the corresponding spacer sequences were introduced into a *S. thermophilus* strain that was sensitive to the same phage, the bacterial strain now gained resistance to infection by that phage (Barrangou et al., 2007).

The identification of an adaptive immune system in prokaryotes was by itself a groundbreaking discovery, but subsequent experiments that elucidated the mechanistic basis of CRISPR-mediated immunity would serve to unleash its true transformative potential.

3.2   CRISPR-Cas9 as an RNA-guided DNA endonuclease

First, it was shown that transcription of the CRISPR array results in the formation of a precursor CRISPR RNA (crRNA), which is then processed by CRISPR-associated (Cas) proteins to form individual crRNAs. These crRNAs then associate with other Cas proteins to mediate antiviral responses (Brouns et al., 2008). Next, it was demonstrated that CRISPR-Cas systems mediate cleavage of double-stranded bacteriophage and plasmid DNA in a sequence-specific fashion (Garneau et al., 2010).

Although a huge diversity of CRISPR-Cas systems have been discovered across prokaryotic species (Koonin et al., 2017), the most widely used is the type II CRISPR-Cas system, due to its simplicity compared with other subtypes. Sapranauskas and colleagues demonstrated that in type II systems, a single Cas protein, Cas9, was sufficient to mediate CRISPR-encoded interference (Sapranauskas et al., 2011), in contrast to other subtypes that required multiple different Cas proteins to mediate binding and cleavage of the target DNA locus (Brouns et al., 2008; Hale et al., 2012; Sinkunas et al., 2011; Zhang et al., 2012). As a result, type II CRISPR-Cas systems are also commonly referred to as CRISPR-Cas9 systems.

CRISPR-Cas9 systems encode a separate trans-activating CRISPR RNA (tracrRNA) that is required for crRNA maturation as well as for cleavage of target DNA

in combination with Cas9 and a crRNA (Jinek et al., 2012). Jinek and colleagues showed that a chimeric single guide RNA (sgRNA), which combines both a crRNA and a tracrRNA, can replace both RNAs for binding to Cas9 and guiding it to target DNA sites (Jinek et al., 2012). This was a significant development because it meant that only two components, Cas9 and an sgRNA, were required to mediate site-specific DNA cleavage. Soon after this, multiple groups successfully adapted the CRISPR-Cas9 system from *Streptococcus pyogenes* for genome engineering in mammalian cells (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013), paving the way for its widespread adoption.

Because of the ease of designing new sgRNA sequences, virtually any location in the genome can be targeted, constrained only by the requirement for a protospacer adjacent motif (PAM) sequence adjacent to the target sequence. Thus, the CRISPR-Cas9 system can be used for any application that requires the creation of a DNA double-stranded break at a specific site in the genome. The most frequent use involves the generation of loss-of-function mutations in target genes. This takes advantage of the cell-intrinsic non-homologous end joining (NHEJ) mechanism for repairing DNA double-stranded breaks. Due to the imprecise, error-prone nature of NHEJ, the repair of double-stranded breaks frequently generates small insertion or deletion mutations, which can result in frameshift mutations when targeted to the coding regions of protein-coding genes (**Fig. 3A**).

The other major mechanism for repairing double-stranded breaks in the cell is homology-direct repair (HDR), in which the cell uses a homologous sequence as a template to repair the broken DNA. This can be utilized together with the CRISPR-Cas9
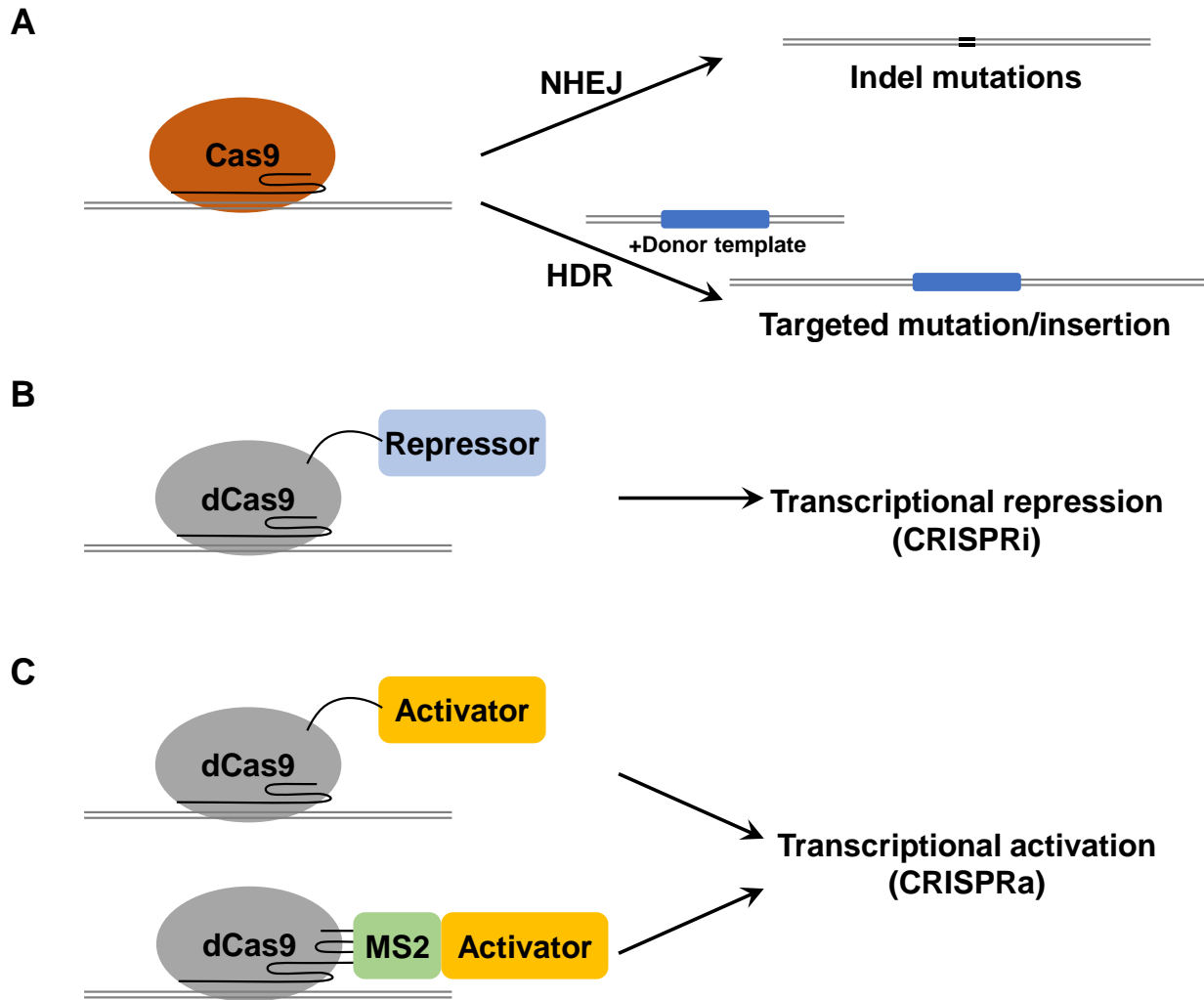
Figure 3: Use of CRISPR-Cas9 for genome editing and transcriptional regulation.
(A) Use of wild-type, catalytically active Cas9 to generate DNA double-stranded breaks, which are repaired by non-homologous end joining (NHEJ) or homology-directed repair (HDR).
(B) Use of catalytically inactive Cas9 (dCas9) fused to a transcriptional repressor for CRISPR interference (CRISPRi).
(C) Two strategies for CRISPR activation (CRISPRa), one using dCas9 fused to a transcriptional activator, and the other using a modified sgRNA scaffold with MS2 binding loops to recruit a MS2-activator fusion protein.

system to generate precise and targeted modifications, by co-introducing a donor DNA template containing the desired modification (**Fig. 3A**). CRISPR-Cas9-mediated homologous recombination occurs at a significantly higher frequency compared with traditional homologous recombination, enabling the simultaneous targeting of multiple loci within a single cell such as a zygote (Wang et al., 2013; Yang et al., 2013).

3.3   Expansion of the CRISPR-Cas toolbox

It was recognized very early on that the ability of sgRNAs to direct Cas9 to specific DNA sequences in the genome could be utilized for purposes other than genome engineering. By mutating the two nuclease domains in Cas9 that cleave double-stranded DNA, Qi and colleagues created a catalytically inactive version of Cas9, dCas9, and showed that it retained its sgRNA-guided, sequence-specific binding activity (Qi et al., 2013). Therefore, this system can be used as a general RNA-guided DNA recognition platform. For example, dCas9 has been adapted for locus-specific epigenetic regulation, including DNA methylation (Liu et al., 2016; McDonald et al., 2016; Vojta et al., 2016), DNA demethylation (Liu et al., 2016; Morita et al., 2016; Xu et al., 2016), histone acetylation (Hilton et al., 2015) and histone demethylation (Kearns et al., 2015). Fusion of dCas9 with cytidine deaminase enzymes such as APOBEC1 or AID also enables targeted base editing to generate specific point mutations (Komor et al., 2016; Nishida et al., 2016).

Transcriptional activation or repression

One area in which dCas9 has been utilized extensively is to mediate gene-specific transcriptional activation (CRISPR activation, or CRISPRa) or repression

(CRISPR interference, or CRISPRi). For example, dCas9 binding can be used directly to sterically hinder transcriptional elongation, RNA polymerase binding, or transcription factor binding (Qi et al., 2013). Alternatively, dCas9 can be fused with transcriptional repression domains, such as the Krüppel associated box (KRAB) domain from ZNF10, to mediate direct transcriptional inhibition (**Fig. 3B**) (Gilbert et al., 2013). Likewise, transcriptional activation domains, such as the VP16 transcriptional activator from the herpes simplex virus, can be fused to dCas9 to mediate transcriptional activation (**Fig. 3C**) (Gilbert et al., 2013; Maeder et al., 2013; Perez-Pinera et al., 2013). Because of the relatively variable efficiencies of these "first-generation" CRISPRa systems, improved versions that combine multiple transcriptional activation domains have been developed. Zalatan and colleagues modified the sgRNA scaffold to include hairpins derived from the bacteriophage MS2, which allows the recruitment of a fusion protein consisting of the MS2 coat protein fused to the VP64 activator (**Fig. 3C**) (Zalatan et al., 2015). Similarly, Konermann and colleagues combined a dCas9-VP64 fusion protein with an sgRNA-MS2 hairpin that recruits an MS2-p65-HSF1 fusion protein (referred to as the synergistic activation mediator, or SAM), and demonstrated that the recruitment of three different transcriptional activators results in potent transcriptional activation (Konermann et al., 2015). Tanenbaum and colleagues took a different approach with their SunTag system, using dCas9 fused to tandem copies of the GCN4 peptide, which recruits multiple copies of GCN4-binding single-chain variable fragment (scFv) antibodies fused to VP64 (Tanenbaum et al., 2014).

An alternative approach for transcription activation, which utilizes wild-type, catalytically active Cas9, was developed by Dahlman and colleagues. This involves the

use of truncated sgRNAs with 14- or 15-nucleotide target sequences instead of the usual 20-nucleotide sequences, which was shown to abolish the nuclease activity of Cas9. These truncated sgRNAs are engineered with MS2 binding loops to recruit the MS2-p65-HSF1 activation complex (Dahlman et al., 2015). This system offers the advantage of allowing both gene knockout and gene activation to be performed simultaneously using the same Cas9 protein, eliminating the need to express both Cas9 and dCas9.

Other CRISPR-Cas systems

As discussed earlier, there is a huge diversity in CRISPR-Cas systems across different species, with different variants having different properties. For example, Cpf1 (also known as Cas12a) is a class 2, type V CRISPR-Cas enzyme that requires only a crRNA molecule, without a tracrRNA, for targeting DNA (Zetsche et al., 2015). Cpf1 is also able to process its own precursor crRNA without the need for additional components (Fonfara et al., 2016), greatly simplifying the delivery of multiple crRNAs for multiplexed genome editing (Zetsche et al., 2015). Furthermore, Cpf1 has been found to have significantly higher specificity compared with Cas9 (Kim et al., 2016; Kleinstiver et al., 2016). Unsurprisingly, catalytically inactive Cpf1 has been adapted for transcriptional activation (Tak et al., 2017).

In addition, some CRISPR-Cas systems target RNA instead of DNA, which can be exploited for post-transcriptional control of gene expression in a similar manner as RNAi. For example, Cas13a (previously known as C2c2), a class 2, type VI CRISPR-Cas enzyme, exhibits RNA-guided RNA endonuclease activity, and has been used in mammalian cells for RNA knockdown (Abudayyeh et al., 2016, 2017). A catalytically

inactive version of Cas13a has also been fused to the adenosine deaminase ADAR2, allowing site-specific RNA editing (Cox et al., 2017). Like Cpf1, Cas13a is able to process its own precursor crRNAs, facilitating multiplexed gene targeting (East-Seletsky et al., 2016). Ongoing efforts will likely uncover additional CRISPR-Cas variants that have unique useful properties.

3.4   Use of CRISPR-Cas9 in cancer biology

The CRISPR-Cas9 system has been used extensively for various applications in cancer biology (reviewed in Sánchez-Rivera and Jacks, 2015). As discussed earlier, CRISPR-Cas9 has been used for performing genetic screens in cancer cell lines to identify cancer type and genotype-specific vulnerabilities. CRISPR-mediated knockout of genes provides several key advantages over RNAi-mediated approaches, including higher consistency, lower noise, and lower off-target effects (Evers et al., 2016). Consequently, CRISPR-based screens uncover both fewer false-negative and fewer false-positive hits compared with RNAi-based screens (Hart et al., 2015; Munoz et al., 2016).

Most screens have used catalytically active Cas9 to generate loss-of-function mutations in genes via NHEJ-mediated repair of double-stranded breaks. However, such screens have been shown to generate false-positive hits in amplified genomic regions, as the multiple DNA cleavage events induce a potent DNA-damage response that leads to cell cycle arrest (Aguirre et al., 2016; Munoz et al., 2016). This can potentially be overcome by using CRISPRi-based transcriptional repression in place of CRISPR knockout. Furthermore, CRISPRi can also be used to repress the transcription

of noncoding RNAs (Gilbert et al., 2014). In addition, CRISPRa-based transcriptional

activation enables gain-of-function screens to be performed. Both genome-scale

CRISPRi and CRISPRa libraries have been developed for use in genetic screens

(Gilbert et al., 2014), and such approaches are likely to complement CRISPR knockout

screens.

In vivo cancer modeling using CRISPR-Cas9

Due to the ease of generating multiple targeted mutations in cells, the CRISPR-

Cas9 system has enabled the rapid generation of new *in vivo* models of cancer. For

example, Yang and colleagues demonstrated the feasibility of simultaneously inserting

two *loxP* sites flanking a target gene in zygotes, enabling one-step generation of mice

with novel conditional alleles (Yang et al., 2013). This approach can also be used for

generating multiple genetic alterations in ES cells, which is potentially useful when

combined with the use of non-germline GEMMs of cancer.

The high efficiency of genome engineering by CRISPR-Cas9 has also expanded

the use of somatic genome engineering approaches for modeling cancer *in vivo*.

Because genome editing by CRISPR-Cas9 requires only transient expression of Cas9

and an sgRNA, this expands the range of possible delivery methods beyond stably

inherited, continuously expressed systems such as lentiviruses or transposons. For

example, Xue and colleagues delivered plasmids expressing Cas9 and sgRNAs

targeting *Pten* and *p53* to the liver of wild-type mice by hydrodynamic tail-veil injection,

and demonstrated that transient expression of these components was sufficient to

induce liver tumor formation (Xue et al., 2014). Likewise, by using a single adenoviral

vector expressing both Cas9 and two sgRNAs targeting specific sites in the *Eml4* and

Alk genes, Maddalo and colleagues were able to generate lung tumors driven by the

Eml4-Alk fusion gene in wild-type mice. This recapitulates a gain-of-function oncogenic

event observed in a subset of human non-small cell lung cancers (Maddalo et al.,

2014). Many other non-viral delivery methods, such as nanoparticle-based systems,

have been, and continue to be, developed for delivery of CRISPR-Cas9 components *in*

vivo, both in mice and humans (Dowdy, 2017).

Combining CRISPR-Cas9 with existing GEMMs of cancer allows combinations of

multiple genetic events to be rapidly modeled. For example, Sánchez-Rivera and

colleagues generated a tri-functional lentiviral vector expressing Cas9, sgRNA and Cre

recombinase, which allows the CRISPR-Cas9 system to be used in existing Cre-*loxP*-

based GEMMs, such as the $Kras^{LSL-G12D}$; $p53^{flox/flox}$ model of lung adenocarcinoma

(Sánchez-Rivera et al., 2014).

Mice carrying constitutive or inducible alleles of Cas9 have been developed by

several groups (Chiou et al., 2015; Dow et al., 2015; Platt et al., 2014). This allows the

use of certain delivery systems, such as adeno-associated viruses (AAVs), which were

previously not feasible due to the large size of the Cas9 gene (4,104 bp. excluding the

stop codon, for *Streptococcus pyogenes* Cas9). These Cas9 alleles have been used to

model various cancers in mice, including lung adenocarcinoma (Platt et al., 2014),

pancreatic ductal adenocarcinoma (Chiou et al., 2015), colorectal carcinoma (Dow et

al., 2015; Roper et al., 2017), and soft tissue sarcoma (Huang et al., 2017). In addition,

CRISPRa systems have recently been adapted for use *in vivo* (Liao et al., 2017;

Wangensteen et al., 2017).

In summary, the versatility of the CRISPR-Cas9 system has vastly increased the speed at which new *in vivo* cancer models can be generated, and its use is likely to continue to be expanded to other cancer types in the future.

Part II: Small cell lung cancer

4.   Background

Lung cancer is the leading cause of cancer deaths in the United States as well as worldwide, accounting for around 13% of new cancer cases annually, but nearly 25% of cancer-related deaths (Ferlay et al., 2015; Siegel et al., 2018). A large fraction of cases can be attributed to tobacco smoking, although there is a substantial proportion of patients who are never-smokers. Lung cancer can be classified into two major subtypes: small cell lung cancer (SCLC), which comprises around 13-15% of lung cancers (Govindan et al., 2006), and non-small cell lung cancer (NSCLC), which comprises the remaining cases and can be further divided into several major types, including lung adenocarcinoma, squamous cell carcinoma, and large cell carcinoma.

Despite decades of research that have significantly advanced our understanding of the molecular drivers of lung cancer progression, treatment outcomes for patients diagnosed with lung cancer remain poor, with an overall 5-year survival of 18.7% (Howlader et al.). The prognosis is even worse for patients with SCLC (5-year survival of 6.3%). In recognition of this, SCLC was one of two cancer types, along with pancreatic ductal adenocarcinoma, to be included under the Recalcitrant Cancer Research Act of 2012 by the United States Congress, which aimed to further advance research into these deadly cancers. The work in the rest of this thesis will focus specifically on SCLC.

4.1   Characteristics of SCLC

SCLC is a high-grade neuroendocrine lung carcinoma that is characterized by rapid growth and early, widespread metastasis, with around 70% of patients presenting with metastatic disease at the point of diagnosis (Califano et al., 2012). As with other subtypes of lung cancer, this late diagnosis results in very poor prognosis for most patients.

Histologically, SCLC tumors are characterized by small cell size (less than the diameter of 3 small resting lymphocytes) with scant cytoplasm, nuclei with finely granular nuclear chromatin and absent or faint nucleoli, high mitotic rate, and frequent necrosis (Travis et al., 2004). SCLC tumors also stain for characteristic neuroendocrine markers, such as achaete-scute homolog 1 (ASCL1), chromogranin A (CHGA), calcitonin gene-related peptide (CGRP), and synaptophysin (SYP).

4.2   Treatment of SCLC

Clinically, SCLC patients are categorized separately from NSCLC patients due to differences in management of the disease. The standard of care for SCLC is combination cytotoxic chemotherapy (most frequently cisplatin and etoposide), as well as radiotherapy in cases without metastatic spread beyond the thorax. Treatment options for SCLC have not changed significantly for over thirty years. Patients often exhibit robust initial responses to chemotherapy, but relapse almost invariably occurs, and second-line therapy options are generally ineffective (Demedts et al., 2010).

Numerous targeted therapies have been tested in SCLC patients over the years, such as receptor tyrosine kinase inhibitors, inhibitors of the PI3K/AKT/mTOR pathway,

inducers of apoptosis, and angiogenesis inhibitors (Santarpia et al., 2016). A search on the NIH Clinical Trials database in April 2018 listed over 150 ongoing clinical trials involving SCLC. Unfortunately, no targeted therapies have thus far shown significant benefit over the current standard of care in clinical trials.

The use of immunotherapy in cancer treatment has shown significant promise in recent years. Checkpoint inhibitors such as ipilimumab, nivolumab and pembrolizumab, which target specific inhibitory receptors on T cells such as PD-1 and CTLA-4, have been approved for the treatment of a number of cancers, including metastatic melanoma and NSCLC. Multiple trials with various immunotherapy agents are currently ongoing in SCLC (Reck et al., 2016), some with promising early results. For example, in a phase 1/2 trial studying nivolumab plus ipilimumab (targeting both PD-1 and CTLA-4) or nivolumab alone (targeting PD-1 only) in patients with recurrent SCLC after at least one previous platinum-containing regimen, there was an improvement in the objective response rate (25% vs. 11%), median overall survival (7.9 months vs. 4.1 months), as well as 2-year overall survival (30% vs. 17%) for patients receiving the combination compared with nivolumab alone (Hellmann et al., 2017).

In summary, given the relative lack of progress in the treatment of SCLC, especially compared with the growing number of targeted therapies and the success of immunotherapy for treating NSCLC (Byers and Rudin, 2015), there is a significant need to develop novel approaches to treat this disease.

5.   Genetics of SCLC

Because most patients with SCLC present with metastatic disease, surgical resection is rarely performed. This has resulted in a paucity of human tumor samples for molecular analyses, which is reflected in the exclusion of SCLC from the list of cancers selected for study by the TCGA project. Early efforts to identify genetic alterations in SCLC tumors depended largely on cell lines established from human tumors, as well as a limited collection of histological samples (Lai et al., 1995; Onuki et al., 1999; Wistuba et al., 2001). More recently, George, Lim and colleagues performed comprehensive molecular profiling of a large panel of human SCLC tumors, including whole-genome sequencing of 110 tumors, transcriptome sequencing of 81 tumors, and SNP array analysis of 142 tumors, yielding an unprecedented amount of information (George et al., 2015). Chief among the alterations present in SCLC tumors are near-universal inactivation of the tumor suppressor genes *TP53* and *RB1* (George et al., 2015; Wistuba et al., 2001). Other frequently observed genetic alterations include mutations in the known tumor suppressor gene *PTEN*, histone acetyltransferase genes *CREBBP* and *EP300*, inactivating mutations in *NOTCH* family genes, and amplifications in *MYC* family genes, namely *MYC, MYCL1* and *MYCN* (George et al., 2015). Mutations in *RBL1 and RBL2* were also observed in subsets of tumors.

Due to the key roles of *TP53* and *RB1* in SCLC, and more broadly in many other cancers, I will discuss these genes in further detail below.

5.1   *TP53* in cancer

Discovery of TP53

TP53 is a tumor suppressor gene that is mutated in around 40-50% of all human tumors (Bailey et al., 2018; Kandoth et al., 2013). It encodes the p53 protein, which was first identified as a cellular protein that co-precipitated with the SV40 large T antigen (Lane and Crawford, 1979; Linzer and Levine, 1979). The p53 protein was found to be expressed at high levels in various transformed cells, including chemically-transformed cells, but not in normal cells such as fibroblasts, lymphoid cells or hematopoietic cells, suggesting a potential role of the protein in transformation (DeLeo et al., 1979). Additional experiments showed that the murine *Trp53* gene could cooperate with *ras* to transform normal murine fibroblasts *in vitro* (Eliyahu et al., 1984; Parada et al., 1984), which led to the classification of *TP53* as an oncogene. However, it soon became apparent that many transformed cell lines harbored inactivating mutations in *Trp53* (Ben David et al., 1988; Mowat et al., 1985; Wolf and Rotter, 1984, 1985); this was also observed in human colorectal carcinoma samples, where both copies of a specific region on human chromosome 17p that encompasses *TP53* are frequently lost or mutated (Baker et al., 1989). Finally, transfection of constructs expressing wild-type *Trp53* suppressed the ability of *ras* and other oncogenes, such as *myc* or adenoviral *E1A*, to transform primary rat embryo fibroblasts (Eliyahu et al., 1989; Finlay et al., 1989). These experiments established *TP53* as a tumor suppressor gene rather than an oncogene.

Functions of p53

p53 is a transcriptional factor that is activated by a wide variety of cellular stress signals, including DNA damage, oncogene activation, ribosomal stress, hypoxia, and oxidative stress (**Fig. 4**). In the absence of stress signals, p53 is usually rapidly degraded in the cell; this is mediated by its binding to the protein MDM2 (Momand et al., 1992), which targets p53 for ubiquitination and subsequent degradation. MDM2 is itself a downstream transcriptional target of p53, thereby forming a negative feedback loop that prevents p53 from accumulating under normal circumstances. Upon activation by cellular stresses, p53 undergoes phosphorylation by kinases, such as CHK1 and CHK2. This prevents p53 from binding to MDM2, resulting in the rapid accumulation of p53, translocation to the nucleus and activation of its transcriptional targets. p53 regulates many different processes in the cell, including many that contribute to its role in tumor suppression, such as cell cycle arrest, DNA repair, apoptosis, senescence and autophagy (**Fig. 4**) (reviewed in detail in Bieging et al., 2014). For example, in response to DNA damage, p53 causes cell cycle arrest at the G1/S checkpoint, allowing the cell to repair the damage before DNA replication occurs. Because of this role, p53 has been dubbed the "guardian of the genome" (Lane, 1992). p53 also has transcription-independent roles, most notably activation of mitochondrial apoptosis by interaction with both anti-apoptotic and pro-apoptotic proteins such as Bcl-xL, Bcl-2 and Bax (Chipuk et al., 2004; Mihara et al., 2003).

Loss of p53 in vivo

In addition to the initial experiments discussed above, several other lines of evidence confirmed the role of *p53* as a tumor suppressor. Firstly, inheritance of a

55
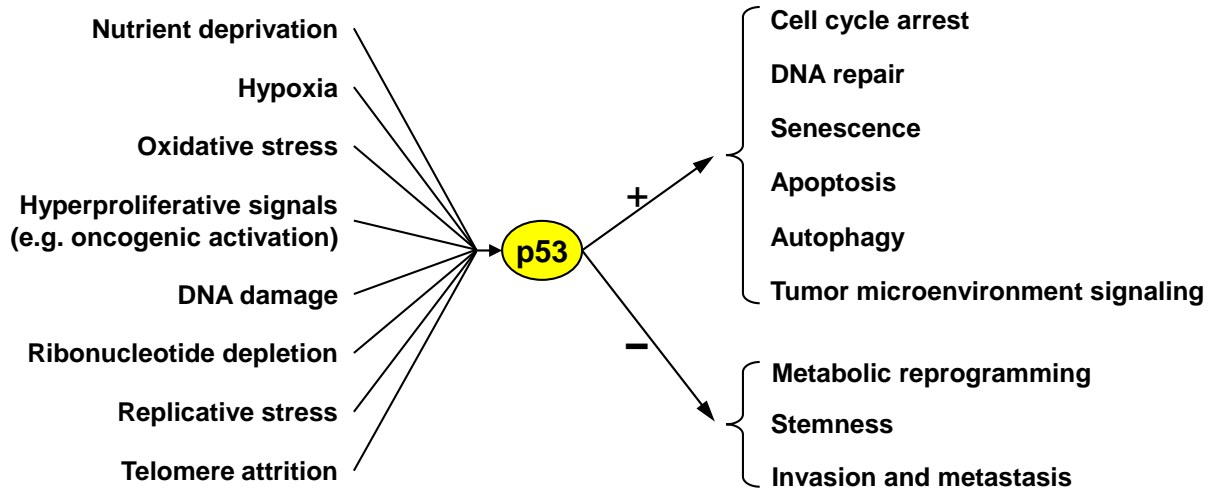
Figure 4: The p53 network.
(Left) Cellular stresses that result in the activation of p53. (Right) p53-regulated processes that contribute to its role as a tumor suppressor. Figure adapted from Bieging, Mello and Attardi, 2014.

mutated copy of *TP53* results in Li-Fraumeni syndrome (Malkin et al., 1990; Srivastava et al., 1990), which is characterized by the early onset of breast cancer, sarcomas, brain tumors, acute leukemias and adrenocortical tumors (Hisada et al., 1998). Secondly, mice with germline null mutations in *Trp53* develop normally, but are predisposed to developing lymphomas and various sarcomas, such as osteosarcoma, rhabdomyosarcoma, hemangiosarcoma, and teratoma (Donehower et al., 1992; Jacks et al., 1994a).

Intriguingly, mice engineered with germline point mutations in *Trp53* corresponding to one of two common hot spot mutations in *TP53*, R175H and R273H, are also predisposed to tumor formation, but these mice develop a different spectrum of tumors compared with mice harboring null alleles of *Trp53* (Lang et al., 2004; Olive et al., 2004). In particular, these mice displayed a significant increase in incidence of carcinomas in a variety of different tissues, including lung, small intestine, colon, breast, skin, liver, and pancreas. This points to potential gain-of-function roles for mutant p53 proteins, beyond the loss of wild-type p53 function (reviewed in Freed-Pastor and Prives, 2012).

p53, p63 and p73

Vertebrates harbor two additional homologs of *TP53*, *TP63* and *TP73* (encoding p63 and p73, respectively). Both p63 and p73 play important roles in development, with *p63*-null mice showing defects in limb, craniofacial and epithelial development (Mills et al., 1999; Yang et al., 1999), and *p73*-null mice showing defects in neurogenesis and pheromone sensory pathways (Yang et al., 2000). This contrasts with the lack of developmental defects in *p53*-null mice (Donehower et al., 1992; Jacks et al., 1994a).

Although mutations in *TP63* or *TP73* are far less frequent in cancer compared with mutations in *TP53*, p63 and p73 have both been demonstrated to have tumor suppressive functions. Mice harboring heterozygous mutations in both *p53* and *p63*, or both *p53* and *p73*, develop a broader spectrum of tumors compared with mice harboring heterozygous mutations in *p53* alone (Flores et al., 2005). In addition, compound-mutant mice exhibited higher tumor burden and developed tumors that were more aggressive and metastatic. Interestingly, a subset of human SCLC tumors were observed to harbor intra-chromosomal deletion events within *TP73* that result in the expression of a truncated version of p73, which is hypothesized to confer a dominant-negative phenotype on wild-type p73, suggesting that loss of p73 function may be functionally relevant in SCLC (George et al., 2015).

5.2   *RB1* in cancer

As discussed earlier in this chapter, the retinoblastoma gene, or *RB1*, was the first tumor suppressor gene to be identified in human cancers (Friend et al., 1986). Mutation of *RB1* occurs at very high frequency in retinoblastoma, osteosarcoma and SCLC, and at lower frequencies in many other cancer types (Bailey et al., 2018; Kandoth et al., 2013). Although the most well-studied function of *RB1* and its protein product, pRB, is in the regulation of cell cycle progression, it has become clear that the gene has many other functions in the cell, all of which may contribute to its function as a tumor suppressor gene.

pRB and cell cycle regulation

pRB, like p53, is a key regulator of the G1/S transition in the cell cycle. In G1, pRB exists in a hypophosphorylated state due to phosphorylation by cyclin D-CDK4/6, which allows it to bind to and sequester E2F family transcription factors in the cytoplasm. During late G1, pRB becomes hyperphosphorylated by cyclin E-CDK2 (**Fig. 5A**). This causes it to release E2F transcription factors, which translocate to the nucleus to activate the transcription of genes involved in progression to S phase (reviewed in Dimova and Dyson, 2005). Accordingly, loss of pRB results in dysregulation of the cell cycle.

Consistent with the importance of this pathway in cancers, a number of other genes in the pRB pathway are also frequently altered in cancer (**Fig. 5B**). For example, loss of *CDKN2A*, which encodes both the CDK4/6 inhibitor p16^{INK4A} and the MDM2 inhibitor p14^{ARF}, occurs at significant frequencies in multiple cancers, including glioblastoma (60% of cases) (Brennan et al., 2013; Kandoth et al., 2013). In addition, cyclin D1, CDK4 and CDK6 are also often overexpressed in breast cancer.

Other functions of pRB

Beyond cell cycle regulation, pRB plays important roles in other cellular processes, including differentiation, stem cell regulation, metabolism and apoptosis (Dyson, 2016). For example, to examine the effects of pRB loss on differentiation, Calo and colleagues compared *Meox2-Cre;Rb1^{flox/flox}* mice, in which *Rb1* is deleted in epiblast-derived tissues beginning at embryonic day 6.5, with *Meox2-Cre;Rb1^{+/+}* mice. *Meox2-Cre;Rb1^{flox/flox}* mice exhibited significant impairment of bone differentiation together with expansion of the fat compartment at embryonic day 18.5 compared with
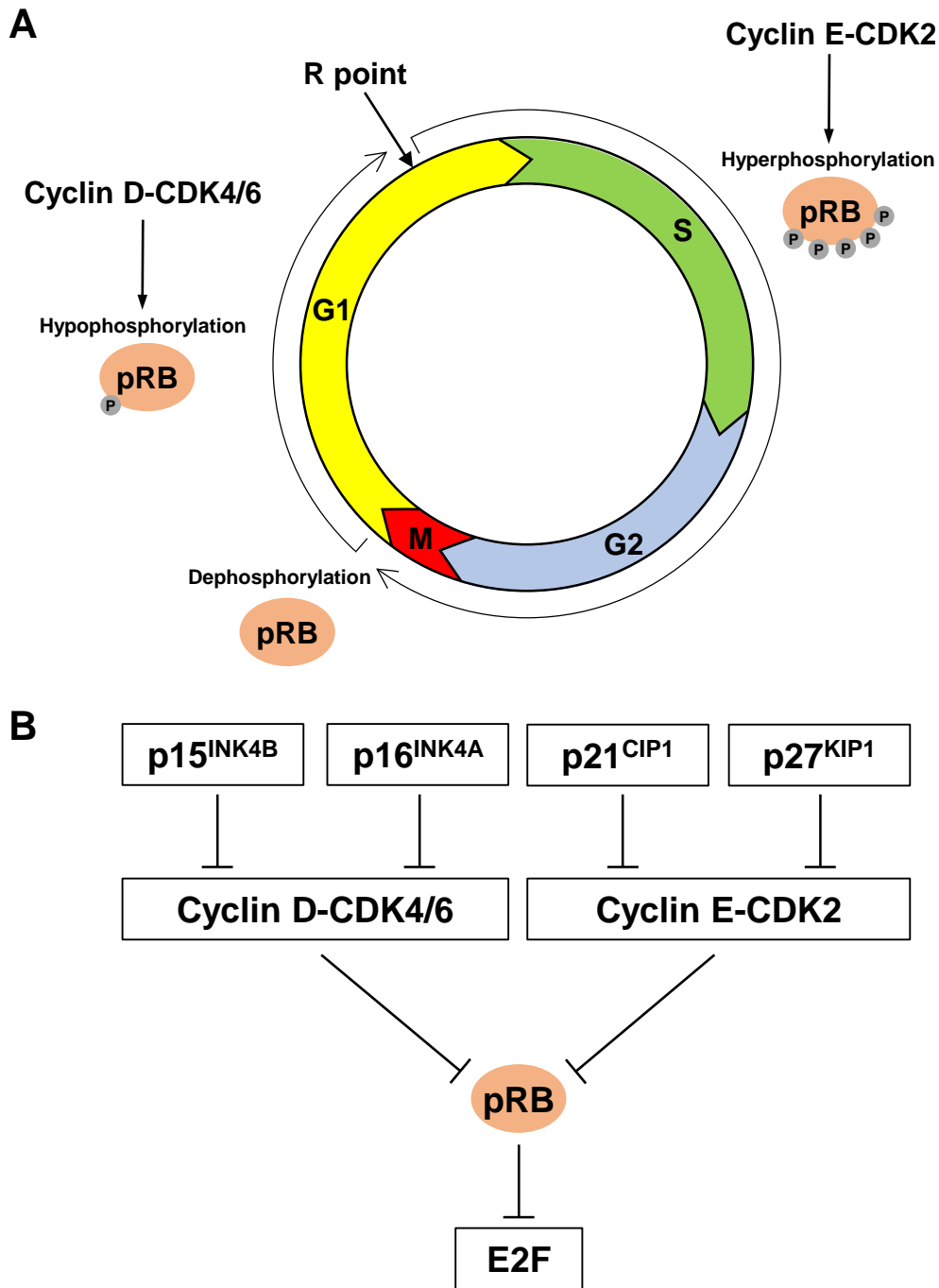
Figure 5: pRB and regulation of the cell cycle.
(A) Phosphorylation state of pRB at different phases of the cell cycle.
(B) Major upstream regulators of pRB, which are commonly altered in cancer.

Meox2-Cre;Rb1⁺/⁺ mice, showing that *Rb1* loss promotes adipogenesis over

osteogenesis *in vivo* (Calo et al., 2010).

pRB has been shown to induce apoptosis both transcriptionally as well as

through direct protein interactions. In response to DNA damage or oncogenic stress,

pRB-E2F1 complexes promote transcriptional activation of pro-apoptotic genes (Ianari

et al., 2009). In addition, a distinct population of pRB in the cell has been shown to be

localized to mitochondria, where it can directly bind and activate the pro-apoptotic

protein Bax to induce apoptosis (Hilgendorf et al., 2013). The existence of a separate

population of pRB in the cell supports the hypothesis that multiple pools of pRB may

exist within the cell, each of which performs functions distinct from its canonical role in

cell cycle regulation (Dyson, 2016).

pRB has also been shown to regulate multiple metabolic processes, such as

glucose tolerance, oxidative metabolism, the tricarboxylic acid (TCA) cycle, and

nucleotide synthesis (Nicolay and Dyson, 2013). Nicolay and colleagues performed

quantitative proteomic analysis to assess the effects of acute *Rb1* loss in different

tissues in adult mice, and found that *Rb1* knockout led to a significant decrease in

mitochondrial protein levels, decreased mitochondrial mass and reduced mitochondrial

function in both lung and colon (Nicolay et al., 2015). Consistent with this finding, they

demonstrated that *RB1*-knockout human retinal cells displayed increased sensitivity to

mitochondrial stress, pointing to a potential vulnerability in *RB1*-null cells that could be

exploited therapeutically.

pRB, p107 and p130

pRB is part of a family that includes two closely related proteins, p107 and p130 (encoded by *RBL1* and *RBL2*, respectively), which are collectively known as the pocket protein family. Like pRB, both proteins were first discovered due to their ability to bind to viral oncoproteins such as adenoviral E1A (Dyson et al., 1989; Ewen et al., 1989; Harlow et al., 1986). Subsequent isolation and characterization of the genes coding for these two proteins revealed substantial homology with *RB1* (Ewen et al., 1991; Hannon et al., 1993; Li et al., 1993). All three proteins are involved in cell cycle regulation through the binding of E2F transcription factors, but while pRB preferentially binds to E2F1, E2F2 and E2F3, both p107 and p130 preferentially bind to E2F4 and E2F5 (Beijersbergen et al., 1994; Ginsberg et al., 1994; Hijmans et al., 1995; Vairo et al., 1995). In addition, while p130 levels are highest in quiescent G0 cells and decline rapidly as cells are stimulated to re-enter the cell cycle, p107 levels show an inverse trend, being low in quiescent cells and rising as cells progress through G1 and S phases (Cobrinik et al., 1993; Devoto et al., 1992; Shirodkar et al., 1992). Thus, all three proteins are thought to play distinct but overlapping roles in cell cycle regulation.

Loss of pRB, p107 and p130 in vivo

In vivo, loss of pRB, p107 or p130 results in very different phenotypes. *Rb1* homozygous knockout in mice leads to severe defects in erythropoiesis as well as neuronal cell death, and results in embryonic lethality by E14.5-15.5 (Jacks et al., 1992; Lee et al., 1992). In contrast, both *p107* homozygous knockout mice and *p130* homozygous knockout mice are viable and appear developmentally normal (Cobrinik et al., 1996; Lee et al., 1996). There appears to be some level of functional redundancy

among the three proteins – *Rb1;p107* double knockout embryos exhibit earlier

embryonic lethality at E10.5-E12.5, compared with E14.5-E15.5 for *Rb1* knockout alone

(Lee et al., 1996), while *p107;p130* double knockout embryos die soon after birth with

defects in chondrocyte proliferation and bone development (Cobrinik et al., 1996).

Notably, there appears to be significant strain-dependent differences in phenotypes.

The previous studies were all performed on a mixed 129/Sv;C57BL/6 background. In

contrast, in the BALB/c background, loss of *p107* alone results in viable and fertile mice,

but with significantly impaired growth rates as well as myeloproliferative disorders, while

loss of *p130* alone is embryonic lethal, with embryos displaying impaired neurogenesis

and myogenesis (LeCouter et al., 1998a, 1998b).

Heterozygous loss of *Rb1* in mice results in a predisposition towards pituitary

tumors (Jacks et al., 1992) and medullary thyroid carcinomas (Williams et al., 1994).

One big surprise, however, was that *Rb1$^{+/-}$* mice do not develop retinoblastoma, in

contrast to cases of familial retinoblastoma in humans. This indicates the existence of

species-specific differences between mice and humans in terms of pRB function, at

least in retinal cells.

Unlike *RB1*, *p107* and *p130* are rarely mutated in cancers. For example, *RB1* is

mutated in around 90% of human SCLC tumors, but *p107* and *p130* are mutated in only

around 6% of SCLC tumors each (George et al., 2015). However, *in vivo* studies in

knockout mice suggest that both *p107* and *p130* have tumor suppressive effects in

certain contexts. In contrast to *Rb1$^{+/-}$* mice, *Rb1$^{+/-}$;p107$^{-/-}$* mice develop dysplastic

lesions in the retina (Lee et al., 1996), suggesting that p107 may compensate for pRB

loss in retinal cells in the mouse. Subsequent studies utilizing either retinal-specific

mutation of *Rb1* in a *p107*-null or *p130*-null background, or chimeric mice carrying

compound *Rb1/p107* or *Rb1/p130* mutations, demonstrated conclusively that the

additional loss of *p107* or *p130* on top of *Rb1* loss results in retinoblastoma formation

(Dannenberg et al., 2004; MacPherson et al., 2004, 2007; Robanus-Maandag et al.,

1998).

In addition to their differing roles in cell cycle regulation, p107 and p130 appear

to have different functions *in vivo* as well. For example, lung-specific ablation of *Rb1* in

a *p107*-null background results in increased cell proliferation in the lung epithelium

compared with ablation of *Rb1* in a wild-type p107 background. In contrast, ablation of

Rb1 in a *p130*-null background has no effect on proliferation, but results in decreased

apoptosis, which is not observed in the *p107*-null background (Simpson et al., 2009).

Further investigation of pRB, p107 and p130 functions in other *in vivo* contexts is likely

to yield insights into functional differences among the three proteins.

6.   Genetically engineered mouse models of SCLC

As previously discussed, most patients with SCLC present with metastatic

disease. This has made it challenging to study early stages of SCLC progression using

human tumor samples. The development of GEMMs of SCLC has greatly facilitated this

aspect of SCLC research. In this section, I will discuss the various GEMMs of SCLC

that have been developed, as well as some of the key biological insights that have been

gained from the use of these models.

6.1 *Trp53/Rb1* double knockout model

Consistent with the near-universal inactivation of *TP53* and *RB1* in human SCLC, *Rb1*$^{+/-}$;*p53*$^{-/-}$ mice were observed to develop bronchial neuroendocrine hyperplasia (Williams et al., 1994). However, these hyperplastic lesions did not develop into frank tumors, as the mice succumbed to other tumors by two to six months of age. The first model of SCLC was developed by Meuwissen and colleagues, who combined the use of conditional alleles of *Trp53* and *Rb1* with an adenoviral vector expressing Cre recombinase (Meuwissen et al., 2003). The conditional *Trp53* allele harbors *loxP* sites flanking exons 2 to 10 of the gene (Jonkers et al., 2001), while the conditional *Rb1* allele harbors *loxP* sites flanking exon 19 (Vooijs et al., 2002). Crucially, to achieve deletion of both genes specifically in the lung epithelium, the adenoviral vector was administered through the trachea of the mice (**Fig. 6A**). This resulted in the development of multiple independent tumors in the lungs with a median latency of 210 days. These tumors displayed the key histological features of human SCLC, such as scant cytoplasm, hyperchromatic nuclei, high mitotic activity and fields of necrosis. They also stained for neuroendocrine markers such as synaptophysin, neural cell adhesion molecule (Ncam1), calcitonin gene-related peptide (CGRP), and neuron-specific enolase. In addition, these tumors frequently metastasized to distant organs such as the bone, brain, adrenal gland, ovary, and liver (**Fig. 6A**), with the metastatic lesions having identical morphological appearance to the primary tumors. Therefore, this model recapitulates the key features of human SCLC. A different conditional allele of *Rb1*, in which *loxP* sites flank exon 3 instead of exon 19 (Sage et al., 2003), has also been used in this model, with no difference in the resulting phenotype (Schaffer et al., 2010).

A

B
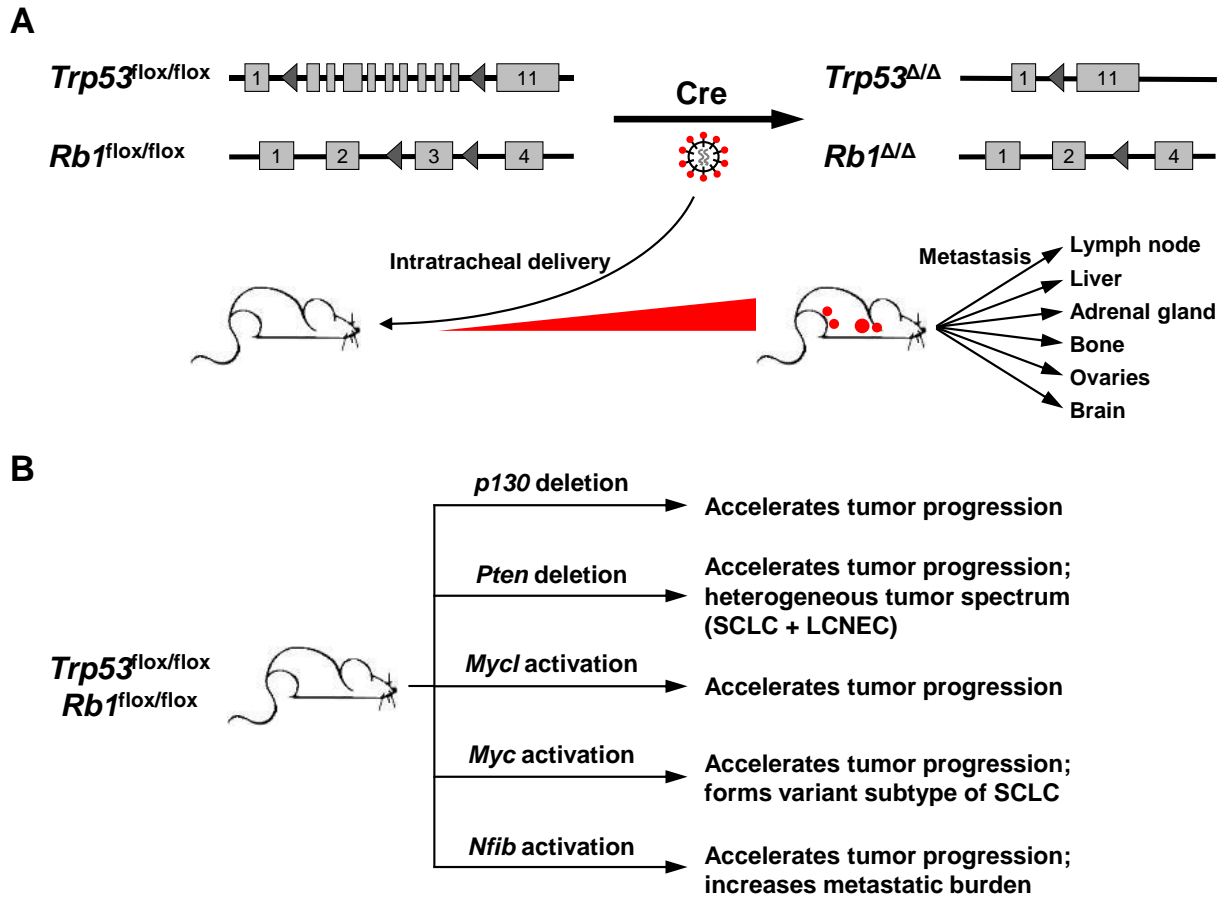
Figure 6: Mouse models of SCLC.

(A) *Trp53*^{flox/flox}; *Rb1*^{flox/flox} model of SCLC. Exons 2-10 of *Trp53* and exon 3 of *Rb1* (exon 19 in the original model by Meuwissen et al., 2003) are flanked by *loxP* sites. Sites of metastasis observed in the model are indicated.

(B) Additional genetic alterations that have been developed on the *Trp53*^{flox/flox}; *Rb1*^{flox/flox} background.

Genomic analysis of murine SCLC tumors isolated from the *Trp53/Rb1* double knockout model revealed that these tumors harbored significantly fewer point mutations compared with human SCLC tumors (McFadden et al., 2014). This is likely due to the absence of carcinogens such as cigarette smoke in the mouse model. However, the murine tumors developed DNA copy number alterations and genomic rearrangements at a similar frequency as human SCLC. In addition, certain genomic alterations, such as point mutations in *Pten* and focal amplification of *Mycl*, are observed in both human and mouse tumors. Therefore, this model of SCLC also recapitulates several key aspects of the genomic landscape in human SCLC.

6.2 Derivatives of the *Trp53/Rb1* model

Several groups have built upon the initial *Trp53/Rb1* double knockout model of SCLC by incorporating additional genetic alterations that are observed in human SCLC (**Fig. 6B**). Schaffer and colleagues developed the *Trp53/Rb1/p130* model of SCLC by combining a conditional allele of *p130* (*Rbl2*) with the existing double knockout model (Schaffer et al., 2010), and demonstrated that loss of *p130* dramatically accelerates tumor progression in these mice. Likewise, two different groups demonstrated that loss of *Pten* accelerates tumor progression in the *Trp53/Rb1* double knockout background (Cui et al., 2014; McFadden et al., 2014). Because both *p130* (6%) and *PTEN* (9%) are mutated in human SCLC, these studies validated their roles as functional tumor suppressors in SCLC.

Amplification of all three *MYC* family genes (*MYC, MYCN* and *MYCL)* occur in a significant fraction of human SCLC tumors, suggesting a role for the activation of *MYC*

family genes in SCLC. Two separate studies demonstrated distinct phenotypes resulting from activation of either *Myc* or *Mycl* in murine SCLC. To model activation of *Mycl* in SCLC, Huijbers and colleagues used the GEMM-ESC approach combined with RMCE, as described earlier in this chapter, to introduce a Cre-inducible allele of *Mycl* into the *Trp53/Rb1* double knockout model, allowing the overexpression of Mycl in SCLC tumors (Huijbers et al., 2014). Overexpression of Mycl resulted in a significant acceleration of tumor progression, reducing the median survival of the mice from 235 days to 140 days. Mollaoglu and colleagues took a different approach to model activation of *Myc* in SCLC, generating a Cre-activated *Myc*[T58A] allele and crossing it into the *Trp53/Rb1* double knockout background (Mollaoglu et al., 2017). The T58A mutation has been shown to stabilize Myc by blocking its phosphorylation by GSK3β, thereby preventing Myc from being targeted for ubiquitination (Sears et al., 2000). Activation of Myc drastically accelerated tumor progression (median survival of 60 days), but the resulting tumors predominantly exhibited histological features that were consistent with a variant form of SCLC, with slightly larger cells, prominent nucleoli and well-defined cytoplasm. Thus, Mycl and Myc appear to play distinct roles in SCLC.

6.3   Biological insights from mouse models of SCLC

Many key insights into SCLC biology have been learnt from studies using the various GEMMs of SCLC. The ability to study early stages of disease progression, as well as the relative ease of obtaining tumor tissue from primary tumors and metastases, has enabled detailed mechanistic studies that are not possible using human tumor samples or cell lines.

Cell of origin for SCLC

As SCLC is a neuroendocrine tumor, the cell of origin for SCLC has been hypothesized to be pulmonary neuroendocrine cells in the bronchi and bronchioles of the lung, which exist either as solitary cells or as clusters called neuroepithelial bodies (NEBs). Using the *Trp53/Rb1* model, Sutherland and colleagues were able to experimentally determine the cell of origin for SCLC. They generated different adenoviral vectors that expressed Cre recombinase driven by different cell type-specific promoters (Sutherland et al., 2011). The *CC10* promoter was used to target club cells, the *SPC* promoter was used to target alveolar type 2 (AT2) cells, and the *CGRP* promoter was used to target pulmonary neuroendocrine cells. Using these viruses to infect *Trp53/Rb1* mice, they demonstrated that most of the animals infected with CGRP-Cre virus developed neuroendocrine tumors, supporting the role of neuroendocrine cells as the predominant cell of origin for SCLC. Interestingly, a significant fraction of mice infected with SPC-Cre virus also developed neuroendocrine tumors, albeit at a lower frequency compared with CGRP-Cre-infected mice, suggesting that a subset of SPC-positive cells can also serve as the cell of origin for SCLC.

Tumor heterogeneity in SCLC

Several studies have examined the role of tumor heterogeneity in SCLC progression and metastasis. Cell lines derived from murine SCLC tumors contain two distinct populations, one that grows as suspended aggregates and expresses high levels of neuroendocrine markers (neuroendocrine, or NE), and the other that grows as a monolayer of cells attached to the cell culture dish and expresses high levels of mesenchymal markers such as *Nestin*, *Vimentin* and *CD44* (non-neuroendocrine, or

non-NE). Calbo and colleagues transplanted these cells subcutaneously into immunocompromised mice and assessed the metastatic potential of the resulting tumors. When either cell population was transplanted in isolation, no metastasis was observed. However, transplantation of a mixture of NE and non-NE cells resulted in metastasis of the NE cells to the liver, suggesting that the presence of non-NE cells in the tumor promotes metastasis of NE cells (Calbo et al., 2011). Subsequent work by Kwon and colleagues showed that this interaction was mediated by paracrine signaling between both populations, resulting in the upregulation of the Pea3 transcription factor that enhances metastasis (Kwon et al., 2015).

Jahchan and colleagues investigated a different aspect of tumor heterogeneity in murine SCLC tumors from the *Trp53/Rb1/p130* triple knockout model, identifying a population of highly tumorigenic tumor-propagating cells (TPCs) in these tumors that express high CD24, low CD44 and high EpCAM levels. This subpopulation of tumor cells expressed elevated levels of Mycl, and treatment of these cells with the BET bromodomain inhibitor JQ1, which has been shown to reduce MYC levels in other tumor models, significantly decreased their tumorigenic potential (Jahchan et al., 2016).

Tumor heterogeneity in SCLC tumors also appears to be regulated by Notch signaling. Lim and colleagues observed heterogeneous expression of Hes1, a transcriptional target of the Notch signaling pathway, in tumors from *Trp53/Rb1/p130* mice. Using a *Hes1^{GFP}* reporter allele, they demonstrated that Notch-high tumor cells, which corresponded to the non-neuroendocrine population, grew more slowly and were relatively resistant to chemotherapy compared with Notch-low cells. Combining chemotherapy with a Notch antagonist significantly suppressed tumor growth both in

murine transplant models and patient-derived xenograft models of SCLC (Lim et al.,

2017).

Role of Nfib in metastatic spread in SCLC

NFIB, which encodes nuclear factor I B, was discovered to be frequently

amplified in murine SCLC tumors, as well as a subset of human SCLC cell lines (Dooley

et al., 2011). Semenova and colleagues demonstrated that overexpression of Nfib in

Trp53/Rb1 tumors accelerated tumor progression and enhanced metastatic spread to

distant organs (Semenova et al., 2016). In a separate study, Denny and colleagues

profiled the chromatin accessibility landscape of primary tumors and liver metastases

from *Trp53/Rb1/p130* mice to identify differentially accessible regions between primary

tumors and liver metastases, and found that these regions were enriched for NFI family

binding motifs. Nfib was found to be highly expressed in invasive primary tumors and

metastases. Nfib knockdown reduced liver metastasis formation in a subcutaneous

transplant model of SCLC; conversely, Nfib overexpression increased liver metastasis

formation (Denny et al., 2016). Collectively, these data demonstrate that Nfib

expression promotes metastasis in SCLC.

Conclusion

Ever since the recognition of small cell lung cancer as a distinct form of lung cancer (Azzopardi, 1959), huge advances have been made in understanding the molecular basis of the disease. Despite this, treatment options for SCLC have not progressed at the same rate, as is true for many other types of cancer. The existing chemotherapeutic regimen for treating SCLC uses cytotoxic drugs that have been in use for decades – cisplatin and etoposide were first approved by the United States Food and Drug Administration (FDA) for the treatment of cancers in 1978 and 1983 respectively. The failure of any new drugs to show improved efficacy over the cisplatin/etoposide combination over this time serves to highlight the urgent need to develop better treatment options for SCLC.

The work presented in this thesis describes the use of the CRISPR-Cas9 system to advance our understanding of SCLC and thereby identify potential new approaches to target the disease. In Chapter 2, I describe our efforts to adapt the CRISPR-Cas9 system for modeling candidate tumor suppressor genes in mouse models of SCLC. In particular, we demonstrate that *p107*, which is mutated in a subset of human SCLC tumors, is a functional tumor suppressor in SCLC. In Chapter 3, I describe a CRISPR-mediated genetic screen that we have performed in murine SCLC cell lines to identify novel SCLC-specific genetic vulnerabilities. This has allowed us to identify novel metabolic vulnerabilities in SCLC, including enhanced sensitivity towards disruption of the pyrimidine biosynthesis pathway. Collectively, these experiments demonstrate the utility of the CRISPR-Cas9 system for carrying out functional interrogation of SCLC.

References

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B.T., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science *353*, aaf5573.

Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P., Han, S., Joung, J., Belanto, J.J., Verdine, V., Cox, D.B.T., Kellner, M.J., Regev, A., et al. (2017). RNA targeting with CRISPR-Cas13. Nature *550*, 280–284.

Aguirre, A.J., Meyers, R.M., Weir, B.A., Vazquez, F., Zhang, C.Z., Ben-David, U., Cook, A., Ha, G., Harrington, W.F., Doshi, M.B., et al. (2016). Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. Cancer Discov. *6*, 914–929.

Ames, B.N., Lee, F.D., and Durston, W.E. (1973b). An Improved Bacterial Test System for the Detection and Classification of Mutagens and Carcinogens. Proc. Natl. Acad. Sci. *70*, 782–786.

Ames, B.N., Durston, W.E., Yamasaki, E., and Lee, F.D. (1973a). Carcinogens are Mutagens: A Simple Test System Combining Liver Homogenates for Activation and Bacteria for Detection. Proc. Natl. Acad. Sci. *70*, 2281–2285.

Azzopardi, J.G. (1959). Oat-cell carcinoma of the bronchus. J. Pathol. Bacteriol. *78*, 513–519.

Baer, A., and Bode, J. (2001). Coping with kinetic and thermodynamic barriers: RMCE, an efficient strategy for the targeted integration of transgenes. Curr. Opin. Biotechnol. *12*, 473–480.

Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371–385.e18.

Baker, S.J., Fearon, E.R., Nigro, J.M., Hamilton, S.R., Preisinger, A.C., Jessup, J.M., VanTuinen, P., Ledbetter, D.H., Barker, D.F., Nakamura, Y., et al. (1989). Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. Science *244*, 217–221.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. a, and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Beijersbergen, R.L., Kerkhoven, R.M., Zhu, L., Carlee, L., Voorhoeve, P.M., and Bernards, R. (1994). E2F-4, a new member of the E2F gene family, has oncogenic activity and associates with p107 in vivo. Genes Dev. *8*, 2680–2690.

Berns, K., Hijmans, E.M., Mullenders, J., Brummelkamp, T.R., Velds, A., Heimerikx, M., Kerkhoven, R.M., Madiredjo, M., Nijkamp, W., Weigelt, B., et al. (2004). A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature *428*, 431–437.

Bieging, K.T., Mello, S.S., and Attardi, L.D. (2014). Unravelling mechanisms of p53-mediated tumour suppression. Nat. Rev. Cancer *14*, 359–370.

Bode, J., Schlake, T., Iber, M., Schübeler, D., Seibler, J., Snezhkov, E., and Nikolaev, L. (2000). The Transgeneticists Toolbox: Novel Methods for the Targeted Modification of Eukaryotic Genomes. Biol. Chem. *381*, 801–813.

Bolotin, A., Quinquis, B., Sorokin, A., and Dusko Ehrlich, S. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551–2561.

Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature *309*, 255–256.

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

Brinster, R.L., Chen, H.Y., Messing, A., van Dyke, T., Levine, A.J., and Palmiter, R.D. (1984). Transgenic mice harboring SV40 t-antigen genes develop characteristic brain tumors. Cell *37*, 367–379.

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E. V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960–964.

Brown, S.D.M., and Moore, M.W. (2012). Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. Dis. Model. Mech. *5*, 289–292.

Brummelkamp, T.R., Bernards, R., and Agami, R. (2002). A system for stable expression of short interfering RNAs in mammalian cells. Science *296*, 550–553.

Byers, L.A., and Rudin, C.M. (2015). Small cell lung cancer: Where do we go from here? Cancer *121*, 664–672.

Calbo, J., van Montfort, E., Proost, N., van Drunen, E., Beverloo, H.B., Meuwissen, R., and Berns, A. (2011). A Functional Role for Tumor Cell Heterogeneity in a Mouse Model of Small Cell Lung Cancer. Cancer Cell *19*, 244–256.

Califano, R., Abidin, A.Z., Peck, R., Faivre-Finn, C., and Lorigan, P. (2012). Management of Small Cell Lung Cancer. Drugs *72*, 471–490.

Calo, E., Quintero-Estades, J.A., Danielian, P.S., Nedelcu, S., Berman, S.D., and Lees, J.A. (2010). Rb regulates fate choice and lineage commitment in vivo. Nature *466*, 1110–1114.

Cancer Genome Atlas Research Network (2014). Comprehensive molecular profiling of lung adenocarcinoma. Nature *511*, 543–550.

Carette, J.E., Guimaraes, C.P., Varadarajan, M., Park, A.S., Wuethrich, I., Godarova, A., Kotecki, M., Cochran, B.H., Spooner, E., Ploegh, H.L., et al. (2009). Haploid genetic screens in human cells identify host factors used by pathogens. Science *326*, 1231–1235.

Cavenee, W.K., Dryja, T.P., Phillips, R.A., Benedict, W.F., Godbout, R., Gallie, B.L., Murphree, A.L., Strong, L.C., and White, R.L. (1983). Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. Nature *305*, 779–784.

Chin, L., Tam, A., Pomerantz, J., Wong, M., Holash, J., Bardeesy, N., Shen, Q., O'Hagan, R., Pantginis, J., Zhou, H., et al. (1999). Essential role for oncogenic Ras in tumour maintenance. Nature *400*, 468–472.

Chiou, S., Winters, I.P., Wang, J., Naranjo, S., Dudgeon, C., Tamburini, F.B., Brady, J.J., Yang, D., Grüner, B.M., Chuang, C., et al. (2015). Pancreatic cancer modeling using retrograde viral vector delivery and in vivo CRISPR/Cas9-mediated somatic genome editing. Genes Dev. *29*, 1576–1585.

Chipuk, J.E., Kuwana, T., Bouchier-Hayes, L., Droin, N.M., Newmeyer, D.D., Schuler, M., and Green, D.R. (2004). Direct activation of Bax by p53 mediates mitochondrial membrane permeabilization and apoptosis. Science *303*, 1010–1014.

Christophorou, M.A., Martin-Zanca, D., Soucek, L., Lawlor, E.R., Brown-Swigart, L., Verschuren, E.W., and Evan, G.I. (2005). Temporal dissection of p53 function in vitro and in vivo. Nat. Genet. *37*, 718–726.

Christophorou, M.A., Ringshausen, I., Finch, A.J., Swigart, L.B., and Evan, G.I. (2006). The pathological response to DNA damage does not contribute to p53-mediated tumour suppression. Nature *443*, 214–217.

Cobrinik, D., Whyte, P., Peeper, D.S., Jacks, T., and Weinberg, R. a (1993). Cell cycle-specific association of E2F with the p130 E1A-binding protein. Genes Dev. *7*, 2392–2404.

Cobrinik, D., Lee, M.H., Hannon, G., Mulligan, G., Bronson, R.T., Dyson, N., Harlow, E., Beach, D., Weinberg, R.A., and Jacks, T. (1996). Shared role of the pRB-related p130 and p107 proteins in limb development. Genes Dev. *10*, 1633–1644.

Collins, F.S., Lander, E.S., Rogers, J., and Waterson, R.H. (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931–945.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J., and Zhang, F. (2017). RNA editing with CRISPR-Cas13. Science *358*, 1019–1027.

Cui, M., Augert, A., Rongione, M., Conkrite, K., Parazzoli, S., Nikitin, A.Y., Ingolia, N., and MacPherson, D. (2014). PTEN Is a Potent Suppressor of Small Cell Lung Cancer. Mol. Cancer Res. *12*, 654–659.

Dahlman, J.E., Abudayyeh, O.O., Joung, J., Gootenberg, J.S., Zhang, F., and Konermann, S. (2015). Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. Nat. Biotechnol. *33*, 1159–1161.

Danielian, P.S., White, R., Hoare, S.A., Fawell, S.E., and Parker, M.G. (1993). Identification of residues in the estrogen receptor that confer differential sensitivity to estrogen and hydroxytamoxifen. Mol. Endocrinol. *7*, 232–240.

Danielian, P.S., Muccino, D., Rowitch, D.H., Michael, S.K., and McMahon, A.P. (1998). Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. Curr. Biol. *8*, 1323-S2.

Dannenberg, J.-H., Schuijff, L., Dekker, M., van der Valk, M., and te Riele, H. (2004). Tissue-specific tumor suppressor activity of retinoblastoma gene homologs p107 and p130. Genes Dev. *18*, 2952–2962.

Ben David, Y., Prideaux, V.R., Chow, V., Benchimol, S., and Bernstein, A. (1988). Inactivation of the p53 oncogene by internal deletion or retroviral integration in erythroleukemic cell lines induced by Friend leukemia virus. Oncogene *3*, 179–185.

DeLeo, A.B., Jay, G., Appella, E., Dubois, G.C., Law, L.W., and Old, L.J. (1979). Detection of a transformation-related antigen in chemically induced sarcomas and other transformed cells of the mouse. Proc. Natl. Acad. Sci. *76*, 2420–2424.

Demedts, I.K., Vermaelen, K.Y., and Van Meerbeeck, J.P. (2010). Treatment of extensive-stage small cell lung carcinoma: Current status and future prospects. Eur. Respir. J. *35*, 202–215.

Denny, S.K., Yang, D., Chuang, C.-H., Brady, J.J., Lim, J.S., Grüner, B.M., Chiou, S.-H., Schep, A.N., Baral, J., Hamard, C., et al. (2016). Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility. Cell *166*, 328–342.

Der, C.J., Krontiris, T.G., and Cooper, G.M. (1982). Transforming genes of human bladder and lung carcinoma cell lines are homologous to the ras genes of Harvey and Kirsten sarcoma viruses. Proc. Natl. Acad. Sci. *79*, 3637–3640.

Devoto, S.H., Mudryj, M., Pines, J., Hunter, T., and Nevins, J.R. (1992). A cyclin A-protein kinase complex possesses sequence-specific DNA binding activity: p33cdk2 is a component of the E2F-cyclin A complex. Cell *68*, 167–176.

Dimova, D.K., and Dyson, N.J. (2005). The E2F transcriptional network: old acquaintances with new faces. Oncogene *24*, 2810–2826.

Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature *330*, 576–578.

Donehower, L.A., Harvey, M., Slagle, B.L., McArthur, M.J., Montgomery, C.A., Butel, J.S., and Bradley, A. (1992). Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. Nature *356*, 215–221.

Dooley, A.L., Winslow, M.M., Chiang, D.Y., Banerji, S., Stransky, N., Dayton, T.L., Snyder, E.L., Senna, S., Whittaker, C.A., Bronson, R.T., et al. (2011). Nuclear factor I/B is an oncogene in small cell lung cancer. Genes Dev. *25*, 1470–1475.

Dow, L.E., Fisher, J., O'Rourke, K.P., Muley, A., Kastenhuber, E.R., Livshits, G., Tschaharganeh, D.F., Socci, N.D., and Lowe, S.W. (2015). Inducible in vivo genome editing with CRISPR-Cas9. Nat. Biotechnol. *33*, 390–394.

Dowdy, S.F. (2017). Overcoming cellular barriers for RNA therapeutics. Nat. Biotechnol. *35*, 222–229.

Dujon, B. (1996). The yeast genome project: what did we learn? Trends Genet. *12*, 263–270.

Dyson, N.J. (2016). RB1 : a prototype tumor suppressor and an enigma. Genes Dev. *30*, 1492–1502.

Dyson, N., Buchkovich, K., Whyte, P., and Harlow, E. (1989). The cellular 107K protein that binds to adenovirus E1A also associates with the large T antigens of SV40 and JC virus. Cell *58*, 249–255.

East-Seletsky, A., O'Connell, M.R., Knight, S.C., Burstein, D., Cate, J.H.D., Tjian, R., and Doudna, J.A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. Nature *538*, 270–273.

Eilers, M., Picard, D., Yamamoto, K.R., and Bishop, J.M. (1989). Chimaeras of Myc oncoprotein and steroid receptors cause hormone-dependent transformation of cells. Nature *340*, 66–68.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001a). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. *15*, 188–200.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001b). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494–498.

Eliyahu, D., Raz, A., Gruss, P., Givol, D., and Oren, M. (1984). Participation of p53 cellular tumour antigen in transformation of normal embryonic cells. Nature *312*, 646–649.

Eliyahu, D., Michalovitz, D., Eliyahu, S., Pinhasi-Kimhi, O., and Oren, M. (1989). Wild-type p53 can inhibit oncogene-mediated focus formation. Proc. Natl. Acad. Sci. *86*, 8763–8767.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156.

Evers, B., Jastrzebski, K., Heijmans, J.P.M., Grernrum, W., Beijersbergen, R.L., and Bernards, R. (2016). CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. Nat. Biotechnol. *34*, 631–633.

Ewen, M.E., Ludlow, J.W., Marsilio, E., DeCaprio, J.A., Millikan, R.C., Cheng, S.H., Paucha, E., and Livingston, D.M. (1989). An N-Terminal transformation-governing sequence of SV40 large T antigen contributes to the binding of both p110Rb and a second cellular protein, p120. Cell *58*, 257–267.

Ewen, M.E., Xing, Y., Lawrence, J.B., and Livingston, D.M. (1991). Molecular cloning, chromosomal mapping, and expression of the cDNA for p107, a retinoblastoma gene product-related protein. Cell *66*, 1155–1164.

Feil, R., Brocard, J., Mascrez, B., LeMeur, M., Metzger, D., and Chambon, P. (1996). Ligand-activated site-specific recombination in mice. Proc. Natl. Acad. Sci. *93*, 10887–10890.

Feil, R., Wagner, J., Metzger, D., and Chambon, P. (1997). Regulation of Cre Recombinase Activity by Mutated Estrogen Receptor Ligand-Binding Domains. Biochem. Biophys. Res. Commun. *237*, 752–757.

Felsher, D.W., and Bishop, J.M. (1999). Reversible Tumorigenesis by MYC in Hematopoietic Lineages. Mol. Cell *4*, 199–207.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer *136*, E359–E386.

Finlay, C.A., Hinds, P.W., and Levine, A.J. (1989). The p53 proto-oncogene can act as a suppressor of transformation. Cell *57*, 1083–1093.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806–811.

Flores, E.R., Sengupta, S., Miller, J.B., Newman, J.J., Bronson, R., Crowley, D., Yang, A., McKeon, F., and Jacks, T. (2005). Tumor predisposition in mice mutant for p63 and p73: Evidence for broader tumor suppressor functions for the p53 family. Cancer Cell *7*, 363–373.

Fodde, R., Edelmann, W., Yang, K., van Leeuwen, C., Carlson, C., Renault, B., Breukel, C., Alt, E., Lipkin, M., and Khan, P.M. (1994). A targeted chain-termination mutation in the mouse Apc gene results in multiple intestinal tumors. Proc. Natl. Acad. Sci. *91*, 8969–8973.

Fonfara, I., Richter, H., BratoviÄ, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature *532*, 517–521.

Freed-Pastor, W.A., and Prives, C. (2012). Mutant p53: one name, many proteins. Genes Dev. *26*, 1268–1286.

Frese, K.K., and Tuveson, D. a (2007). Maximizing mouse cancer models. Nat. Rev. Cancer *7*, 654–658.

Friend, S.H., Bernards, R., Rogelj, S., Weinberg, R.A., Rapaport, J.M., Albert, D.M., and Dryja, T.P. (1986). A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature *323*, 643–646.

Furth, P.A., St Onge, L., Boger, H., Gruss, P., Gossen, M., Kistner, A., Bujard, H., and Hennighausen, L. (1994). Temporal control of gene expression in transgenic mice by a tetracycline-responsive promoter. Proc. Natl. Acad. Sci. *91*, 9302–9306.

Garneau, J.E., Dupuis, M.-È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67–71.

George, J., Lim, J.S., Jang, S.J., Cun, Y., Ozretić, L., Kong, G., Leenders, F., Lu, X., Fernández-Cuesta, L., Bosco, G., et al. (2015). Comprehensive genomic profiles of small cell lung cancer. Nature *524*, 47–53.

Giaever, G., and Nislow, C. (2014). The yeast deletion collection: A decade of functional genomics. Genetics *197*, 451–465.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387–391.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell *154*, 442–451.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Ginsberg, D., Vairo, G., Chittenden, T., Xiao, Z.X., Xu, G., Wydner, K.L., DeCaprio, J.A., Lawrence, J.B., and Livingston, D.M. (1994). E2F-4, a new member of the E2F transcription factor family, interacts with p107. Genes Dev. *8*, 2665–2679.

Goldfarb, M., Shimizu, K., Perucho, M., and Wigler, M. (1982). Isolation and preliminary characterization of a human transforming gene from T24 bladder carcinoma cells. Nature *296*, 404–409.

Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc. Natl. Acad. Sci. *89*, 5547–5551.

Gossen, M., Freundlieb, S., Bender, G., Müller, G., Hillen, W., and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. Science *268*, 1766–1769.

Govindan, R., Page, N., Morgensztern, D., Read, W., Tierney, R., Vlahiotis, A., Spitznagel, E.L., and Piccirillo, J. (2006). Changing epidemiology of small-cell lung cancer in the United States over the last 30 years: analysis of the surveillance, epidemiologic, and end results database. J. Clin. Oncol. *24*, 4539–4544.

Gu, H., Marth, J.D., Orban, P.C., Mossmann, H., and Rajewsky, K. (1994). Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. Science *265*, 103–106.

Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V.C., Graveley, B.R., Terns, R.M., et al. (2012). Essential Features and Rational Design of CRISPR RNAs that Function with the Cas RAMP Module Complex to Cleave RNAs. Mol. Cell *45*, 292–302.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950–952.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature *404*, 293–296.

Hanahan, D. (1985). Heritable formation of pancreatic beta-cell tumours in transgenic mice expressing recombinant insulin/simian virus 40 oncogenes. Nature *315*, 115–122.

Hanahan, D., and Weinberg, R.A. (2000). The Hallmarks of Cancer. Cell *100*, 57–70.

Hannon, G.J., Demetrick, D., and Beach, D. (1993). Isolation of the Rb-related p130 through its interaction with CDK2 and cyclins. Genes Dev. *7*, 2378–2391.

Harlow, E., Whyte, P., Franza, B.R., and Schley, C. (1986). Association of adenovirus early-region 1A proteins with cellular polypeptides. Mol. Cell. Biol. *6*, 1579–1589.

Harris, H. (1971). Cell fusion and the analysis of malignancy. Proc. R. Soc. B Biol. Sci. *179*, 1–20.

Hart, T., Brown, K.R., Sircoulomb, F., Rottapel, R., and Moffat, J. (2014). Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. Mol. Syst. Biol. *10*, 733–733.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell *163*, 1515–1526.

Hart, T., Tong, A.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A., et al. (2017). Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. G3 (Bethesda). *7*, 2719–2727.

Hellmann, M.D., Ott, P.A., Zugazagoitia, J., Ready, N.E., Hann, C.L., Braud, F.G. De, Antonia, S.J., Ascierto, P.A., Moreno, V., Atmaca, A., et al. (2017). Nivolumab (nivo) ± ipilimumab (ipi) in advanced small-cell lung cancer (SCLC): First report of a randomized expansion cohort from CheckMate 032. J. Clin. Oncol. *35*, 8503.

Hijmans, E.M., Voorhoeve, P.M., Beijersbergen, R.L., van 't Veer, L.J., and Bernards, R. (1995). E2F-5, a new E2F family member that interacts with p130 in vivo. Mol. Cell. Biol. *15*, 3082–3089.

Hilgendorf, K.I., Leshchiner, E.S., Nedelcu, S., Maynard, M.A., Calo, E., Ianari, A., Walensky, L.D., and Lees, J.A. (2013). The retinoblastoma protein induces apoptosis directly at the mitochondria. Genes Dev. *27*, 1003–1015.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Hingorani, S.R., Wang, L., Multani, A.S., Combs, C., Deramaudt, T.B., Hruban, R.H., Rustgi, A.K., Chang, S., and Tuveson, D.A. (2005). Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. Cancer Cell *7*, 469–483.

Hinnen, A., Hicks, J.B., and Fink, G.R. (1978). Transformation of yeast. Proc. Natl. Acad. Sci. *75*, 1929–1933.

Hisada, M., Garber, J.E., Li, F.P., Fung, C.Y., and Fraumeni, J.F. (1998). Multiple Primary Cancers in Families With Li-Fraumeni Syndrome. JNCI J. Natl. Cancer Inst. *90*, 606–611.

Holliday, R. (1964). A mechanism for gene conversion in fungi. Genet. Res. *5*, 282.

Howlader, N., Noone, A., Krapcho, M., Miller, D., Bishop, K., Kosary, C., Yu, M., Ruhl, J., Tatalovich, Z., Mariotto, A., et al. SEER Cancer Statistics Review, 1975-2014, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2014/, based on November 2016 SEER data submission, posted to the SEER web site, April 2017.

Huang, J., Chen, M., Whitley, M.J., Kuo, H., Xu, E.S., Walens, A., Mowery, Y.M., Van Mater, D., Eward, W.C., Cardona, D.M., et al. (2017). Generation and comparison of CRISPR-Cas9 and Cre-mediated genetically engineered mouse models of sarcoma. Nat. Commun. *8*, 15999.

Huijbers, I.J., Krimpenfort, P., Berns, A., and Jonkers, J. (2011). Rapid validation of cancer genes in chimeras derived from established genetically engineered mouse models. BioEssays *33*, 701–710.

Huijbers, I.J., Bin Ali, R., Pritchard, C., Cozijnsen, M., Kwon, M.-C.C., Proost, N., Song, J.-Y.Y., de Vries, H., Badhai, J., Sutherland, K., et al. (2014). Rapid target gene validation in complex cancer mouse models using re-derived embryonic stem cells. EMBO Mol. Med. *6*, 212–225.

Ianari, A., Natale, T., Calo, E., Ferretti, E., Alesse, E., Screpanti, I., Haigis, K., Gulino, A., and Lees, J.A. (2009). Proapoptotic Function of the Retinoblastoma Tumor Suppressor Protein. Cancer Cell *15*, 184–194.

Indra, A.K., Warot, X., Brocard, J., Bornert, J.-M., Xiao, J.-H., Chambon, P., and Metzger, D. (1999). Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ERT and Cre-ERT2 recombinases. Nucleic Acids Res. *27*, 4324–4327.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J. Bacteriol. *169*, 5429–5433.

Jacks, T., Fazeli, A., Schmitt, E.M., Bronson, R.T., Goodell, M.A., and Weinberg, R.A. (1992). Effects of an Rb mutation in the mouse. Nature *359*, 295–300.

Jacks, T., Remington, L., Williams, B.O., Schmitt, E.M., Halachmi, S., Bronson, R.T., and Weinberg, R.A. (1994a). Tumor spectrum analysis in p53-mutant mice. Curr. Biol. *4*, 1–7.

Jacks, T., Shih, T.S., Schmitt, E.M., Bronson, R.T., Bernards, A., and Weinberg, R. a (1994b). Tumour predisposition in mice heterozygous for a targeted mutation in Nf1. Nat. Genet. *7*, 353–361.

Jackson, E.L., Olive, K.P., Tuveson, D.A., Bronson, R., Crowley, D., Brown, M., and Jacks, T. (2005). The differential effects of mutant p53 alleles on advanced murine lung cancer. Cancer Res. *65*, 10280–10288.

Jaenisch, R. (1976). Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. Proc. Natl. Acad. Sci. *73*, 1260–1264.

Jaenisch, R., and Mintz, B. (1974). Simian Virus 40 DNA Sequences in DNA of Healthy Adult Mice Derived from Preimplantation Blastocysts Injected with Viral DNA. Proc. Natl. Acad. Sci. *71*, 1250–1254.

Jahchan, N.S., Lim, J.S., Bola, B., Morris, K., Seitz, G., Tran, K.Q., Xu, L., Trapani, F., Morrow, C.J., Cristea, S., et al. (2016). Identification and Targeting of Long-Term Tumor-Propagating Cells in Small Cell Lung Cancer. Cell Rep. *16*, 644–656.

Jansen, R., Embden, J.D.A. van, Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. *43*, 1565–1575.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. Elife *2*, e00471.

Jonkers, J., Meuwissen, R., van der Gulden, H., Peterse, H., van der Valk, M., and Berns, A. (2001). Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. Nat. Genet. *29*, 418–425.

Junttila, M.R., Karnezis, A.N., Garcia, D., Madriles, F., Kortlever, R.M., Rostker, F., Brown Swigart, L., Pham, D.M., Seo, Y., Evan, G.I., et al. (2010). Selective activation of p53-mediated tumour suppression in high-grade tumours. Nature *468*, 567–571.

Kaelin, W.G. (2005). The concept of synthetic lethality in the context of anticancer therapy. Nat. Rev. Cancer *5*, 689–698.

Kaelin, W.G. (2012). Molecular biology. Use and abuse of RNAi to study mammalian gene function. Science *337*, 421–422.

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

Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat. Methods *12*, 401–403.

Kim, D., Kim, J., Hur, J.K., Been, K.W., Yoon, S.H., and Kim, J.S. (2016). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nat. Biotechnol. *34*, 863–868.

Kinzler, K.W., and Vogelstein, B. (1996). Lessons from Hereditary Colorectal Cancer. Cell *87*, 159–170.

Kirsch, D.G., Dinulescu, D.M., Miller, J.B., Grimm, J., Santiago, P.M., Young, N.P., Nielsen, G.P., Quade, B.J., Chaber, C.J., Schultz, C.P., et al. (2007). A spatially and temporally restricted mouse model of soft tissue sarcoma. Nat. Med. *13*, 992–997.

Kistner, A., Gossen, M., Zimmermann, F., Jerecic, J., Ullmer, C., Lubbert, H., and Bujard, H. (1996). Doxycycline-mediated quantitative and tissue-specific control of gene expression in transgenic mice. Proc. Natl. Acad. Sci. *93*, 10933–10938.

Kleinstiver, B.P., Tsai, S.Q., Prew, M.S., Nguyen, N.T., Welch, M.M., Lopez, J.M., McCaw, Z.R., Aryee, M.J., and Joung, J.K. (2016). Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nat. Biotechnol. *34*, 869–874.

Knudson, A.G. (1971). Mutation and Cancer: Statistical Study of Retinoblastoma. Proc. Natl. Acad. Sci. *68*, 820–823.

Koller, B.H., Hagemann, L.J., Doetschman, T., Hagaman, J.R., Huang, S., Williams, P.J., First, N.L., Maeda, N., and Smithies, O. (1989). Germ-line transmission of a planned alteration made in a hypoxanthine phosphoribosyltransferase gene by homologous recombination in embryonic stem cells. Proc. Natl. Acad. Sci. *86*, 8927–8931.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420–424.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

Koonin, E. V., Makarova, K.S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. *37*, 67–78.

Kotecki, M., Reddy, P.S., and Cochran, B.H. (1999). Isolation and characterization of a near-haploid human cell line. Exp. Cell Res. *252*, 273–280.

Kwon, M.C., Proost, N., Song, J.Y., Sutherland, K.D., Zevenhoven, J., and Berns, A. (2015). Paracrine signaling between tumor subclones of mouse sclc: A critical role of ets transcription factor pea3 in facilitating metastasis. Genes Dev. *29*, 1587–1592.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853–858.

Lai, S.L., Brauch, H., Knutsen, T., Johnson, B.E., Nau, M.M., Mitsudomi, T., Tsai, C.M., Whang-Peng, J., Zbar, B., and Kaye, F.J. (1995). Molecular genetic characterization of neuroendocrine lung cancer cell lines. Anticancer Res. *15*, 225–232.

Lakso, M., Sauer, B., Mosinger, B., Lee, E.J., Manning, R.W., Yu, S.H., Mulder, K.L., and Westphal, H. (1992). Targeted oncogene activation by site-specific recombination in transgenic mice. Proc. Natl. Acad. Sci. *89*, 6232–6236.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Lane, D.P. (1992). p53, guardian of the genome. Nature *358*, 15–16.

Lane, D.P., and Crawford, L. V (1979). T antigen is bound to a host protein in SV40-transformed cells. Nature *278*, 261–263.

Lang, G.A., Iwakuma, T., Suh, Y.A., Liu, G., Rao, V.A., Parant, J.M., Valentin-Vega, Y.A., Terzian, T., Caldwell, L.C., Strong, L.C., et al. (2004). Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. Cell *119*, 861–872.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858–862.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G. V, Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

LeCouter, J.E., Kablar, B., Whyte, P.F., Ying, C., and Rudnicki, M.A. (1998a). Strain-dependent embryonic lethality in mice lacking the retinoblastoma-related p130 gene. Development *125*, 4669–4679.

LeCouter, J.E., Kablar, B., Hardy, W.R., Ying, C., Megeney, L.A., May, L.L., and Rudnicki, M.A. (1998b). Strain-dependent myeloid hyperplasia, growth deficiency, and accelerated cell cycle in mice lacking the Rb-related p107 gene. Mol. Cell. Biol. *18*, 7455–7465.

Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. Science *294*, 862–864.

Lee, E.Y.-H.P., Chang, C.-Y., Hu, N., Wang, Y.-C.J., Lai, C.-C., Herrup, K., Lee, W.-H., and Bradley, A. (1992). Mice deficient for Rb are nonviable and show defects in neurogenesis and haematopoiesis. Nature *359*, 288–294.

Lee, M.H., Williams, B.O., Mulligan, G., Mukai, S., Bronson, R.T., Dyson, N., Harlow, E., and Jacks, T. (1996). Targeted disruption of p107: Functional overlap between p107 and Rb. Genes Dev. *10*, 1621–1632.

Li, Y., Graham, C., Lacy, S., Duncan, A.M.V., and Whyte, P. (1993). The adenovirus E1A-associated 130-kD protein is encoded by a member of the retinoblastoma gene family and physically interacts with cyclins A and E. Genes Dev. *7*, 2366–2377.

Liao, H.-K., Hatanaka, F., Araoka, T., Reddy, P., Wu, M.-Z., Sui, Y., Yamauchi, T., Sakurai, M., O'Keefe, D.D., Núñez-Delicado, E., et al. (2017). In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans -epigenetic Modulation. Cell 1–13.

Lim, J.S., Ibaseta, A., Fischer, M.M., Cancilla, B., O'Young, G., Cristea, S., Luca, V.C., Yang, D., Jahchan, N.S., Hamard, C., et al. (2017). Intratumoural heterogeneity generated by Notch signalling promotes small-cell lung cancer. Nature *545*, 360–364.

Linzer, D.I.H., and Levine, A.J. (1979). Characterization of a 54K Dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. Cell *17*, 43–52.

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233–247.e17.

MacPherson, D., Sage, J., Kim, T., Ho, D., McLaughlin, M.E., and Jacks, T. (2004). Cell type-specific effects of Rb deletion in the murine retina. Genes Dev. *18*, 1681–1694.

MacPherson, D., Conkrite, K., Tam, M., Mukai, S., Mu, D., and Jacks, T. (2007). Murine bilateral retinoblastoma exhibiting rapid-onset, metastatic progression and N-myc gene amplification. EMBO J. *26*, 784–794.

Maddalo, D., Manchado, E., Concepcion, C.P., Bonetti, C., Vidigal, J.A., Han, Y.-C., Ogrodowski, P., Crippa, A., Rekhtman, N., de Stanchina, E., et al. (2014). In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. Nature *516*, 423–427.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013). CRISPR RNA-guided activation of endogenous human genes. Nat. Methods *10*, 977–979.

Makarova, K.S., Grishin, N. V., Shabalina, S.A., Wolf, Y.I., and Koonin, E. V. (2006). No Title. Biol. Direct *1*, 7.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Malkin, D., Li, F.P., Strong, L.C., Fraumeni, J.F., Nelson, C.E., Kim, D.H., Kassel, J., Gryka, M.A., Bischoff, F.Z., and Tainsky, M.A. (1990). Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science *250*, 1233–1238.

Mansour, S.L., Thomas, K.R., and Capecchi, M.R. (1988). Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature *336*, 348–352.

Martin, G.R., and Evans, M.J. (1975). Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. Proc. Natl. Acad. Sci. *72*, 1441–1445.

Martins, C.P., Brown-Swigart, L., and Evan, G.I. (2006). Modeling the Therapeutic Efficacy of p53 Restoration in Tumors. Cell *127*, 1323–1334.

Marumoto, T., Tashiro, A., Friedmann-Morvinski, D., Scadeng, M., Soda, Y., Gage, F.H., and Verma, I.M. (2009). Development of a novel mouse glioma model using lentiviral vectors. Nat. Med. *15*, 110–116.

McDonald, E.R., de Weck, A., Schlabach, M.R., Billy, E., Mavrakis, K.J., Hoffman, G.R., Belur, D., Castelletti, D., Frias, E., Gampa, K., et al. (2017). Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. Cell *170*, 577–592.e10.

McDonald, J.I., Celik, H., Rois, L.E., Fishberger, G., Fowler, T., Rees, R., Kramer, A., Martens, A., Edwards, J.R., and Challen, G.A. (2016). Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. Biol. Open *5*, 866–874.

McFadden, D.G., Papagiannakopoulos, T., Taylor-Weiner, A., Stewart, C., Carter, S.L., Cibulskis, K., Bhutkar, A., McKenna, A., Dooley, A., Vernon, A., et al. (2014). Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. Cell *156*, 1298–1311.

Meselson, M.S., and Radding, C.M. (1975). A general model for genetic recombination. Proc. Natl. Acad. Sci. *72*, 358–361.

Meuwissen, R., Linn, S.C., Linnoila, R.I., Zevenhoven, J., Mooi, W.J., and Berns, A. (2003). Induction of small cell lung cancer by somatic inactivation of both Trp53 and Rb1 in a conditional mouse model. Cancer Cell *4*, 181–189.

Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N. V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat. Genet. *49*, 1779–1784.

Mihara, M., Erster, S., Zaika, A., Petrenko, O., Chittenden, T., Pancoska, P., and Moll, U.M. (2003). p53 Has a Direct Apoptogenic Role at the Mitochondria. Mol. Cell *11*, 577–590.

Mills, A.A., Zheng, B., Wang, X.-J., Vogel, H., Roop, D.R., and Bradley, A. (1999). p63 is a p53 homologue required for limb and epidermal morphogenesis. Nature *398*, 708–713.

Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K., et al. (2006). A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen. Cell *124*, 1283–1298.

Mojica, F.J.M., Diez-Villasenor, C., Soria, E., and Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. Mol. Microbiol. *36*, 244–246.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. *60*, 174–182.

Mollaoglu, G., Guthrie, M.R., Böhm, S., Brägelmann, J., Can, I., Ballieu, P.M., Marx, A., George, J., Heinen, C., Chalishazar, M.D., et al. (2017). MYC Drives Progression of Small Cell Lung Cancer to a Variant Neuroendocrine Subtype with Vulnerability to Aurora Kinase Inhibition. Cancer Cell 1–16.

Momand, J., Zambetti, G.P., Olson, D.C., George, D., and Levine, A.J. (1992). The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. Cell *69*, 1237–1245.

Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., Sakai, A., Nakashima, H., Hata, K., Nakashima, K., et al. (2016). Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. Nat. Biotechnol. *34*, 1060–1065.

Mowat, M., Cheng, A., Kimura, N., Bernstein, A., and Benchimol, S. (1985). Rearrangements of the cellular p53 gene in erythroleukaemic cells transformed by Friend virus. Nature *314*, 633–636.

Munoz, D.M., Cassiani, P.J., Li, L., Billy, E., Korn, J.M., Jones, M.D., Golji, J., Ruddy, D.A., Yu, K., McAllister, G., et al. (2016). CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. Cancer Discov. *6*, 900–913.

Nicolay, B.N., and Dyson, N.J. (2013). The multiple connections between pRB and cell metabolism. Curr. Opin. Cell Biol. *25*, 735–740.

Nicolay, B.N., Danielian, P.S., Kottakis, F., Lapek, J.D., Sanidas, I., Miles, W.O., Dehnad, M., Tschöp, K., Gierut, J.J., Manning, A.L., et al. (2015). Proteomic analysis of pRb loss highlights a signature of decreased mitochondrial oxidative phosphorylation. Genes Dev. *29*, 1875–1889.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science *353*, aaf8729-aaf8729.

Olive, K.P., Tuveson, D.A., Ruhe, Z.C., Yin, B., Willis, N.A., Bronson, R.T., Crowley, D., and Jacks, T. (2004). Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. Cell *119*, 847–860.

Onuki, N., Wistuba, I.I., Travis, W.D., Virmani, A.K., Yashima, K., Brambilla, E., Hasleton, P., and Gazdar, A.F. (1999). Genetic changes in the spectrum of neuroendocrine lung tumors. Cancer *85*, 600–607.

Orban, P.C., Chui, D., and Marth, J.D. (1992). Tissue- and site-specific DNA recombination in transgenic mice. Proc. Natl. Acad. Sci. *89*, 6861–6865.

Orr-Weaver, T.L., Szostak, J.W., and Rothstein, R.J. (1981). Yeast transformation: a model system for the study of recombination. Proc. Natl. Acad. Sci. *78*, 6354–6358.

Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J., and Conklin, D.S. (2002). Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. Genes Dev. *16*, 948–958.

Paddison, P.J., Silva, J.M., Conklin, D.S., Schlabach, M., Li, M., Aruleba, S., Balija, V., O'Shaughnessy, A., Gnoj, L., Scoble, K., et al. (2004). A resource for large-scale RNA-interference-based screens in mammals. Nature *428*, 427–431.

Palmiter, R.D., Chen, H.Y., Messing, A., and Brinster, R.L. (1985). SV40 enhancer and large-T antigen are instrumental in development of choroid plexus tumours in transgenic mice. Nature *316*, 457–460.

Papaioannou, V.E., McBurney, M.W., Gardner, R.L., and Evans, M.J. (1975). Fate of teratocarcinoma cells injected into early mouse embryos. Nature *258*, 70–73.

Parada, L.F., Tabin, C.J., Shih, C., and Weinberg, R.A. (1982). Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma virus ras gene. Nature *297*, 474–478.

Parada, L.F., Land, H., Weinberg, R.A., Wolf, D., and Rotter, V. (1984). Cooperation between gene encoding p53 tumour antigen and ras in cellular transformation. Nature *312*, 649–651.

Paul, C.P., Good, P.D., Winer, I., and Engelke, D.R. (2002). Effective expression of small interfering RNA in human cells. Nat. Biotechnol. *20*, 505–508.

Pelengaris, S., Littlewood, T., Khan, M., Elia, G., and Evan, G. (1999). Reversible activation of c-Myc in skin: Induction of a complex neoplastic phenotype by a single oncogenic lesion. Mol. Cell *3*, 565–577.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., et al. (2013). RNA-guided gene activation by CRISPR-Cas9–based transcription factors. Nat. Methods *10*, 973–976.

Perucho, M., Goldfarb, M., Shimizu, K., Lama, C., Fogh, J., and Wigler, M. (1981). Human-tumor-derived cell lines contain common and different transforming genes. Cell *27*, 467–476.

Picard, D., Salser, S.J., and Yamamoto, K.R. (1988). A movable and regulable inactivation function within the steroid binding domain of the glucocorticoid receptor. Cell *54*, 1073–1080.

Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M., et al. (2014). CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Cell *159*, 440–455.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology *151*, 653–663.

Pulciani, S., Santos, E., Lauver, a V, Long, L.K., Robbins, K.C., and Barbacid, M. (1982). Oncogenes in human tumor cell lines: molecular cloning of a transforming gene from human bladder carcinoma cells. Proc. Natl. Acad. Sci. *79*, 2845–2849.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell *152*, 1173–1183.

Reck, M., Heigener, D., and Reinmuth, N. (2016). Immunotherapy for small-cell lung cancer: emerging evidence. Futur. Oncol. *12*, 931–943.

Reddy, E.P., Reynolds, R.K., Santos, E., and Barbacid, M. (1982). A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. Nature *300*, 149–152.

Renan, M.J. (1993). How many mutations are required for tumorigenesis? Implications from human cancer data. Mol. Carcinog. *7*, 139–146.

Robanus-Maandag, E., Dekker, M., Van Der Valk, M., Carrozza, M.L., Jeanny, J.C., Dannenberg, J.H., Berns, A., and Riele, H. Te (1998). p107 is a suppressor of retinoblastoma development in pRB-deficient mice. Genes Dev. *12*, 1599–1609.

Roper, J., Tammela, T., Cetinbas, N.M., Akkad, A., Roghanian, A., Rickelt, S., Almeqdadi, M., Wu, K., Oberli, M.A., Sánchez-Rivera, F., et al. (2017). In vivo genome editing and organoid transplantation models of colorectal cancer and metastasis. Nat. Biotechnol. *35*, 569–576.

Rous, P. (1911). A sarcoma of the fowl transmissible by an agent separable from the tumor cells. J. Exp. Med. *13*, 397–411.

Sage, J., Miller, A.L., Pérez-Mancera, P.A., Wysocki, J.M., and Jacks, T. (2003). Acute mutation of retinoblastoma gene function is sufficient for cell cycle re-entry. Nature *424*, 223–228.

Sánchez-Rivera, F.J., and Jacks, T. (2015). Applications of the CRISPR–Cas9 system in cancer biology. Nat. Rev. Cancer *15*, 387–395.

Sánchez-Rivera, F.J., Papagiannakopoulos, T., Romero, R., Tammela, T., Bauer, M.R., Bhutkar, A., Joshi, N.S., Subbaraj, L., Bronson, R.T., Xue, W., et al. (2014). Rapid modelling of cooperating genetic events in cancer through somatic genome editing. Nature *516*, 428–431.

Santarpia, M., Daffinà, M.G., Karachaliou, N., González-Cao, M., Lazzari, C., Altavilla, G., and Rosell, R. (2016). Targeted drugs in small-cell lung cancer. Transl. Lung Cancer Res. *5*, 51–70.

Santos, E., Tronick, S.R., Aaronson, S.A., Pulciani, S., and Barbacid, M. (1982). T24 human bladder carcinoma oncogene is an activated form of the normal human homologue of BALB- and Harvey-MSV transforming genes. Nature *298*, 343–347.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res. *39*, 9275–9282.

Schaffer, B.E., Park, K.S., Yiu, G., Conklin, J.F., Lin, C., Burkhart, D.L., Karnezis, A.N., Sweet-Cordero, E.A., and Sage, J. (2010). Loss of p130 Accelerates Tumor Development in a Mouse Model for Human Small-Cell Lung Carcinoma. Cancer Res. *70*, 3877–3883.

Schwenk, F., Kühn, R., Angrand, P.O., Rajewsky, K., and Stewart, A.F. (1998). Temporally and spatially regulated somatic mutagenesis in mice. Nucleic Acids Res. *26*, 1427–1432.

Sears, R., Nuckolls, F., Haura, E., Taya, Y., Tamai, K., and Nevins, J.R. (2000). Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. Genes Dev. *14*, 2501–2514.

Semenova, E.A., Kwon, M., Monkhorst, K., Song, J.-Y., Bhaskaran, R., Krijgsman, O., Kuilman, T., Peters, D., Buikhuisen, W.A., Smit, E.F., et al. (2016). Transcription Factor NFIB Is a Driver of Small Cell Lung Cancer Progression in Mice and Marks Metastatic Disease in Patients. Cell Rep. *16*, 631–643.

Shibata, H., Toyama, K., Shioya, H., Ito, M., Hirota, M., Hasegawa, S., Matsumoto, H., Takano, H., Akiyama, T., Toyoshima, K., et al. (1997). Rapid colorectal adenoma formation initiated by conditional targeting of the Apc gene. Science *278*, 120–123.

Shih, C., and Weinberg, R.A. (1982). Isolation of a transforming sequence from a human bladder carcinoma cell line. Cell *29*, 161–169.

Shih, C., Padhy, L.C., Murray, M., and Weinberg, R.A. (1981). Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts. Nature *290*, 261–264.

Shirodkar, S., Ewen, M., DeCaprio, J.A., Morgan, J., Livingston, D.M., and Chittenden, T. (1992). The transcription factor E2F interacts with the retinoblastoma product and a p107-cyclin A complex in a cell cycle-regulated manner. Cell *68*, 157–166.

Siegel, R.L., Miller, K.D., and Jemal, A. (2018). Cancer statistics, 2018. CA. Cancer J. Clin. *68*, 7–30.

Silva, J.M., Li, M.Z., Chang, K., Ge, W., Golding, M.C., Rickles, R.J., Siolas, D., Hu, G., Paddison, P.J., Schlabach, M.R., et al. (2005). Second-generation shRNA libraries covering the mouse and human genomes. Nat. Genet. *37*, 1281–1288.

Simpson, D.S., Mason-Richie, N.A., Gettler, C.A., and Wikenheiser-Brokamp, K.A. (2009). Retinoblastoma family proteins have distinct functions in pulmonary epithelial cells in vivo critical for suppressing cell growth and tumorigenesis. Cancer Res. *69*, 8733–8741.

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO J. *30*, 1335–1342.

Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A., and Kucherlapati, R.S. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature *317*, 230–234.

Spector, D.H., Varmus, H.E., and Bishop, J.M. (1978). Nucleotide sequences related to the transforming gene of avian sarcoma virus are present in DNA of uninfected vertebrates. Proc. Natl. Acad. Sci. *75*, 4102–4106.

Srivastava, S., Zou, Z., Pirollo, K., Blattner, W., and Chang, E.H. (1990). Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li–Fraumeni syndrome. Nature *348*, 747–749.

Stehelin, D., Varmus, H.E., Bishop, J.M., and Vogt, P.K. (1976). DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. Nature *260*, 170–173.

Stewart, T.A., Pattengale, P.K., and Leder, P. (1984). Spontaneous mammary adenocarcinomas in transgenic mice that carry and express MTV/myc fusion genes. Cell *38*, 627–637.

Sutherland, K.D., Proost, N., Brouns, I., Adriaensen, D., Song, J.Y., and Berns, A. (2011). Cell of origin of small cell lung cancer: Inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. Cancer Cell *19*, 754–764.

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J., and Stahl, F.W. (1983). The double-strand-break repair model for recombination. Cell *33*, 25–35.

Tabin, C.J., Bradley, S.M., Bargmann, C.I., Weinberg, R.A., Papageorge, A.G., Scolnick, E.M., Dhar, R., Lowy, D.R., and Chang, E.H. (1982). Mechanism of activation of a human oncogene. Nature *300*, 143–149.

Tak, Y.E., Kleinstiver, B.P., Nuñez, J.K., Hsu, J.Y., Horng, J.E., Gong, J., Weissman, J.S., and Joung, J.K. (2017). Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. Nat. Methods *14*, 1163–1166.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. Cell *159*, 635–646.

Taparowsky, E., Suard, Y., Fasano, O., Shimizu, K., Goldfarb, M., and Wigler, M. (1982). Activation of the T24 bladder carcinoma transforming gene is linked to a single amino acid change. Nature *300*, 762–765.

Thomas, K.R., Folger, K.R., and Capecchi, M.R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. Cell *44*, 419–428.

Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L., and Melton, D.W. (1989). Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. Cell *56*, 313–321.

Travis, W.D., Brambilla, E., Müller-Hermelink, H.K., and Harris, C.C. (2004). World Health Organization Classification of Tumours: Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart. IARC Press Lyon.

Tschida, B.R., Largaespada, D.A., and Keng, V.W. (2014). Mouse models of cancer: Sleeping Beauty transposons for insertional mutagenesis screens and reverse genetic studies. Semin. Cell Dev. Biol. *27*, 86–95.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. Cell *170*, 564–576.e16.

Vairo, G., Livingston, D.M., and Ginsberg, D. (1995). Functional interaction between E2F-4 and p130: evidence for distinct mechanisms underlying growth suppression by different retinoblastoma protein family members. Genes Dev. *9*, 869–881.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. Science *291*, 1304–1351.

Vogelstein, B., and Kinzler, K.W. (1993). The multistep nature of cancer. Trends Genet. *9*, 138–141.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558.

Vojta, A., Dobrinic, P., Tadic, V., Bockor, L., Korac, P., Julg, B., Klasic, M., and Zoldos, V. (2016). Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res. *44*, 5615–5628.

Vooijs, M., te Riele, H., van der Valk, M., and Berns, A. (2002). Tumor formation in mice with somatic inactivation of the retinoblastoma gene in interphotoreceptor retinol binding protein-expressing cells. Oncogene *21*, 4635–4645.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910–918.

Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. Science *350*, 1096–1101.

Wang, T., Yu, H., Hughes, N.W., Liu, B., Kendirli, A., Klein, K., Chen, W.W., Lander, E.S., and Sabatini, D.M. (2017). Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. Cell *168*, 890–903.e15.

Wangensteen, K.J., Wilber, A., Keng, V.W., He, Z., Matise, I., Wangensteen, L., Carson, C.M., Chen, Y., Steer, C.J., McIvor, R.S., et al. (2008). A facile method for somatic, lifelong manipulation of multiple genes in the mouse liver. Hepatology *47*, 1714–1724.

Wangensteen, K.J., Wang, Y.J., Dou, Z., Wang, A.W., Mosleh-Shirazi, E., Horlbeck, M.A., Gilbert, L.A., Weissman, J.S., Berger, S.L., and Kaestner, K.H. (2017). Combinatorial genetics in liver repopulation and carcinogenesis with a novel in vivo CRISPR activation platform. Hepatology . Accepted Author Manuscript. doi:10.1002/hep.2962.

Wiesner, S.M., Decker, S.A., Larson, J.D., Ericson, K., Forster, C., Gallardo, J.L., Long, C., Demorest, Z.L., Zamora, E.A., Low, W.C., et al. (2009). De novo induction of genetically engineered brain tumors in mice using plasmid DNA. Cancer Res. *69*, 431–439.

Williams, B.O., Remington, L., Albert, D.M., Mukai, S., Bronson, R.T., and Jacks, T. (1994). Cooperative tumorigenic effects of germline mutations in Rb and p53. Nat. Genet. *7*, 480–484.

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., et al. (1999). Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science *285*, 901–906.

Wistuba, I., Gazdar, A., and Minna, J. (2001). Molecular genetics of small cell lung carcinoma. Semin. Oncol. *28*, 3–13.

Wolf, D., and Rotter, V. (1984). Inactivation of p53 gene expression by an insertion of Moloney murine leukemia virus-like DNA sequences. Mol. Cell. Biol. *4*, 1402–1410.

Wolf, D., and Rotter, V. (1985). Major deletions in the gene encoding the p53 tumor antigen cause lack of p53 expression in HL-60 cells. Cell Biol. *82*, 790–794.

Xu, X., Tao, Y., Gao, X., Zhang, L., Li, X., Zou, W., Ruan, K., Wang, F., Xu, G., and Hu, R. (2016). A CRISPR-based approach for targeted DNA demethylation. Cell Discov. *2*, 16009.

Xue, W., Chen, S., Yin, H., Tammela, T., Papagiannakopoulos, T., Joshi, N.S., Cai, W., Yang, G., Bronson, R., Crowley, D.G., et al. (2014). CRISPR-mediated direct mutation of cancer genes in the mouse liver. Nature *514*, 380–384.

Yamagiwa, K., and Ichikawa, K. (1918). Experimental Study of the Pathogenesis of Carcinoma. J. Cancer Res. *3*, 1–29.

Yang, A., Schweitzer, R., Sun, D., Kaghad, M., Walker, N., Bronson, R.T., Tabin, C., Sharpe, A., Caput, D., Crum, C., et al. (1999). p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. Nature *398*, 714–718.

Yang, A., Walker, N., Bronson, R., Kaghad, M., Oosterwegel, M., Bonnin, J., Vagner, C., Bonnet, H., Dikkes, P., Sharpe, A., et al. (2000). p73-deficient mice have neurological, pheromonal and inflammatory defects but lack spontaneous tumours. Nature *404*, 99–103.

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., and Jaenisch, R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell *154*, 1370–1379.

Yang, X., Boehm, J.S., Yang, X., Salehi-Ashtiani, K., Hao, T., Shen, Y., Lubonja, R., Thomas, S.R., Alkan, O., Bhimdi, T., et al. (2011). A public genome-scale lentiviral expression library of human ORFs. Nat. Methods *8*, 659–661.

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S., et al. (2015). Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. Cell *160*, 339–350.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell *101*, 25–33.

Zender, L., Xue, W., Zuber, J., Semighini, C.P., Krasnitz, A., Ma, B., Zender, P., Kubicka, S., Luk, J.M., Schirmacher, P., et al. (2008). An Oncogenomics-Based In Vivo RNAi Screen Identifies Tumor Suppressors in Liver Cancer. Cell *135*, 852–864.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., Van Der Oost, J., Regev, A., et al. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell *163*, 759–771.

Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.-V., et al. (2012). Structure and Mechanism of the CMR Complex for CRISPR-Mediated Antiviral Immunity. Mol. Cell *45*, 303–313.

Zhou, Y., Rideout, W.M., Zi, T., Bressel, A., Reddypalli, S., Rancourt, R., Woo, J.-K., Horner, J.W., Chin, L., Chiu, M.I., et al. (2010). Chimeric mouse tumor models reveal differences in pathway activation between ERBB family– and KRAS-dependent lung adenocarcinomas. Nat. Biotechnol. *28*, 71–78.

CHAPTER 2

CRISPR-mediated modeling and functional validation of candidate tumor suppressor genes in small cell lung cancer

Sheng Rong Ng[1,2], William M. Rideout III[1], Elliot H. Akama-Garren[1], Roderick T. Bronson[3] and Tyler Jacks[1,2,4]

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

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

[3] Department of Pathology, Tufts University School of Veterinary Medicine, North Grafton, MA 01536

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

S.R.N. and T.J. designed the study; S.R.N performed all experiments except for the following: W.M.R. performed embryonic stem cell targeting experiments, and E.H.A-G. cloned the Rosa26-CAGGS-LSL-Cas9-GFP-Csy4 allele; R.T.B. performed histopathological analyses. All experiments were performed in the laboratory of T.J.

ABSTRACT

Small cell lung cancer (SCLC) is a highly aggressive subtype of lung cancer that remains among the most lethal of solid tumor malignancies. Recent genomic sequencing studies have identified multiple recurrently mutated genes in human SCLC tumors. However, the functional roles of most of these genes remain to be validated. The CRISPR-Cas9 system has been utilized in genetically engineered mouse models of cancer to generate somatic genetic alterations in tumor cells *in vivo*, greatly accelerating efforts to study the roles of candidate genes of interest in various cancers. Here, we have adapted the CRISPR-Cas9 system to a well-established murine model of SCLC. We generated a Cre-activated allele of Cas9 and crossed it into the *Trp53/Rb1* double knockout model of SCLC. Using adenoviral vectors co-expressing Cre recombinase and a target sgRNA, we showed that loss of p107 (Rbl1), which is mutated in a significant fraction of human SCLC tumors, resulted in acceleration of tumor progression in the model, to a similar extent as loss of p130 (Rbl2). We observed notable differences between the effects of loss of p107 and p130, with loss of p107 resulting in fewer initiated tumors but larger mean tumor size, as well as the earlier appearance of metastatic spread, compared with loss of p130. We also took advantage of unique patterns of CRISPR-induced mutations in each tumor to infer lineage relationships between primary and metastatic tumors in the same animal. In summary, we have demonstrated the feasibility of using the CRISPR-Cas9 system to rapidly model loss-of-function mutations in candidate tumor suppressor genes in SCLC, and we anticipate that this approach will greatly speed up efforts to investigate mechanisms driving progression of this deadly disease.

INTRODUCTION

Small cell lung cancer (SCLC) is a highly aggressive neuroendocrine lung carcinoma that comprises around 13-15% of all diagnosed lung cancer cases (Govindan et al., 2006). The disease is characterized by rapid growth and early, widespread metastasis, with the majority of patients presenting with extensive stage disease (Califano et al., 2012). Although SCLC patients often exhibit robust initial responses to cytotoxic chemotherapy, relapse almost invariably occurs, and no effective second-line therapies currently exist (Demedts et al., 2010). Despite decades of research, no new therapies have demonstrated efficacy in SCLC patients, in contrast to the growing number of targeted therapies available for treating non-small cell lung cancer (Byers and Rudin, 2015).

Genetically engineered mouse models (GEMMs) of SCLC have been used extensively to study the molecular mechanisms of tumor progression in SCLC. Based on the fact that inactivating mutations in the tumor suppressor genes *TP53* and *RB1* are found in almost all human SCLC tumors (George et al., 2015; Wistuba et al., 2001), a murine model of SCLC (mSCLC) was developed by conditionally deleting *Trp53* and *Rb1* in the murine lung epithelium (Meuwissen et al., 2003). This model faithfully recapitulates the key features of human SCLC, including histopathological appearance, expression of key neuroendocrine markers, and pattern of metastatic spread (Meuwissen et al., 2003). Subsequent studies have utilized the *Trp53/Rb1* double knockout model of SCLC to functionally investigate additional genes, such as *Rbl2* (a.k.a. *p130*), *Pten*, *Mycl1*, *Nfib* and *Myc* (Cui et al., 2014; Denny et al., 2016; Dooley et

al., 2011; Huijbers et al., 2014; McFadden et al., 2014; Mollaoglu et al., 2017; Schaffer et al., 2010; Semenova et al., 2016; Wu et al., 2016).

Large-scale cancer genome sequencing studies have generated an extensive catalog of genes that are mutated in numerous cancer types (Vogelstein et al., 2013). It remains a significant challenge to distinguish between driver versus passenger mutations, to identify genes or pathways that are truly important for tumor progression. This is particularly relevant in cancers that have high mutations rates, such as lung cancer (George et al., 2015; Lawrence et al., 2013; Peifer et al., 2012; Rudin et al., 2012). One recent study involving SCLC identified multiple recurrently altered genes in these tumors, including inactivating mutations in the Notch signaling pathway, which was subsequently shown to functionally contribute to SCLC tumor progression (George et al., 2015), However, apart from a few other notable examples, many of the most frequently mutated genes have yet to be functionally validated in SCLC.

The development of the CRISPR-Cas9 system for genome editing in mammalian cells (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013) has revolutionized the field of cancer research, allowing the rapid validation of candidate oncogenes and tumor suppressor genes both *in vitro* as well as *in vivo*. This has been especially powerful when combined with GEMMs of various cancers (Chiou et al., 2015; Dow et al., 2015; Huang et al., 2017; Maddalo et al., 2014; Platt et al., 2014; Roper et al., 2017; Sánchez-Rivera et al., 2014; Xue et al., 2014). By bypassing the need to generate new germline or conditional alleles for each gene of interest, the CRISPR-Cas9 system has vastly increased the speed at which candidate genes, such as those identified from cancer genome sequencing studies or genetic screens, can be functionally validated in relevant

preclinical models of cancer. These systems also streamline the development of *in vivo* models with which to examine the biological effects of multiple tumor-associated mutations.

In this study, we have adapted the CRISPR-Cas9 system to the *Trp53/Rb1* model of SCLC. We demonstrate the utility of this system to rapidly model loss of function of candidate tumor suppressor genes in SCLC. In particular, we show that loss of p107 (a.k.a. Rbl1), a member of the retinoblastoma family of proteins that is recurrently mutated in a subset of human SCLC tumors (George et al., 2015), significantly accelerates tumor progression in the *Trp53/Rb1*-mutant background. Notably, loss of p107 appears to have differential effects compared with loss of p130, another member of the retinoblastoma family. We also demonstrate the feasibility of using the CRISPR-Cas9 system to infer lineage relationships between primary and metastatic tumors in the same animal.

RESULTS

Strategy for *in vivo* CRISPR-mediated targeting of genes in mSCLC

We generated a Cre-activated allele of Cas9 targeted to the *Rosa26* locus (**Fig. 1A**). This allele also co-expresses EGFP and Csy4 (Cas6f); the latter is a Type I CRISPR-Cas endonuclease that recognizes and cleaves RNA at a 28-nucleotide sequence (Haurwitz et al., 2010), and has previously been used for multiplexed sgRNA expression from a single RNA transcript (Nissim et al., 2014). We crossed this allele into the *Trp53*^{flox/flox}; *Rb1*^{flox/flox} background to generate Trp53/Rb1/Cas9 animals. To enable monitoring of tumor progression *in vivo*, we also crossed a Cre-activated luciferase reporter allele into these animals (Dooley et al., 2011; McFadden et al., 2014).

To restrict CRISPR-Cas9 activity specifically to initiated tumor cells *in vivo*, we generated an adenoviral vector, Ad5-USEC, that expresses an sgRNA targeting a gene of interest together with Cre recombinase (**Fig. 1B**). Cre activity in the vector was validated *in vitro* using the Green-Go reporter cell line previously generated in our laboratory, in which GFP is activated upon Cre expression (Sánchez-Rivera et al., 2014) (**Fig. 1C**). SCLC tumors were initiated by intratracheal administration of Ad5-USEC into the lungs of animals (DuPage et al., 2009).

Loss of p107 accelerates tumor progression in SCLC

To validate this system, we chose to target *p107* and *p130*, both of which are members of the retinoblastoma family of proteins that are recurrently mutated in around 6% of human SCLC tumors each (George et al., 2015). In particular, *p130* was chosen as a positive control because germline p130 conditional alleles have been previously
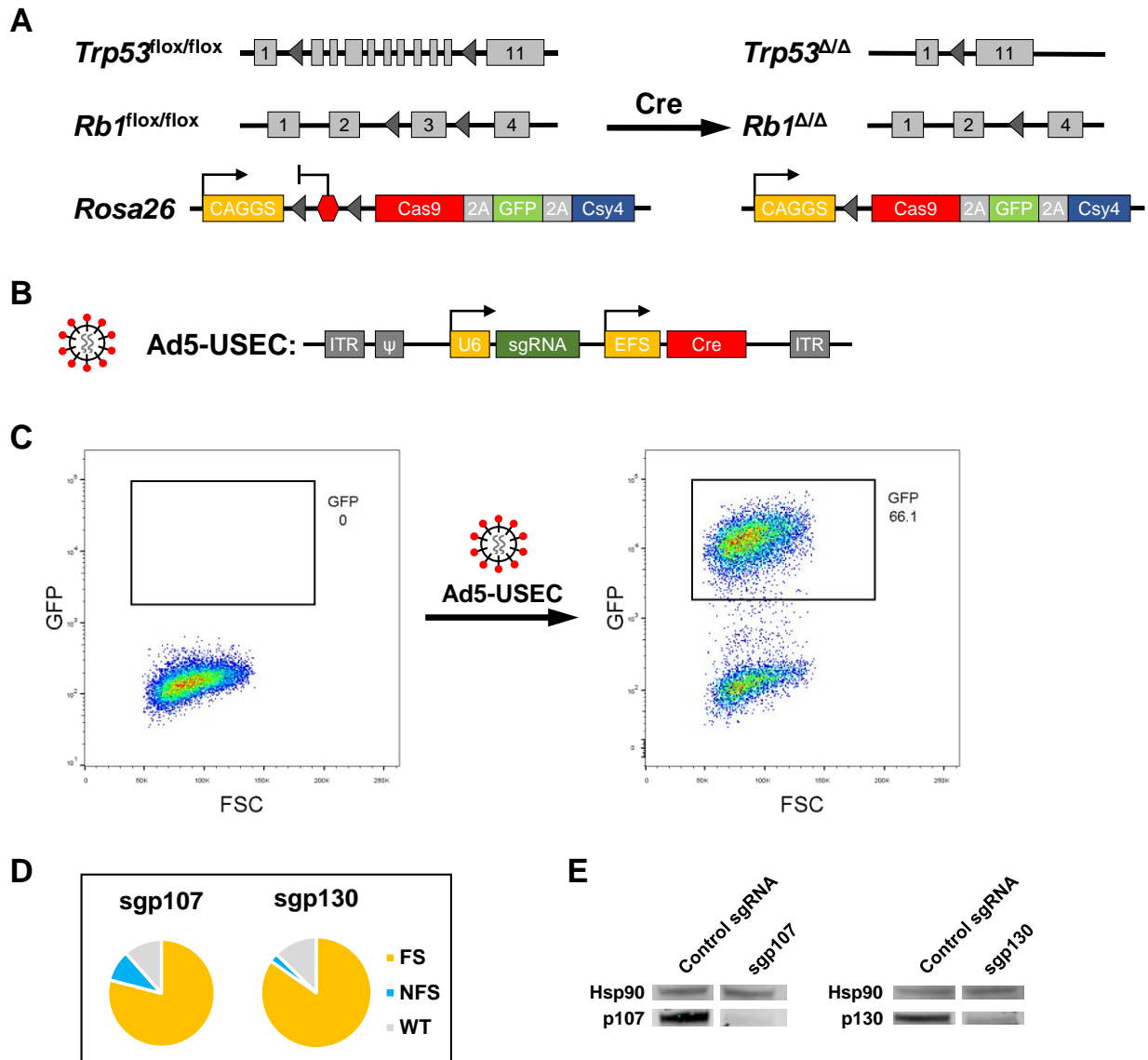
Figure 1: Strategy for *in vivo* CRISPR-mediated targeting of genes in mSCLC.

(A) Schematic of the Rosa26-CAGGS-LSL-Cas9-GFP-Csy4 allele crossed into the *Trp53/Rb1* double knockout model of SCLC.

(B) Schematic of the Ad5-USEC vector.

(C) *In vitro* validation of the Ad5-USEC vector in the Green-Go reporter cell line, which harbors a Cre-activated GFP cassette.

(D) Deep sequencing of *p107* and *p130* genomic loci, respectively, from Green-Go-Cas9 cells infected with Ad5-USEC-sgp107 or Ad5-USEC-sgp130. FS: frameshift mutation; NFS: non-frameshift mutation; WT: wild-type allele.

(E) Western blot analysis of p107 and p130 protein levels in Green-Go-Cas9 cells infected with Ad5-USEC-sgp107 or Ad5-USEC-sgp130.

used to accelerate tumor progression in SCLC (Schaffer et al., 2010). We designed

sgRNAs targeting *p107* and *p130*, and validated their activity *in vitro* in Green-Go cells

that were transduced with a Cas9-expressing lentivirus (Sanjana et al., 2014), both by

deep sequencing of the respective target genomic loci to assess the efficiency of

generation of mutations (**Fig. 1D**), as well as by Western blot to confirm the decrease in

protein levels (**Fig. 1E**).

To test our system *in vivo*, we infected Trp53/Rb1/Cas9 animals with Ad5-USEC

vectors expressing sgRNAs targeting *p107*, *p130*, or a control unannotated region of the

genome (sgp107, sgp130, control sg, respectively). 5.5 months post-tumor initiation, we

performed *in vivo* bioluminescence imaging to track tumor progression in these animals.

We observed significantly higher luciferase activity in both sgp130 and sgp107-infected

animals compared with control animals (**Fig. 2A-B**), consistent with a significant

acceleration in tumor progression. Furthermore, both sgp130 (232 days) and sgp107-

infected animals (205 days) showed significantly reduced median survival compared

with control animals (267 days; **Fig. 2C**). The acceleration of tumor progression that we

observed in sgp130-infected animals recapitulates the results obtained in a previous

study using conditional *Trp53/Rb1/p130* triple knockout SCLC mice (Schaffer et al.,

2010). Importantly, the tumors that developed in all animals retained the characteristic

histological features of SCLC, including positive staining for the neuroendocrine marker

CGRP (data not shown), confirming that the acceleration of tumor progression was not

a result of a change in the tumor spectrum.

To determine whether the observed phenotypes were a result of loss of *p130* or

p107 gene function, we isolated genomic DNA from tumors dissected from infected
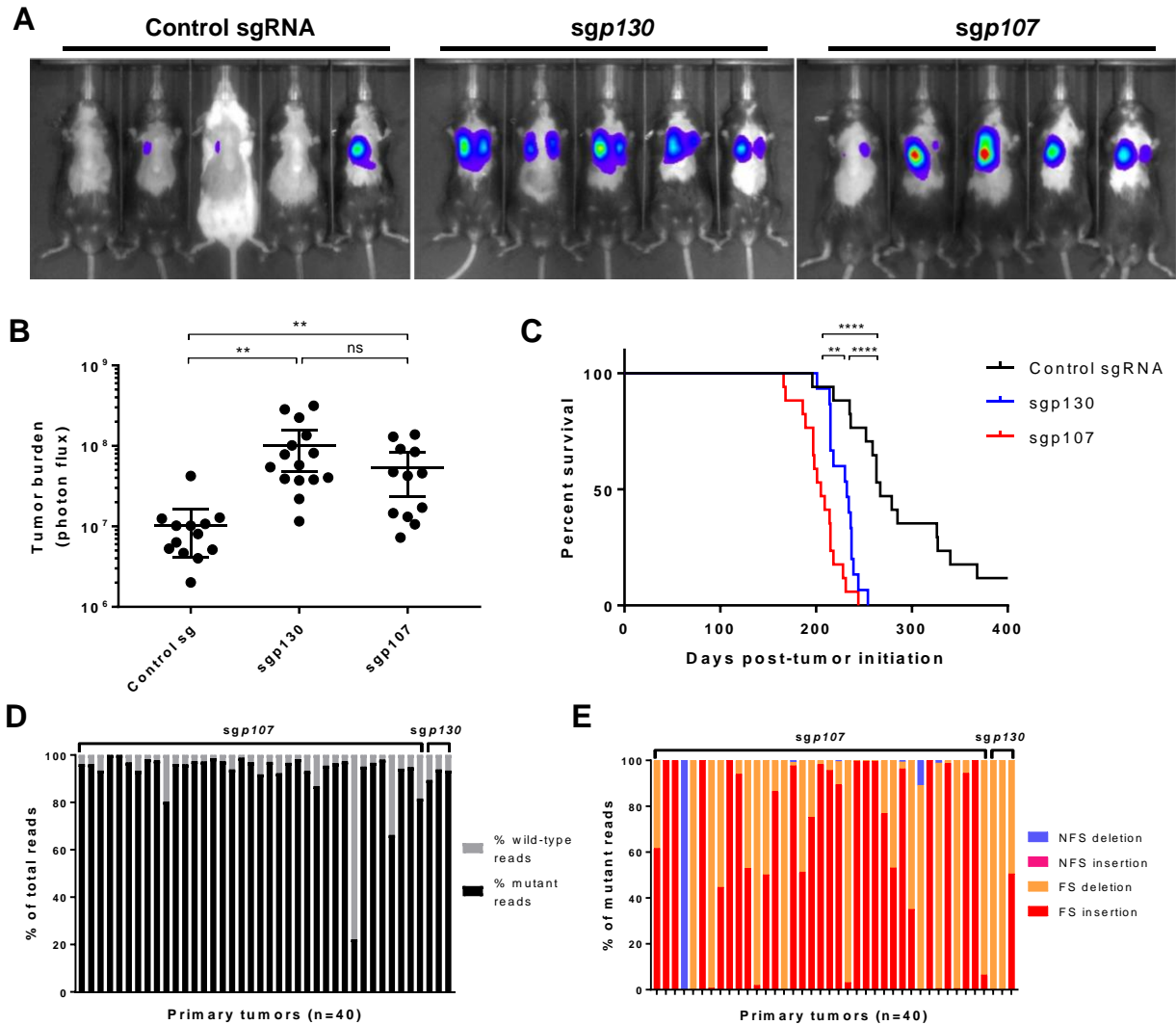
Figure 2: Loss of p107 accelerates tumor progression in SCLC.

(A) Representative images from *in vivo* bioluminescence imaging of Trp53/Rb1/Cas9 animals infected with Ad5-USEC harboring the indicated sgRNAs. Animals were imaged approximately 5.5 months post-tumor initiation.

(B) Quantification of tumor burden in Trp53/Rb1/Cas9 animals infected Ad5-USEC harboring control sgRNA (n = 13), sgp130 (n = 15) or sgp107 (n = 13) by *in vivo* bioluminescence imaging. Data are presented as means, with error bars denoting the 95% confidence interval. ** $p<0.005$, ns: not significant, two-tailed Student's *t*-test.

(C) Survival analysis for Trp53/Rb1/Cas9 animals infected with control sgRNA (n = 17), sgp130 (n = 15) or sgp107 (n = 17). ** $p<0.005$, **** $p<0.0001$, log-rank (Mantel-Cox) test.

(D) Deep sequencing of *p107* or *p130* genomic loci from tumors isolated from Trp53/Rb1/Cas9 animals infected with the respective sgRNAs, showing the proportion of mutant and wild-type sequencing reads, as well as the proportion of each indicated type of mutation for each tumor. FS: frameshift, NFS: non-frameshift.

animals, then performed targeted deep sequencing of the genomic loci targeted by the

respective sgRNAs. We observed that the vast majority of detected sequences

contained either frameshift insertions or deletions (**Fig. 2D-E, and Supplementary Fig.**

1), with tumors harboring 1-4 different mutant alleles each. This result is consistent with

a strong positive selection pressure for loss-of-function alleles in these tumors.

Collectively, these data validate our approach for modeling loss-of-function mutations in

this model, and demonstrate that p107, like p130, is a functional tumor suppressor in

SCLC.

Differential effects of loss of p107 and p130 on tumor progression

To more closely examine changes in tumor progression upon loss of p107 or

p130, we sacrificed a cohort of infected Trp53/Rb1/Cas9 mice 6 months post-tumor

initiation for histopathologic analysis. Analysis of hematoxylin and eosin (H&E) stained

lung sections from these animals showed that loss of either p107 or p130 significantly

increases overall lung tumor burden compared with control animals (**Fig. 3B**),

consistent with the earlier *in vivo* bioluminescence imaging data. In addition, there was

no significant difference in tumor burden between sgp107 and sgp130-infected animals.

However, there was a marked difference in the tumor phenotype observed in both

groups of animals. sgp107-infected animals developed significantly fewer tumors, but

with a larger mean tumor size, compared with sgp130-infected animals (**Fig. 3A, C-D**).

Notably, sgp107-infected animals also displayed a greater incidence of mediastinal

lymph node metastasis compared with sgp130-infected animals at this time point (**Fig.**

3E). This finding is consistent with the observation that sgp107-infected animals showed
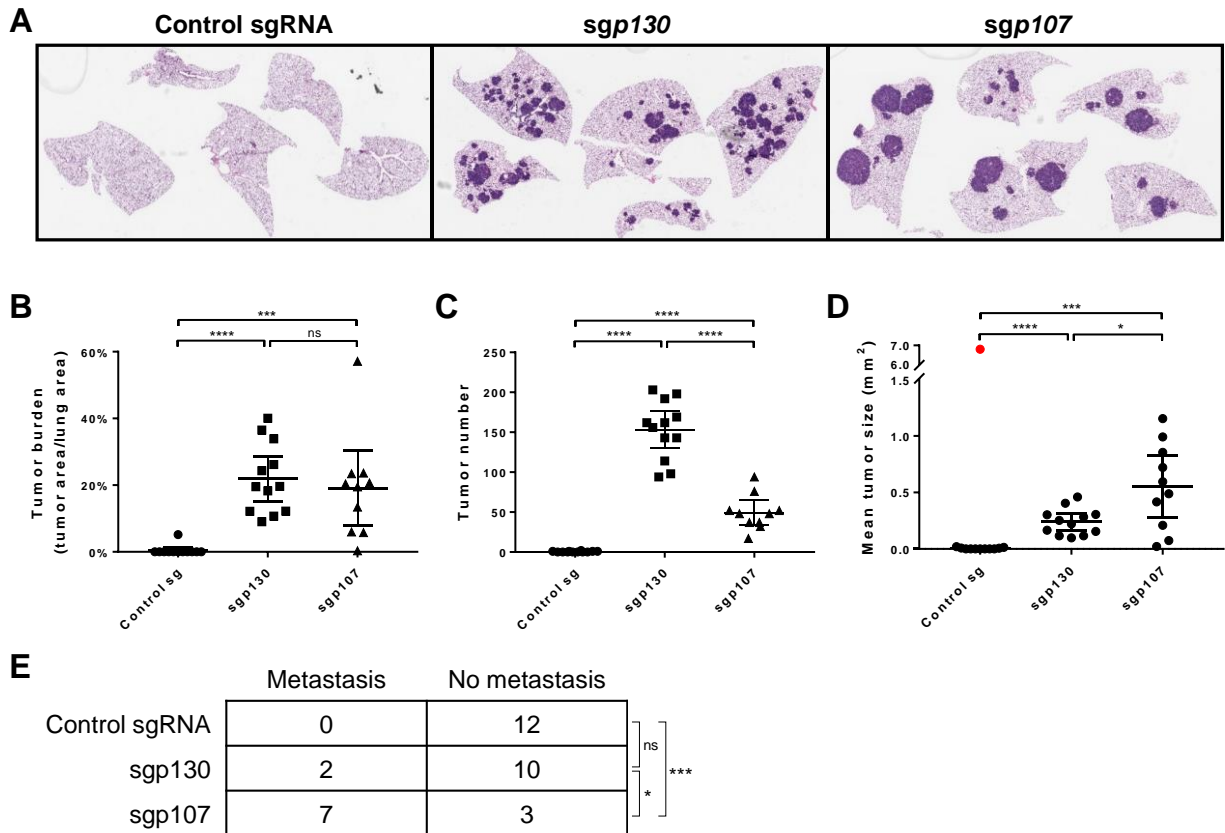
Figure 3: Differential effects of loss of p107 and p130 on tumor progression.

(A) Representative hematoxylin and eosin (H&E) sections of Trp53/Rb1/Cas9 animals infected with Ad5-USEC harboring the indicated sgRNAs. Animals were sacrificed 6 months post-tumor initiation.

(B-D) Quantification of tumor burden **(B)**, tumor number **(C)** and mean tumor size **(D)** in Trp53/Rb1/Cas9 animals infected with Ad5-USEC harboring control sgRNA (n = 12), sgp130 (n = 12) or sgp107 (n = 10), 6 months post-tumor initiation. For **(D)**, red indicates a data point that was identified as an outlier (Grubbs' test, α<0.001), and excluded from subsequent analyses. Data are presented as means, with error bars denoting the 95% confidence interval. ns: not significant, * p<0.05, *** p<0.001, **** p<0.0001, two-tailed Student's *t*-test.

(E) Quantification of number of animals presenting with mediastinal lymph node metastasis 6 months post-tumor initiation. ns: not significant, * p<0.05, *** p<0.001, two-sided Fisher's exact test.

slightly reduced median survival compared with sgp130-infected animals (**Fig. 2C**), as the mediastinal lymph node metastases accelerated the onset of breathing difficulties in these animals. Collectively, these data demonstrate that loss of p107 appears to have differential effects on SCLC development compared with loss of p130.

Loss of p107 in the *Trp53/Rb1/p130*-null background does not accelerate tumor progression

Because of the differential effects of p107 and p130 loss on tumor progression, we decide to investigate whether loss of both p107 and p130 would affect tumor progression. To do this, we crossed the Cas9 allele into the *Trp53*$^{flox/flox}$; *Rb1*$^{flox/flox}$; *p130*$^{flox/flox}$ background to generate Trp53/Rb1/p130/Cas9 animals (**Fig. 4A**), then infected these animals with Ad5-USEC vectors expressing sgp107 or control sgRNA. We observed no significant difference in survival between sgp107-infected animals and control animals, with the same median survival of 174 days for both groups (**Fig. 4B**). Consistent with this, analysis of tumor burden at 5 months post-tumor initiation showed no significant difference between sgp107-infected animals and control animals (**Fig. 4C**). Therefore, loss of both p107 and p130 does not appear to accelerate SCLC tumor progression over loss of p130 alone.

Inferring lineage relationships between primary and metastatic tumors

Because the majority of primary tumors sequenced harbored a combination of different mutant alleles, we sought to determine whether this could be used as a marker to infer lineage relationships between primary and metastatic tumors from the same
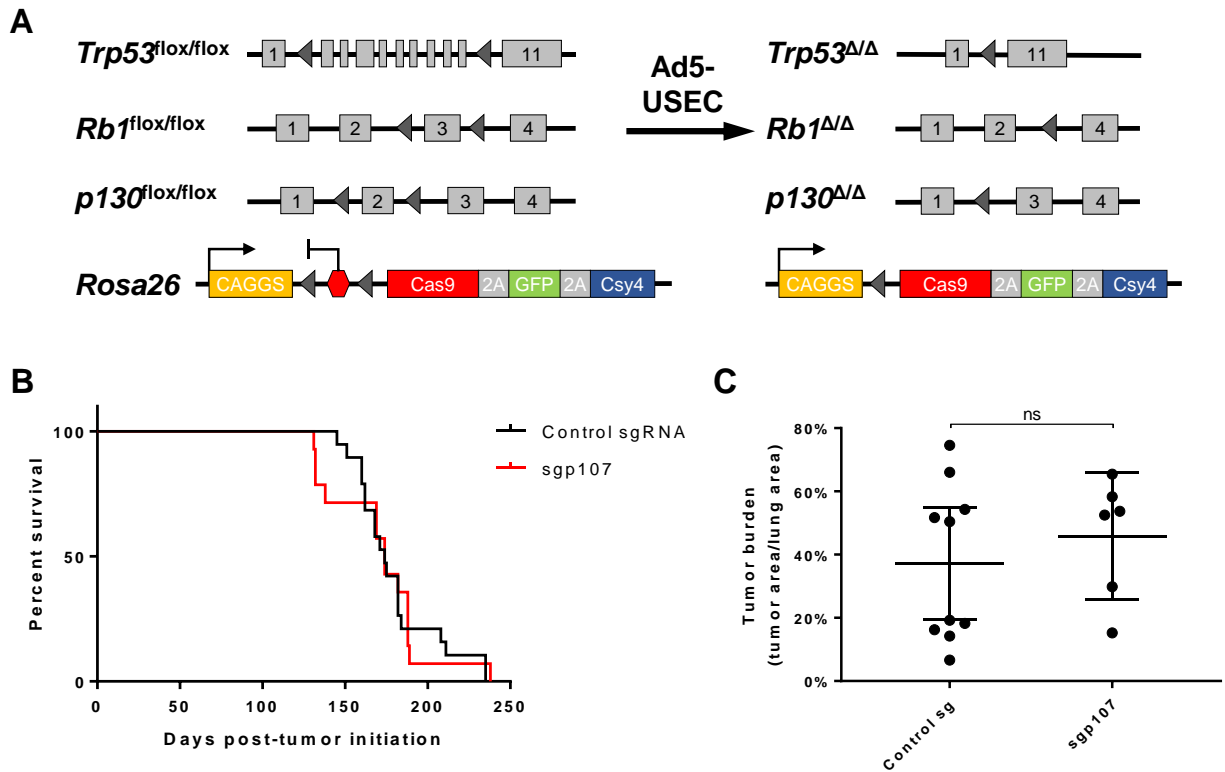
Figure 4: Loss of p107 in the *Trp53/Rb1/p130*-mutant background does not accelerate tumor progression.

(A) Schematic of the Rosa26-CAGGS-LSL-Cas9-GFP-Csy4 allele crossed into the *Trp53/Rb1/p130* triple knockout model of SCLC.

(B) Survival analysis for Trp53/Rb1/p130/Cas9 animals infected with control sgRNA (n = 19) or sgp107 (n = 14).

(C) Quantification of tumor burden in Trp53/Rb1/p130/Cas9 animals infected with Ad5-USEC harboring control sgRNA (n = 10) or sgp107 (n = 6), 5 months post-tumor initiation. Data are presented as means, with error bars denoting the 95% confidence interval. ns: not significant, two-tailed Student's *t*-test.

animal. In animals with a small number of tumors and mutant alleles, this was relatively straightforward to interpret with little ambiguity. For example, in animal 5759, two primary tumors (T1 and T2) and three independent liver metastases (C1, C2, C3) were sequenced, with two different alleles detected across all five tumors (**Fig. 5A and Supplementary Fig. 2A**). 5759-T1 harbored both alleles at an approximately 1:1 ratio, while 5759-T2 harbored only the first allele. All three liver metastases also harbored a 1:1 ratio of both alleles, suggesting that all three tumors were seeded from 5759-T1 rather than 5759-T2.

In animals with more complex allelic combinations, interpretation of lineage relationships was less definitive, but certain trends could nevertheless be inferred. In animal 5634, six primary tumors (T1-T6), one lymph node metastasis (LN) and one liver metastasis (C1) were sequenced, with 10 different alleles detected in total (**Fig. 5B and Supplementary Fig. 2B**). Both 5634-T1 and 5634-C1 harbored alleles A and B at a 1:1 ratio. Because no other sequenced primary tumor harbored these two alleles at a significant fraction, 5634-C1 was most likely seeded from 5634-T1. Given that only the largest primary tumors were sequenced, we cannot formally rule out the possibility that other smaller primary tumors that were not sequenced might also have harbored alleles A and B and seeded the liver metastasis. However, the nearly identical ratio of alleles in both the primary tumor and the liver metastasis in this case suggests that they are likely to be related.

In addition, 5634-LN also harbored alleles A and B at a 1:1 ratio, suggesting seeding from 5634-T1 as well. However, these two alleles comprised only 60% of sequencing reads detected from the tumor, with the remaining 40% comprised of a

110
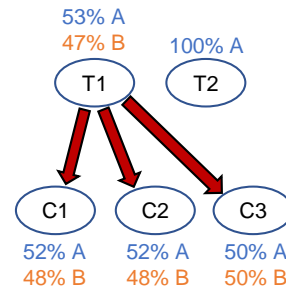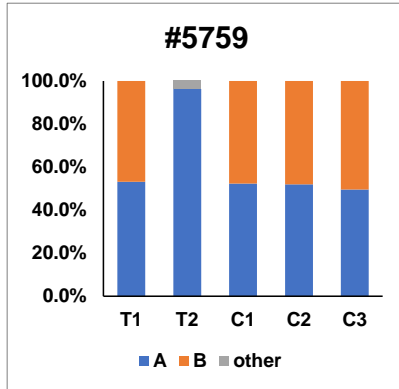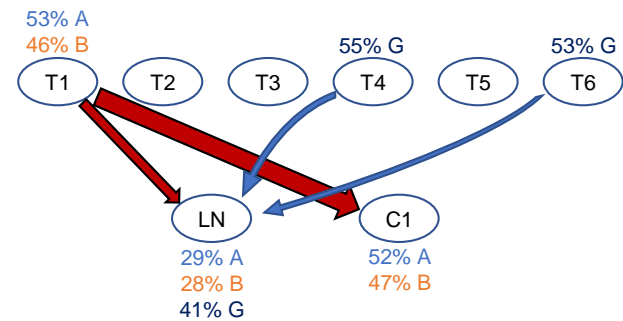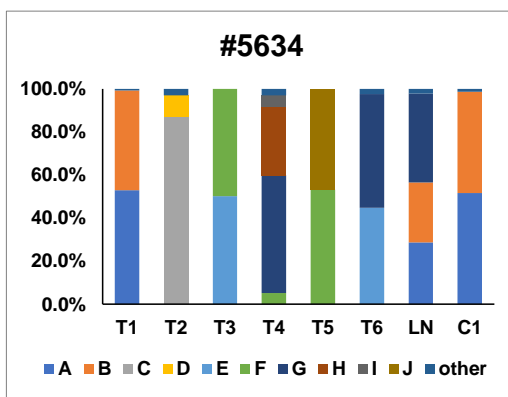
Figure 5: Inferring lineage relationships between primary and metastatic tumors.
(A) Proportions of mutant alleles detected in two primary tumors (T1, T2) and three liver metastases (C1-C3) from animal #5759.
(B) Proportions of mutant alleles detected in six primary tumors (T1-T6), one lymph node metastasis (LN) and one liver metastasis (C1) from animal #5534.
Red arrows indicate high-confidence lineage relationships between primary tumor(s) and metastases, while blue arrows indicate uncertain or ambiguous relationships.

different allele (G) not present in 5634-T1. Instead, allele G is present in 5634-T4 (55%) and 5634-T6 (53%), suggesting that either tumor, or another tumor harboring the same allele that was not sequenced, may also have seeded part of 5634-LN. Therefore, 5634-LN appears to have a polyclonal origin, being seeded from at least two different primary tumors.

DISCUSSION

Functional studies in GEMMs have long been hampered by the need to generate new germline or conditional alleles for each new gene of interest. Approaches such as gene targeting in embryonic stem cells derived from GEMMs (GEMM-ESCs) (Huijbers et al., 2014) have reduced the time needed to generate new mouse models, but still require the dedication of significant time and resources. The ability of the CRISPR-Cas9 system to generate genomic alterations in somatic cells *in vivo* with high efficiency allows this process to be bypassed (Sánchez-Rivera and Jacks, 2015). In this study, we have demonstrated the feasibility of this approach in SCLC by modeling loss of p107 and p130 in the *Trp53/Rb1* double knockout model of SCLC, showing that loss of p107 significantly increased tumor burden, to a similar extent as loss of p130. However, loss of p107 resulted in fewer but larger tumors compared with loss of p130. We also showed that metastatic tumors could be matched to their respective primary tumors within the same animal based on unique patterns of CRISPR-induced mutations.

CRISPR-mediated knockout of genes has been particularly useful for modeling loss-of-function mutations in tumor suppressor genes, such as those as demonstrated here, as well as in other models (Chiou et al., 2015; Huang et al., 2017; Sánchez-Rivera et al., 2014; Xue et al., 2014). Other studies have demonstrated the ability to generate gain-of-function alterations such as chromosomal rearrangements (Maddalo et al., 2014), as well as homologous recombination-mediated activation of oncogenes (Xue et al., 2014), albeit at a very low efficiency in the latter case. The adaptation of CRISPR-Cas9 for transcriptional activation (CRISPRa) or inhibition (CRISPRi) (Dahlman et al., 2015; Gilbert et al., 2013; Kiani et al., 2015; Konermann et al., 2015; Perez-Pinera et

al., 2013; Tanenbaum et al., 2014) has further extended the utility of the CRISPR-Cas9 system for modeling additional gain-of-function mutations. In particular, the ability to perform *in vivo* transcriptional activation using wild-type Cas9 has been reported recently (Liao et al., 2017). We expect that similar approaches will also be possible in SCLC using our system.

Due to the flexibility afforded by the CRISPR-Cas9 system, we were able to rapidly compare the effects of loss of p107 with loss of p130 in SCLC progression, which has not been done previously. Both genes are mutated in about 6% of human SCLC tumors, in a predominantly mutually-exclusive fashion (George et al., 2015). We found that loss of p107 in the *Trp53/Rb1*-mutant background resulted in fewer but larger tumors compared with loss of p130. Furthermore, loss of p107 also accelerated the development of metastatic spread in these animals. These suggest that the two genes may play different roles in SCLC development, with loss of p130 appearing to promote tumor initiation, and loss of p107 promoting tumor progression. p107 and p130 have been shown to exhibit different timings of expression and interaction with E2F family proteins during the cell cycle (Nevins, 1998; Smith et al., 1996), as well as to interact with different downstream effectors (Wirt and Sage, 2010). Interestingly, p107 and p130 have previously been reported to have distinct roles in lung epithelial development (Simpson et al., 2009). Deletion of *Rb1* in the lung epithelium of *p107*$^{-/-}$ animals results in increased pulmonary epithelial proliferation in E18.5 embryos compared with deletion of *Rb1* in *p107*$^{+/+}$ animals, while having no effect on apoptosis. Conversely, deletion of *Rb1* in *p130*$^{-/-}$ animals results in decreased apoptosis but has no effect on proliferation. It is possible that a similar dynamic exists in the context of SCLC development, where

loss of p130 may allow nascent transformed cells to escape apoptosis, resulting in an increase in initiated tumors, whereas loss of p107 results in an increase in proliferation rate of developing tumors without affecting tumor initiation. Curiously, loss of both p107 and p130 did not accelerate tumor progression compared with loss of p130 alone, suggesting that in the *p130*-null background, loss of p107 does not confer a significant growth advantage to the tumors. Further experiments, such as analysis of early lesions isolated from p107-targeted and p130-targeted animals, could shed light on the mechanisms driving these differential effects in SCLC.

The use of autochthonous models of cancer, including SCLC, to study metastasis has led to key insights into the molecular determinants of metastatic spread (Chiou et al., 2017; Chuang et al., 2017; Denny et al., 2016; Winslow et al., 2011). Because multiple independent primary tumors usually develop within a single animal in these models, the use of barcoding techniques has been important to enable metastatic tumors to be matched to their respective seeding primary tumors, so as to distinguish primary tumors based on their metastatic potential (Chuang et al., 2017; Winslow et al., 2011). In the absence of barcoding, relationships between primary and metastatic tumors can also be inferred from shared mutational patterns derived from exome sequencing data (McFadden et al., 2014). By taking advantage of the imprecise repair of CRISPR-mediated double-stranded breaks in the cell by non-homologous end joining, we have demonstrated a straightforward alternative approach for matching primary and metastatic tumors within a single animal based on the patterns of mutations at the sgRNA target site. Although the use of single target sites, such as those presented in our study, leaves some ambiguity in assigning metastatic tumors to

primary tumors in certain cases, we were still able to significantly reduce the number of possible assignments. We anticipate that the inclusion of a second sgRNA into the adenoviral vector that targets a different site in the genome will significantly reduce this ambiguity, by doubling the number of genomic loci available for generating unique combinations of mutations.

In summary, we have demonstrated the feasibility of using the CRISPR-Cas9 system for modeling loss of tumor suppressor genes in an autochthonous mouse model of SCLC. This opens the door for rapid functional validation of other candidate genes of interest that are frequently mutated in SCLC. In addition, we anticipate that this approach, combined with ongoing advances in CRISPR-based tools for genetic and transcriptomic perturbations, as well as developments in the use of lentiviral vectors with *in vivo* models of SCLC (Xia et al., 2018), will also be useful for validation of therapeutic targets for SCLC in the future.

MATERIALS AND METHODS

Mice

All animal studies were approved by the MIT Institutional Animal Care and Use Committee. All animals were maintained on a mixed B6;129 background. $Trp53^{flox/flox}$; $Rb1^{flox/flox}$; $Rosa26^{LSL-Luciferase/LSL-Luciferase}$ mice have been described previously (Dooley et al., 2011; McFadden et al., 2014). Tumors were initiated by intratracheal delivery of 2×10^8 plaque-forming units (p.f.u.) of adenovirus expressing Cre recombinase, as previously described (DuPage et al., 2009). Adenovirus stocks were prepared and titered as described below. For *in vivo* bioluminescence imaging, mice were anesthetized by isoflurane inhalation, administered with 150 mg/kg of D-luciferin (PerkinElmer #122799) by intraperitoneal injection, then imaged 10 minutes post-administration using the IVIS Spectrum In Vivo Imaging System (PerkinElmer). Visualization and quantification of bioluminescence signal was performed using Living Image (PerkinElmer).

Generation of Rosa26-CAGGS-LSL-Cas9-GFP-Csy4 allele

The Rosa26-CAGGS-LSL-Cas9-GFP-Csy4 targeting vector was cloned using constructs generated for the Gibson assembly-based modular assembly platform (GMAP), as described previously (Akama-Garren et al., 2016). In brief, the 3xFLAG-NLS-hSpCas9-2A-GFP-2A-Csy4 gene C (gC) fragment was cloned into the R26TV CAG LSL 2-5 targeting vector by Gibson assembly, using the conditions described for GMAP. The full sequences of both fragments are listed in Supplementary Table 1.

The assembled Rosa26-CAGGS-LSL-Cas9-GFP-Csy4 targeting vector was linearized by digestion with BsmBI (New England Biolabs). 40 μg of the linearized vector was transfected by electroporation into mixed B6;129 embryonic stem cells, followed by selection with 350 μg/mL G418 (Life Technologies) for 7 days. Clones were screened by PCR and sequenced to confirm correct targeting into the *Rosa26* locus, then injected into CD1 donor blastocysts to generate chimeric mice. High-degree chimeric mice were subsequently crossed into the *Trp53flox/flox; Rb1flox/flox; Rosa26LSL-Luciferase/LSL-Luciferase* background to obtain stocks of *Trp53flox/flox; Rb1flox/flox; Rosa26LSL-Cas9-GFP-Csy4/LSL-Luciferase* mice for experiments.

Generation of Ad5-USEC vectors

To generate a GMAP-compatible adenoviral vector, a filler sequence containing site #1 (GATCAGTGTGAGGGAGTGTAAAGCTGGTTT) and site #5 (AAACGTTGTTGTTT GGGGTTGAATTACTCT) was amplified by PCR using lentiCRISPRv2 (Sanjana et al., 2014) as a template, digested with XhoI and EcoRI (New England Biolabs), then ligated into the XhoI/EcoRI-digested pacAd5 shuttle vector (Anderson et al., 2000). The resulting vector was linearized with BspEI (New England Biolabs) for subsequent Gibson assembly.

sgRNAs targeting *p107*, *p130*, and a control unannotated region on mouse chromosome 4 were designed using the Broad Institute sgRNA Designer tool (Doench et al., 2016). These sgRNAs were cloned into lentiCRISPRv2 using the recommended protocol (Sanjana et al., 2014). GMAP-compatible U6-sgRNA cloning fragments for each sgRNA were amplified by PCR from the corresponding lentiCRISPRv2-sgRNA

vectors using the U6-pA-F and tracrRNA-gA-R primers. Ad5-USEC vectors were then

assembled by Gibson assembly using the BspEI-linearized adenoviral vector, U6-

sgRNA cloning fragments, as well as pEFS promoter B and NLS-Cre gene B parts from

the GMAP collection (Akama-Garren et al., 2016). All vectors were verified by

sequencing before use.

Adenoviral vectors were packaged at the Viral Vector Core of the Horae Gene

Therapy Center, University of Massachusetts Medical School.

Cell culture

All cell lines were maintained in DMEM (Corning #10-013-CV) supplemented with 10%

fetal bovine serum, 2 mM L-glutamine (Gibco #25030) and 50 µg/mL gentamicin (Gibco

#15710). Lentiviral vectors were generated in HEK293T cells. In brief, cells were plated

1 day before transfection, then co-transfected with lentiviral constructs and packaging

plasmids psPAX2 and pMD2.G (Addgene #12260 and #12259; both plasmids were gifts

from Didier Trono). Viral supernatant was harvested 48 and 72 hours after transfection,

then frozen at -80°C. Green-Go reporter cells (Sánchez-Rivera et al., 2014) were

transduced with lentiCas9-Blast virus (Addgene #52962; plasmid was a gift from Feng

Zhang), then selected with 20 µg/mL Blasticidin S (Gibco #A11139) for 7 days.

Flow cytometry

Cells were trysinized, centrifuged at 1,000 rpm (approximately 200 x g) for 5 minutes,

resuspended in FACS buffer (PBS, 0.5% fetal bovine serum, 2 mM EDTA), then filtered

through a 35 μm cell strainer (Corning #352235). Samples were sorted on a BD

FACSAria III system (BD Biosciences).

Immunoblotting

Cells were lysed with RIPA lysis buffer (Thermo Scientific #89900) supplemented with

1x protease and phosphatase inhibitor cocktail (Thermo Scientific #78440), rotated at

4°C for 30 minutes, then centrifuged at 13,000 x g for 30 minutes. Protein concentration

was quantified using the BCA protein assay (Thermo Scientific #23225). 40 μg of

protein was run on a 4-12% Bis-Tris gradient gel (Invitrogen #NP0335), then transferred

onto a nitrocellulose membrane. The following primary antibodies were used for

immunoblotting: rabbit anti-p107 (Santa Cruz #sc-318, 1:100), rabbit anti-p130 (Santa

Cruz #sc-317, 1:200), mouse anti-Hsp90 (BD Biosciences #610418, 1:5,000). Primary

antibodies were detected with the following fluorescence-conjugated secondary

antibodies: IRDye 680RD Donkey anti-Mouse IgG (LI-COR #926-68072, 1:10,000),

IRDye® 800CW Donkey anti-Rabbit IgG (LI-COR # 926-32213, 1:10,000). Immunoblots

were imaged using the LI-COR Odyssey Infrared Imager and quantified using Image

Studio (LI-COR).

Genomic DNA isolation and deep sequencing of target loci

Tumor tissue was dissected from lungs, lymph nodes or liver tissue upon necropsy,

snap-frozen in liquid nitrogen, then stored at -80°C until subsequent processing.

Genomic DNA was isolated from tumors using the Gentra PureGene Tissue Kit

(QIAGEN #158667). Genomic loci were amplified by PCR using either Herculase II

Fusion DNA Polymerase (Agilent #600675) for control sgRNA samples, or KAPA HiFi

DNA Polymerase (KAPA Biosystems #KK2601) for sgp107 and sgp130 samples.

Amplified samples were purified using the QIAquick PCR Purification Kit (QIAGEN

#28104), then submitted for deep sequencing using the CRISPR Sequencing service at

the DNA Core of the Center for Computational & Integrative Biology (CCIB),

Massachusetts General Hospital. Sequence variant detection was performed by the

CCIB DNA Core using their standard algorithm.

Histology

Harvested tissues were fixed in zinc formalin (Polysciences #21516) overnight at room

temperature, then transferred to 70% ethanol until subsequent paraffin embedding.

Hematoxylin and eosin (H&E) staining was performed on 4 µm tissue sections using the

Varistain Gemini ES Automated Slide Stainer (Thermo Shandon).

Tumor burden analysis

H&E-stained slides were imaged using the Aperio AT2 slide scanner (Leica Biosystems)

and visualized using Aperio ImageScope. Tumor regions and total lung area were

outlined and quantified in ImageScope. Tumor burden was calculated as the percentage

of tumor area over the total lung area, as measured from the largest cross-sectional

region of the lung lobes.

ACKNOWLEDGEMENTS

SUPPLEMENTARY TABLES AND FIGURES

Supplementary Table 1: Primers for cloning Ad5-USEC vector

Note: GMAP-compatible overhangs are underlined and in bold.

Primer	Sequence
Ad5-GMAP-filler-F	GGCCAACTCGAG**GATCAGTGTGAGGGAGTGTAAAGCTGGTTT**CCGGACACCGGAGACGGTTGTAAAT
Ad5-GMAP-filler-R	GGCCAAGAATTC**AGAGTAATTCAACCCCAAACAACAACGTTT**CCGGAGAGACGTACAAAAAAGAGCAAG
U6-pA-F	**GATCAGTGTGAGGGAGTGTAAAGCTGGTTT**GAGGGCCTATTTCCCATGATTCC
tracrRNA-gA-R	**AGGCCTCGGGATTcctaggAACAGCGGTTT**AAAAAAGCACCGACTCGGTGCC

Supplementary Table 2: sgRNA sequences used in this study

Target gene	sgRNA sequence (PAM sequence in bold)
p107 (Rbl1)	TAGGCTGTAGTTGCCGCGGA**TGG**
p130 (Rbl2)	GTACGTTCTCGGAAATGTGG**GGG**
Control (unannotated region on chromosome 4)	TGACAAACGTCCGGAAGCGC**AGG**

Supplementary Table 3: Primers for genomic loci amplification

Primer	Sequence
sgp107 forward	GCCGAGCTACACCCACCCTTC
sgp107 reverse	CTGGACGAGGGAAGCGCGGC
sgp130 forward	AAGATGAAGAAGTGGGAAGACATGGCAAAT
sgp130 reverse	ATTTTCCCGGAGAGAAGAGACACATACCTC
Control sgRNA forward	CGATGTTCTTGACCTACCAGTTCTTGAGAC
Control sgRNA reverse	CTTGAACTCAGAAATCCGCCTGCCTCT

Mutation	Indel	Frequency	Sequence
			CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCGCGGATGGCCGTGAAGTCATCCAGGGCTTCG
FS	+1	26/37	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCGC**C**GGATGGCCGTGAAGTCATCCAGGGCTTCG
FS	+1	4/37	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCGC**A**GGATGGCCGTGAAGTCATCCAGGGCTTCG
FS	+1	4/37	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCGC**T**GGATGGCCGTGAAGTCATCCAGGGCTTCG
FS	+1	3/37	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCGC**G**GGATGGCCGTGAAGTCATCCAGGGCTTCG
FS	-1	5/37	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCGC-GATGGCCGTGAAGTCATCCAGGGCTTCG
FS	-1	3/37	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCG-GGATGGCCGTGAAGTCATCCAGGGCTTCG
FS	-2	4/37	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGC--CGGATGGCCGTGAAGTCATCCAGGGCTTCG
FS	-2	2/37	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCGC--ATGGCCGTGAAGTCATCCAGGGCTTCG
NFS	-3	1/37	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCC---GATGGCCGTGAAGTCATCCAGGGCTTCG
FS	-4	2/37	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTG----GGATGGCCGTGAAGTCATCCAGGGCTTCG
FS	-7	3/37	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCC-------GCCGTGAAGTCATCCAGGGCTTCG
FS	-8	1/37	CCTGCCGCCGCTCACCTCTAGGCTGTA--------GGATGGCCGTGAAGTCATCCAGGGCTTCG
FS	-8	1/37	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCGC---T-----TGAAGTCATCCAGGGCTTCG
FS	-10	5/37	CCTGCCGCCGCTCACCTCTAGGCTGTAGTT----------GCCGTGAAGTCATCCAGGGCTTCG
FS	-11	1/37	CCTGCCGCCGCTCACCTCTAGGCT----------GGATGGCCGTGAAGTCATCCAGGGCTTCG
NFS	-12	1/37	CCTGCCGCCGCTCACCTCTAGGCTGTA-----------GGCCGTGAAGTCATCCAGGGCTTCG
FS	-16	1/37	CCTGCCGCCGCTCACCTC---------------CGGATGGCCGTGAAGTCATCCAGGGCTTCG
FS	-16	1/37	CCTGCCGCCGCTCACCTCTA--------------GATGGCCGTGAAGTCATCCAGGGCTTCG
FS	-16	1/37	CCTGCCGCCGCTCACCTCT--------------GGATGGCCGTGAAGTCATCCAGGGCTTCG
FS	-19	2/37	CCTGCCGCCGCTCACCTCTA----------------GGCCGTGAAGTCATCCAGGGCTTCG
FS	-19	1/37	CCTGCCGCCGCTCACCTCTAGGCT------------------GTGAAGTCATCCAGGGCTTCG
FS	-20	1/37	CCTGCCGCCGCTCACCTCTAGG------------------CGTGAAGTCATCCAGGGCTTCG
NFS	-21	1/37	CCTGCCGCCGCTCACCTCTAGGC------------------TGAAGTCATCCAGGGCTTCG
FS	-28	1/37	CCTGCCGCCGCTCACCTCT--------------------------AGTCATCCAGGGCTTCG
FS	-29	1/37	CCTGCCGCCGCTCACCT---------------------------AAGTCATCCAGGGCTTCG
FS	-29	1/37	CCTGCCGCCGCTCACCTCTAGGCT--------------------------CCAGGGCTTCG
FS/SS	-55	1/37	CCTGCCGCCGCT---

**Supplementary Figure 1: Mutant alleles detected in sgp107-targeted tumors.**
Mutant alleles detected by deep sequencing of the *p107* genomic locus from tumors isolated from Trp53/Rb1/Cas9 animals infected with sgp107. The type of mutation, indel size, frequency of occurrence across all sequenced tumors, and sequence of the allele are shown. FS: frameshift, NFS: non-frameshift, SS: splice site.

A
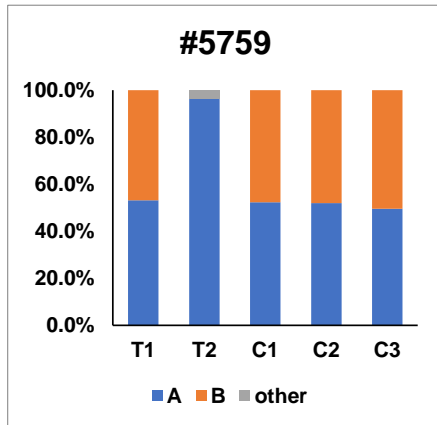
#5759

WT	CCTGCCGCCGCTCACCTC<u>TAGGCTGTAGTTGCCGCGGA</u>TGGCCGTGAAGTCATCCAGGGCTTCG
A	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCGC_CGGATGGCCGTGAAGTCATCCAGGGCTTCG
B	CCTGCCGCCGCTCACCTCTAGGCTGTAGTT----------GCCGTGAAGTCATCCAGGGCTTCG

B

#5634
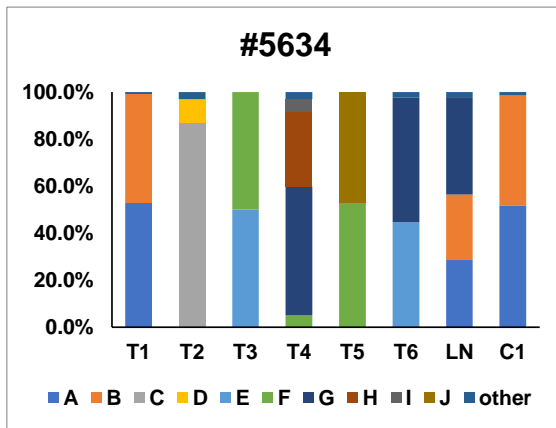
WT	CCTGCCGCCGCTCACCTC<u>TAGGCTGTAGTTGCCGCGGA</u>TGGCCGTGAAGTCATCCAGGGCTTCG
A	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCGC_GGGATGGCCGTGAAGTCATCCAGGGCTTCG
B	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCC-------GCCGTGAAGTCATCCAGGGCTTCG
C	CCTGCCGCCGCTCACCTC---------------CGGATGGCCGTGAAGTCATCCAGGGCTTCG
D	CCTGCCGCCGCT--
E	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCGC_AGGATGGCCGTGAAGTCATCCAGGGCTTCG
F	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCGC-GATGGCCGTGAAGTCATCCAGGGCTTCG
G	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCGC_CGGATGGCCGTGAAGTCATCCAGGGCTTCG
H	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCGC_TGGATGGCCGTGAAGTCATCCAGGGCTTCG
I	CCTGCCGCCGCTCACCTCTA------------------GGCCGTGAAGTCATCCAGGGCTTCG
J	CCTGCCGCCGCTCACCTCTAGGCTGTAGTTGCCG-GGATGGCCGTGAAGTCATCCAGGGCTTCG

Supplementary Figure 2. Sequences of mutant alleles from tumors used to infer lineage relationships in mouse #5759 and #5634.

REFERENCES

Akama-Garren, E.H., Joshi, N.S., Tammela, T., Chang, G.P., Wagner, B.L., Lee, D.-Y., Rideout III, W.M., Papagiannakopoulos, T., Xue, W., and Jacks, T. (2016). A Modular Assembly Platform for Rapid Generation of DNA Constructs. Sci. Rep. *6*, 16836.

Anderson, R.D., Haskell, R.E., Xia, H., Roessler, B.J., and Davidson, B.L. (2000). A simple method for the rapid generation of recombinant adenovirus vectors. Gene Ther. *7*, 1034–1038.

Byers, L.A., and Rudin, C.M. (2015). Small cell lung cancer: Where do we go from here? Cancer *121*, 664–672.

Califano, R., Abidin, A.Z., Peck, R., Faivre-Finn, C., and Lorigan, P. (2012). Management of Small Cell Lung Cancer. Drugs *72*, 471–490.

Chiou, S.-H., Risca, V.I., Wang, G.X., Yang, D., Grüner, B.M., Kathiria, A.S., Ma, R.K., Vaka, D., Chu, P., Kozak, M., et al. (2017). Blimp1 induces transient metastatic heterogeneity in pancreatic cancer. Cancer Discov. CD-17-0250.

Chiou, S., Winters, I.P., Wang, J., Naranjo, S., Dudgeon, C., Tamburini, F.B., Brady, J.J., Yang, D., Grüner, B.M., Chuang, C., et al. (2015). Pancreatic cancer modeling using retrograde viral vector delivery and in vivo CRISPR/Cas9-mediated somatic genome editing. Genes Dev. *29*, 1576–1585.

Chuang, C.H., Greenside, P.G., Rogers, Z.N., Brady, J.J., Yang, D., Ma, R.K., Caswell, D.R., Chiou, S.H., Winters, A.F., Grüner, B.M., et al. (2017). Molecular definition of a metastatic lung cancer state reveals a targetable CD109-Janus kinase-Stat axis. Nat. Med. *23*, 291–300.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Cui, M., Augert, A., Rongione, M., Conkrite, K., Parazzoli, S., Nikitin, A.Y., Ingolia, N., and MacPherson, D. (2014). PTEN Is a Potent Suppressor of Small Cell Lung Cancer. Mol. Cancer Res. *12*, 654–659.

Dahlman, J.E., Abudayyeh, O.O., Joung, J., Gootenberg, J.S., Zhang, F., and Konermann, S. (2015). Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. Nat. Biotechnol. *33*, 1159–1161.

Demedts, I.K., Vermaelen, K.Y., and Van Meerbeeck, J.P. (2010). Treatment of extensive-stage small cell lung carcinoma: Current status and future prospects. Eur. Respir. J. *35*, 202–215.

Denny, S.K., Yang, D., Chuang, C.-H., Brady, J.J., Lim, J.S., Grüner, B.M., Chiou, S.-H., Schep, A.N., Baral, J., Hamard, C., et al. (2016). Nfib Promotes Metastasis through a Widespread Increase in Chromatin Accessibility. Cell *166*, 328–342.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

Dooley, A.L., Winslow, M.M., Chiang, D.Y., Banerji, S., Stransky, N., Dayton, T.L., Snyder, E.L., Senna, S., Whittaker, C.A., Bronson, R.T., et al. (2011). Nuclear factor I/B is an oncogene in small cell lung cancer. Genes Dev. *25*, 1470–1475.

Dow, L.E., Fisher, J., O'Rourke, K.P., Muley, A., Kastenhuber, E.R., Livshits, G., Tschaharganeh, D.F., Socci, N.D., and Lowe, S.W. (2015). Inducible in vivo genome editing with CRISPR-Cas9. Nat. Biotechnol. *33*, 390–394.

DuPage, M., Dooley, A.L., and Jacks, T. (2009). Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. Nat. Protoc. *4*, 1064–1072.

George, J., Lim, J.S., Jang, S.J., Cun, Y., Ozretić, L., Kong, G., Leenders, F., Lu, X., Fernández-Cuesta, L., Bosco, G., et al. (2015). Comprehensive genomic profiles of small cell lung cancer. Nature *524*, 47–53.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell *154*, 442–451.

Govindan, R., Page, N., Morgensztern, D., Read, W., Tierney, R., Vlahiotis, A., Spitznagel, E.L., and Piccirillo, J. (2006). Changing epidemiology of small-cell lung cancer in the United States over the last 30 years: analysis of the surveillance, epidemiologic, and end results database. J. Clin. Oncol. *24*, 4539–4544.

Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J.A. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science *329*, 1355–1358.

Huang, J., Chen, M., Whitley, M.J., Kuo, H., Xu, E.S., Walens, A., Mowery, Y.M., Van Mater, D., Eward, W.C., Cardona, D.M., et al. (2017). Generation and comparison of CRISPR-Cas9 and Cre-mediated genetically engineered mouse models of sarcoma. Nat. Commun. *8*, 15999.

Huijbers, I.J., Bin Ali, R., Pritchard, C., Cozijnsen, M., Kwon, M.-C.C., Proost, N., Song, J.-Y.Y., de Vries, H., Badhai, J., Sutherland, K., et al. (2014). Rapid target gene validation in complex cancer mouse models using re-derived embryonic stem cells. EMBO Mol. Med. *6*, 212–225.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. Elife *2*, e00471.

Kiani, S., Chavez, A., Tuttle, M., Hall, R.N., Chari, R., Ter-Ovanesyan, D., Qian, J., Pruitt, B.W., Beal, J., Vora, S., et al. (2015). Cas9 gRNA engineering for genome editing, activation and repression. Nat. Methods *12*, 1051–1054.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

Liao, H.-K., Hatanaka, F., Araoka, T., Reddy, P., Wu, M.-Z., Sui, Y., Yamauchi, T., Sakurai, M., O'Keefe, D.D., Núñez-Delicado, E., et al. (2017). In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans -epigenetic Modulation. Cell 1–13.

Maddalo, D., Manchado, E., Concepcion, C.P., Bonetti, C., Vidigal, J.A., Han, Y.-C., Ogrodowski, P., Crippa, A., Rekhtman, N., de Stanchina, E., et al. (2014). In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. Nature *516*, 423–427.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

McFadden, D.G., Papagiannakopoulos, T., Taylor-Weiner, A., Stewart, C., Carter, S.L., Cibulskis, K., Bhutkar, A., McKenna, A., Dooley, A., Vernon, A., et al. (2014). Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. Cell *156*, 1298–1311.

Meuwissen, R., Linn, S.C., Linnoila, R.I., Zevenhoven, J., Mooi, W.J., and Berns, A. (2003). Induction of small cell lung cancer by somatic inactivation of both Trp53 and Rb1 in a conditional mouse model. Cancer Cell *4*, 181–189.

Mollaoglu, G., Guthrie, M.R., Böhm, S., Brägelmann, J., Can, I., Ballieu, P.M., Marx, A., George, J., Heinen, C., Chalishazar, M.D., et al. (2017). MYC Drives Progression of Small Cell Lung Cancer to a Variant Neuroendocrine Subtype with Vulnerability to Aurora Kinase Inhibition. Cancer Cell 1–16.

Nevins, J.R. (1998). Toward an understanding of the functional complexity of the E2F and retinoblastoma families. Cell Growth Differ. *9*, 585–593.

Nissim, L., Perli, S.D., Fridkin, A., Perez-Pinera, P., and Lu, T.K. (2014). Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. Mol. Cell *54*, 698–710.

Peifer, M., Fernández-Cuesta, L., Sos, M.L., George, J., Seidel, D., Kasper, L.H., Plenker, D., Leenders, F., Sun, R., Zander, T., et al. (2012). Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat. Genet. *44*, 1104–1110.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., et al. (2013). RNA-guided gene activation by CRISPR-Cas9–based transcription factors. Nat. Methods *10*, 973–976.

Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M., et al. (2014). CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Cell *159*, 440–455.

Roper, J., Tammela, T., Cetinbas, N.M., Akkad, A., Roghanian, A., Rickelt, S., Almeqdadi, M., Wu, K., Oberli, M.A., Sánchez-Rivera, F., et al. (2017). In vivo genome editing and organoid transplantation models of colorectal cancer and metastasis. Nat. Biotechnol. *35*, 569–576.

Rudin, C.M., Durinck, S., Stawiski, E.W., Poirier, J.T., Modrusan, Z., Shames, D.S., Bergbower, E.A., Guan, Y., Shin, J., Guillory, J., et al. (2012). Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. Nat. Genet. *44*, 1111–1116.

Sánchez-Rivera, F.J., and Jacks, T. (2015). Applications of the CRISPR–Cas9 system in cancer biology. Nat. Rev. Cancer *15*, 387–395.

Sánchez-Rivera, F.J., Papagiannakopoulos, T., Romero, R., Tammela, T., Bauer, M.R., Bhutkar, A., Joshi, N.S., Subbaraj, L., Bronson, R.T., Xue, W., et al. (2014). Rapid modelling of cooperating genetic events in cancer through somatic genome editing. Nature *516*, 428–431.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Schaffer, B.E., Park, K.S., Yiu, G., Conklin, J.F., Lin, C., Burkhart, D.L., Karnezis, A.N., Sweet-Cordero, E.A., and Sage, J. (2010). Loss of p130 Accelerates Tumor Development in a Mouse Model for Human Small-Cell Lung Carcinoma. Cancer Res. *70*, 3877–3883.

Semenova, E.A., Kwon, M., Monkhorst, K., Song, J.-Y., Bhaskaran, R., Krijgsman, O., Kuilman, T., Peters, D., Buikhuisen, W.A., Smit, E.F., et al. (2016). Transcription Factor NFIB Is a Driver of Small Cell Lung Cancer Progression in Mice and Marks Metastatic Disease in Patients. Cell Rep. *16*, 631–643.

Simpson, D.S., Mason-Richie, N.A., Gettler, C.A., and Wikenheiser-Brokamp, K.A. (2009). Retinoblastoma family proteins have distinct functions in pulmonary epithelial cells in vivo critical for suppressing cell growth and tumorigenesis. Cancer Res. *69*, 8733–8741.

Smith, E.J., Leone, G., DeGregori, J., Jakoi, L., and Nevins, J.R. (1996). The accumulation of an E2F-p130 transcriptional repressor distinguishes a G0 cell state from a G1 cell state. Mol. Cell. Biol. *16*, 6965–6976.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. Cell *159*, 635–646.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558.

Winslow, M.M., Dayton, T.L., Verhaak, R.G.W., Kim-Kiselak, C., Snyder, E.L., Feldser, D.M., Hubbard, D.D., DuPage, M.J., Whittaker, C. a, Hoersch, S., et al. (2011). Suppression of lung adenocarcinoma progression by Nkx2-1. Nature *473*, 101–104.

Wirt, S.E., and Sage, J. (2010). p107 in the public eye: an Rb understudy and more. Cell Div. *5*, 9.

Wistuba, I., Gazdar, A., and Minna, J. (2001). Molecular genetics of small cell lung carcinoma. Semin. Oncol. *28*, 3–13.

Wu, N., Jia, D., Ibrahim, A.H., Bachurski, C.J., Gronostajski, R.M., and MacPherson, D. (2016). NFIB overexpression cooperates with *Rb/p53* deletion to promote small cell lung cancer. Oncotarget *7*, 57514–57524.

Xia, Y., Zhan, C., Feng, M., Leblanc, M., Ke, E., Yeddula, N., and Verma, I.M. (2018). Targeting CREB Pathway Suppresses Small Cell Lung Cancer. Mol. Cancer Res. *16*, 825–832.

Xue, W., Chen, S., Yin, H., Tammela, T., Papagiannakopoulos, T., Joshi, N.S., Cai, W., Yang, G., Bronson, R., Crowley, D.G., et al. (2014). CRISPR-mediated direct mutation of cancer genes in the mouse liver. Nature *514*, 380–384.

CHAPTER 3

Identification of novel therapeutic targets in small cell lung cancer through CRISPR-based genetic screens

Leanne Li[1] *, **Sheng Rong Ng**[1,2] *, Caterina Colón[1], Arjun Bhutkar[1], Tyler Jacks[1,2,3]

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

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

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

* These authors contributed equally to this work.

ABSTRACT

Small cell lung cancer (SCLC) is a highly aggressive form of lung cancer with an extremely poor survival rate. Despite substantial efforts, the current standard of care has remained unchanged for over thirty years, with no new treatment options having demonstrated significantly improved outcomes. Genomic sequencing studies to identify molecular drivers of SCLC progression have not led to many promising targets for treatment of the disease. Here, we have taken a CRISPR-based genetic screening approach to identify novel genetic vulnerabilities that may serve as potential therapeutic targets. We designed a customized sgRNA library targeting the druggable genome and used this library to perform loss-of-function genetic screens in a panel of SCLC cell lines derived from an autochthonous murine model of SCLC. By performing cross-cancer analyses, we identified potential SCLC-specific targets for downstream functional validation. We found that SCLC cells displayed enhanced sensitivity towards disruption of several key metabolic pathways, including the tricarboxylic acid cycle, oxidative phosphorylation, and pyrimidine biosynthesis. Pharmacological inhibition of Dhodh, a key enzyme in the pyrimidine biosynthesis pathway, reduced the viability of SCLC cells *in vitro* and strongly suppressed SCLC tumor growth *in vivo*. In summary, we have identified a novel metabolic vulnerability in SCLC that represents a promising target for therapeutic intervention in this deadly disease.

INTRODUCTION

Small cell lung cancer (SCLC) is a highly aggressive form of lung cancer that is among the deadliest of all solid tumor malignancies. It is characterized by rapid growth and early, widespread metastasis (Califano et al., 2012), which results in very poor overall prognosis. Despite decades of research, treatment of SCLC still primarily involves the use of cytotoxic chemotherapy regimens, most commonly a combination of cisplatin and etoposide (Demedts et al., 2010). In contrast to the growing number of targeted therapies available for treating non-small cell lung cancer, no new therapies have demonstrated efficacy in SCLC patients (Byers and Rudin, 2015), which highlights a significant need to develop novel approaches for treating the disease.

Large-scale cancer genome sequencing studies have contributed significantly to the development of new therapies for the treatment of various cancers (Vogelstein et al., 2013). By identifying genes that are frequently mutated in different cancers, such studies have enabled the discovery of novel potential therapeutic targets, as well as expanded the use of existing therapies to other cancer types. However, functional validation of these genetic alterations in relevant cancer models remains a significant bottleneck in this process, despite the recent development of novel approaches that have significantly sped up this process, as discussed in Chapter 1.

Genetic screens in mammalian cells have been used to identify and characterize essential genes in the human genome, as well as cancer type-specific and genotype-specific vulnerabilities in cancer cells (Hart et al., 2015; Tzelepis et al., 2016; Wang et al., 2015, 2017). The adoption of the CRISPR-Cas9 system for genetic screens has overcome many of the difficulties associated with the use of RNAi approaches, such as
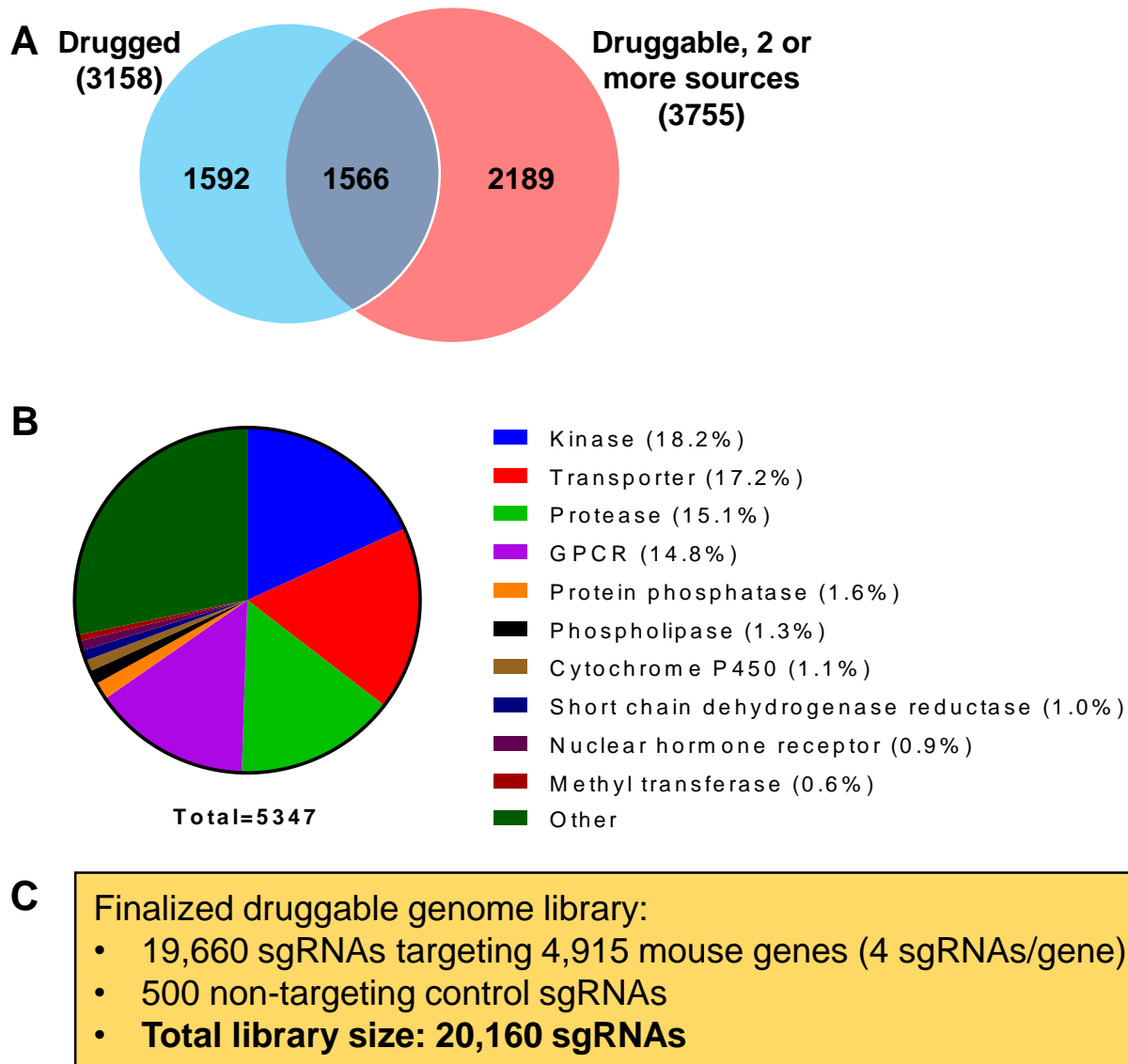
incomplete knockdown of target genes and potentially confounding off-target effects (Echeverri et al., 2006). This has provided a powerful complementary approach to cancer genome sequencing studies for identifying new targets for therapy, especially when combined with subsequent functional validation in relevant preclinical models.

In this study, we utilize a custom sgRNA library targeting the druggable genome to carry out genetic perturbation screens in genetically defined cell lines derived from an autochthonous murine model of SCLC (Meuwissen et al., 2003) to identify novel genetic vulnerabilities. To exclude broadly essential genes and focus on SCLC-specific targets, we compare our results with parallel screens performed in our laboratory in lung adenocarcinoma (LUAD) and pancreatic ductal adenocarcinoma (PDAC) cell lines using the same sgRNA library. Through this approach, we identify potential vulnerabilities in several key metabolic processes, and demonstrate that SCLC cells show significant sensitivity towards disruption of the *de novo* pyrimidine synthesis pathway.

RESULTS

Design of druggable genome sgRNA library

In order to focus on therapeutically relevant genes, we designed a customized sgRNA library that targets the druggable genome (Hopkins and Groom, 2002). This consists of genes that express proteins considered to be good potential drug targets, and includes protein families such as kinases, G-protein-coupled receptors (GPCRs), proteases, nuclear hormone receptors, and phosphodiesterases (Hopkins and Groom, 2002). To obtain a comprehensive consensus set of druggable genes, we used the Drug-Gene Interaction Database (Griffith et al., 2013; Wagner et al., 2016) to define two categories of genes. The first category consists of genes that are known targets of existing drug compounds, which we termed "drugged" genes. The second category consists of genes that belong to druggable gene categories, which we termed "druggable" genes. To generate a high-confidence list of "druggable" genes, we included only genes that were listed in at least two different primary sources (**Supplementary Table 1**). This resulted in a total of 5347 genes (**Fig. 1A**), with 4915 corresponding mouse orthologs. As expected, the large majority of the genes were kinases, transporters, proteases and GPCRs (**Fig. 1B**). The final sgRNA library consisted of 20,160 sgRNAs, with four sgRNAs targeting each gene, and an additional 500 non-targeting control sgRNAs (**Fig. 1C**). Targeting sgRNAs were extracted from the Brie whole-genome sgRNA library or designed using the Broad Institute sgRNA design tool (Doench et al., 2016), while control sgRNAs were taken from the Asiago library from the same study.

A Drugged (3158) Druggable, 2 or more sources (3755)

1592 1566 2189

B

Total=5347

- Kinase (18.2%)
- Transporter (17.2%)
- Protease (15.1%)
- GPCR (14.8%)
- Protein phosphatase (1.6%)
- Phospholipase (1.3%)
- Cytochrome P450 (1.1%)
- Short chain dehydrogenase reductase (1.0%)
- Nuclear hormone receptor (0.9%)
- Methyl transferase (0.6%)
- Other

C Finalized druggable genome library:
- 19,660 sgRNAs targeting 4,915 mouse genes (4 sgRNAs/gene)
- 500 non-targeting control sgRNAs
- **Total library size: 20,160 sgRNAs**

Figure 1: Design of druggable genome sgRNA library.
(A) Number of genes included for each category in the druggable genome library, including the overlap between both categories.
**(B)** Composition of genes within the druggable genome library.
(C) Breakdown of the total number of sgRNAs in the druggable genome library.

136

We cloned the sgRNA library into lentiGuide-Puro, a lentiviral vector that expresses an sgRNA together with a puromycin vector (Sanjana et al., 2014; **Fig. 2A**). This library was used to infect 4 independent murine SCLC cell lines, which were isolated from the $Trp53^{flox/flox}$; $Rb1^{flox/flox}$ model of SCLC (Meuwissen et al., 2003). These lines were infected with a lentiviral construct, lentiCas9-Blast (**Fig. 2A**), to enable stable expression of Cas9, prior to infection with the library. After library infection, an initial population of infected cells was harvested for genomic DNA (PD0), while the remaining cells were passaged for a total of 13 population doublings (PD) before the final population was harvested (PD13) (**Fig. 2B**). To ensure that every sgRNA was sufficiently represented in the population, we maintained at least 20 million cells at every step of the screen (1,000X coverage). Screen deconvolution was carried out by high-throughput sequencing of the integrated sgRNA site, as previously described (Joung et al., 2017).

Identification of SCLC-specific genetic vulnerabilities

For each cell line, we first determined the change in sgRNA representation between the starting population (PD0) and final population (PD13) for each sgRNA, calculated as the difference between the base 2 logarithm of the normalized sequencing counts at PD13 and PD0 (log2 fold change, or L2FC). Next, we defined the gene score for each gene as the median L2FC for all four sgRNAs targeting that gene. A negative gene score corresponds to the depletion of sgRNAs targeting that gene over time, implying that loss of function of the gene negatively impacted cell fitness. Conversely, a positive gene score implies that loss of gene function increased cell fitness.
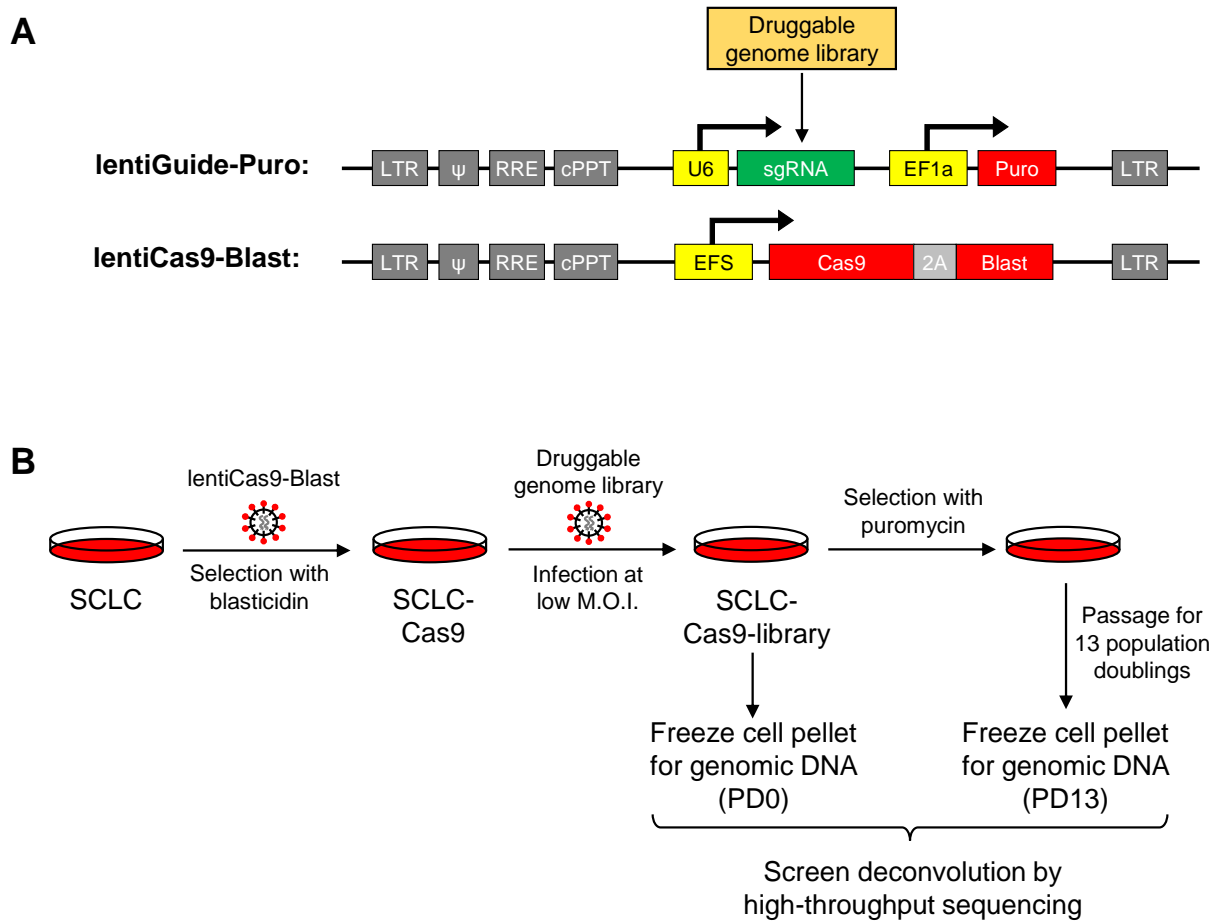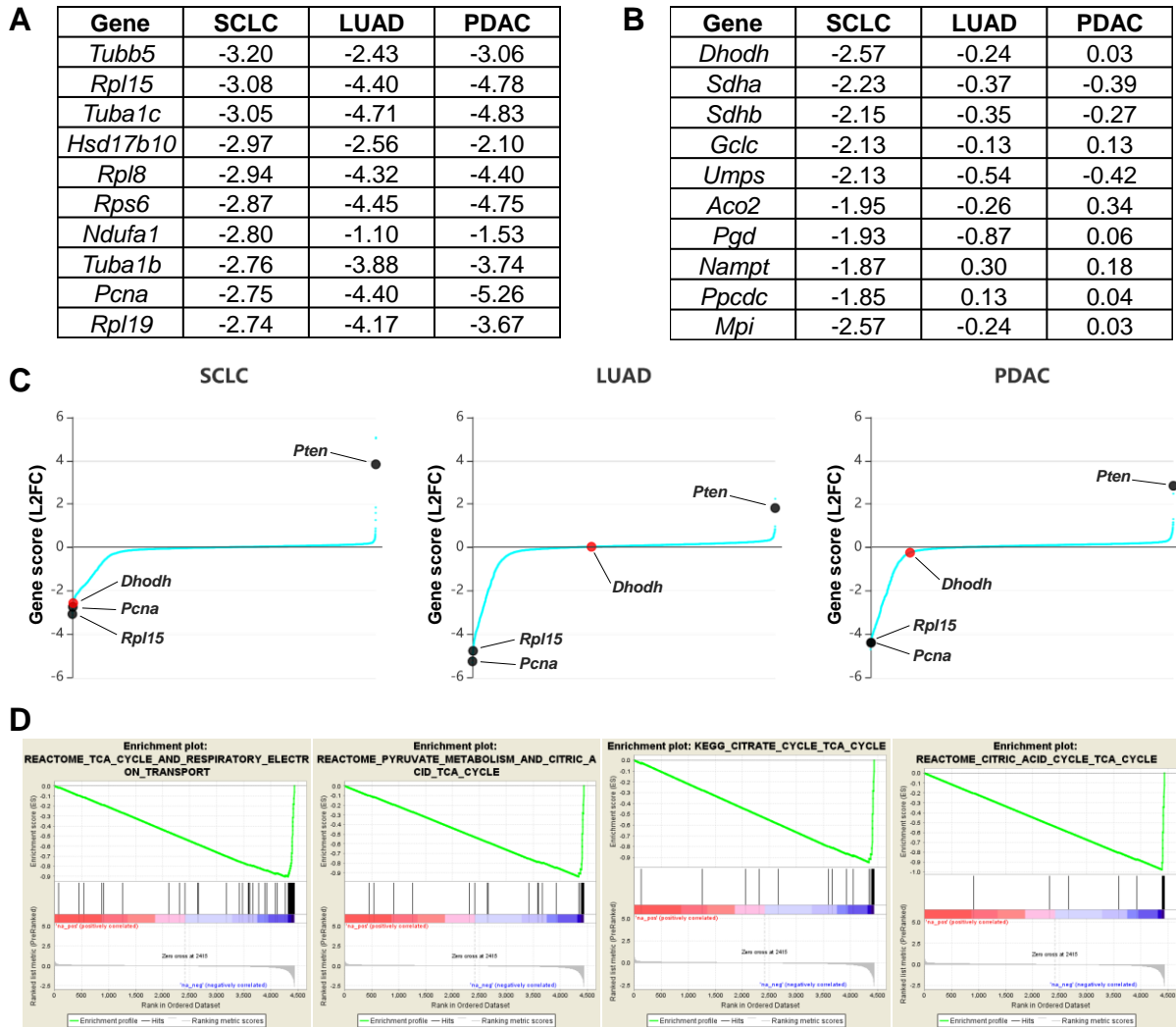
Figure 2: Strategy for CRISPR screen in SCLC cell lines.
(A) Schematic of lentiviral vectors used in the CRISPR screen. Vectors were obtained from Sanjana et al., 2014.
(B) Workflow for CRISPR screen.

To determine whether the screen was performing as expected, we analyzed the list of genes with lowest (most negative) and highest (most positive) gene scores in SCLC cell lines. The genes with the lowest gene scores in SCLC cell lines also tended to exhibit strongly negative gene scores in both LUAD and PDAC cell lines (**Fig. 3A, C**), suggesting that the screen was successfully identifying essential genes. Indeed, many of these genes, such as *Rpl15* and *Pcna*, are involved in essential cell processes such as translation and DNA replication. In addition, sgRNAs targeting *Pten* were found to be strongly enriched in all three cell types (**Fig. 3C**), consistent with its known role as a tumor suppressor gene. To confirm our initial observations, we performed gene set enrichment analysis (GSEA) (Subramanian et al., 2005) on the list of genes ranked by gene score for each of the three cell types. For all three cell types, the most significantly negatively enriched gene sets reflected various essential processes such as tRNA aminoacylation, translational regulation, and cell cycle progression (**Supplementary Fig. 1A-C; Supplementary Tables 2-4**), demonstrating that the screen was identifying functionally relevant genes. Therefore, these data validate that the screen was performing as expected.

Next, to exclude genes that are broadly essential across all three cell types, we filtered out all genes that had median gene scores of less than -0.5 in all three cell types, which removed approximately 400 genes from the list. This allowed us to focus on SCLC-specific vulnerabilities (**Fig. 3B**). When we performed GSEA on the remaining genes ranked by SCLC gene score, we observed that genes sets involving the tricarboxylic acid (TCA) cycle and oxidative phosphorylation (**Fig. 3D**), as well as pyrimidine metabolism (**Supplementary Table 5**), were significantly negatively

A

Gene	SCLC	LUAD	PDAC
Tubb5	-3.20	-2.43	-3.06
Rpl15	-3.08	-4.40	-4.78
Tuba1c	-3.05	-4.71	-4.83
Hsd17b10	-2.97	-2.56	-2.10
Rpl8	-2.94	-4.32	-4.40
Rps6	-2.87	-4.45	-4.75
Ndufa1	-2.80	-1.10	-1.53
Tuba1b	-2.76	-3.88	-3.74
Pcna	-2.75	-4.40	-5.26
Rpl19	-2.74	-4.17	-3.67

B

Gene	SCLC	LUAD	PDAC
Dhodh	-2.57	-0.24	0.03
Sdha	-2.23	-0.37	-0.39
Sdhb	-2.15	-0.35	-0.27
Gclc	-2.13	-0.13	0.13
Umps	-2.13	-0.54	-0.42
Aco2	-1.95	-0.26	0.34
Pgd	-1.93	-0.87	0.06
Nampt	-1.87	0.30	0.18
Ppcdc	-1.85	0.13	0.04
Mpi	-2.57	-0.24	0.03

C

D

Figure 3: Identification of SCLC-specific genetic vulnerabilities.
(A) Genes with the lowest (most negative) gene scores in SCLC cell lines, with the corresponding gene scores for each indicated cell type.
(B) Genes with the lowest gene scores in SCLC cell lines, after exclusion of "broadly essential genes", with the corresponding gene scores for each indicated cell type.
(C) Graph of gene scores for each indicated cell type, ranked from lowest to highest. Gene scores for *Rpl15*, *Pcna*, *Pten* and *Dhodh* are indicated on each graph.
(D) Gene set enrichment analysis (GSEA) of genes ranked by SCLC gene score, after exclusion of broadly essential genes, showing the top four most negatively enriched gene sets.

enriched. This suggests that SCLC cell lines may have an enhanced sensitivity towards perturbations in several key metabolic pathways.

SCLC cells exhibit increased sensitivity to Dhodh inhibition

We decided to focus on the gene *Dhodh*, which was the top candidate among the list of SCLC-specific depleted genes (**Fig. 3B, 4C**). *Dhodh* encodes dihydroorotate dehydrogenase, an enzyme that catalyzes the conversion of dihydroorotate to orotate, a key step in the *de novo* pyrimidine synthesis pathway (**Fig. 4A**). The other steps in this pathway are catalyzed by two multifunctional proteins encoded by the genes *Cad* and *Umps* (**Fig. 4A**). We observed that both *Cad* and *Umps* were also specifically depleted in SCLC compared with LUAD and PDAC (**Fig. 4B, D**). This confirms that SCLC cells specifically display increased sensitivity towards disruption of this pathway.

To directly test the effect of inhibition of Dhodh in SCLC, we utilized two different chemical inhibitors of Dhodh, leflunomide (Davis et al., 1996) and brequinar (Chen et al., 1986). SCLC cell lines displayed significantly higher sensitivity to both leflunomide and brequinar *in vitro*, compared with LUAD and PDAC cell lines (**Fig. 4E-F**). Furthermore, the effect of Dhodh inhibition could be rescued by supplementing the cell culture medium with uridine, which can be taken up by cells to bypass the requirement for *de novo* synthesis of uridine (**Fig. 4E-F**). This confirmed that the observed suppression of growth was a result of inhibition of uridine synthesis, rather than other off-target effects of either drug. Collectively, these results demonstrate that inhibition of Dhodh, and by extension the *de novo* uridine synthesis pathway, may be a feasible strategy to specifically target SCLC.
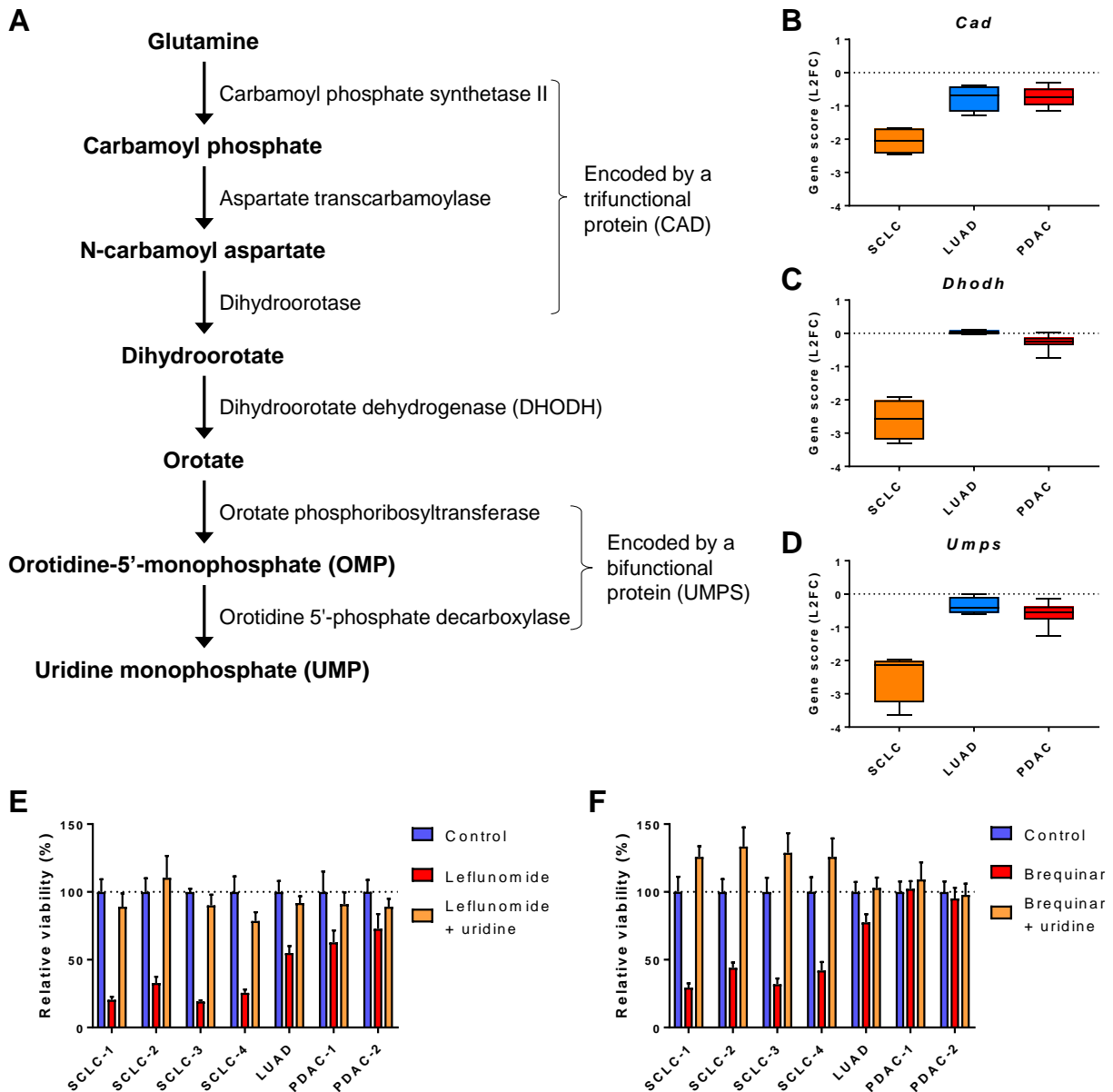
Figure 4: SCLC cells exhibit increased sensitivity to Dhodh inhibition.

**(A)** The *de novo* uridine biosynthesis pathway.

(B-D) Gene scores for the indicated genes for SCLC (n = 4), LUAD (n = 4) and PDAC (n = 8). Data are presented as median gene scores, with boxes denoting the interquartile range and bars denoting the range.

(E-F) Quantification of viable cell counts for each cell line following treatment with leflunomide ± uridine **(E)**, or with brequinar ± uridine **(F)**, as measured using the CellTiter-Glo assay. Results for each cell line are normalized to the control, untreated samples. Data are presented as means of four technical replicates, with error bars denoting the standard deviation.

Dhodh inhibition suppresses growth of SCLC tumors *in vivo*

DHODH inhibition has been demonstrated to be a feasible strategy for targeting

acute myeloid leukemia (AML) in preclinical models (Sykes et al., 2016). To test

whether inhibition of Dhodh affected SCLC tumor growth in an *in vivo* setting, we

utilized an intrasplenic transplant model of SCLC, which models metastatic colonization

of the liver (Khanna and Hunter, 2005), a frequent site of metastasis in SCLC patients

as well as in late-stage autochthonous SCLC tumor-bearing animals. At two weeks

post-transplantation, a time point at which small tumors are visible in the liver, we began

treatment of the animals with 50 mg/kg of brequinar every three days (**Fig. 5A**). This

dose has been previously shown to be well tolerated in animals for up to 72 days

(Sykes et al., 2016).

When we assessed tumor burden by *in vivo* bioluminescence imaging 8 days

after initiation of treatment, we observed a significant attenuation of liver tumor growth in

brequinar-treated animals compared with vehicle-treated animals (**Fig. 5B**). Animals

were sacrificed after approximately three weeks of treatment, when vehicle-treated

animals were moribund. At this time point, brequinar-treated animals had significantly

lower liver tumor burden compared with vehicle-treated animals (**Fig. 5C**). Importantly,

brequinar treatment was well-tolerated by the animals, with no significant changes in

body weight throughout the duration of treatment (data not shown). Therefore, these

data demonstrate that brequinar treatment strongly suppresses growth of SCLC tumors

in vivo, validating DHODH as a candidate therapeutic target for treating SCLC.
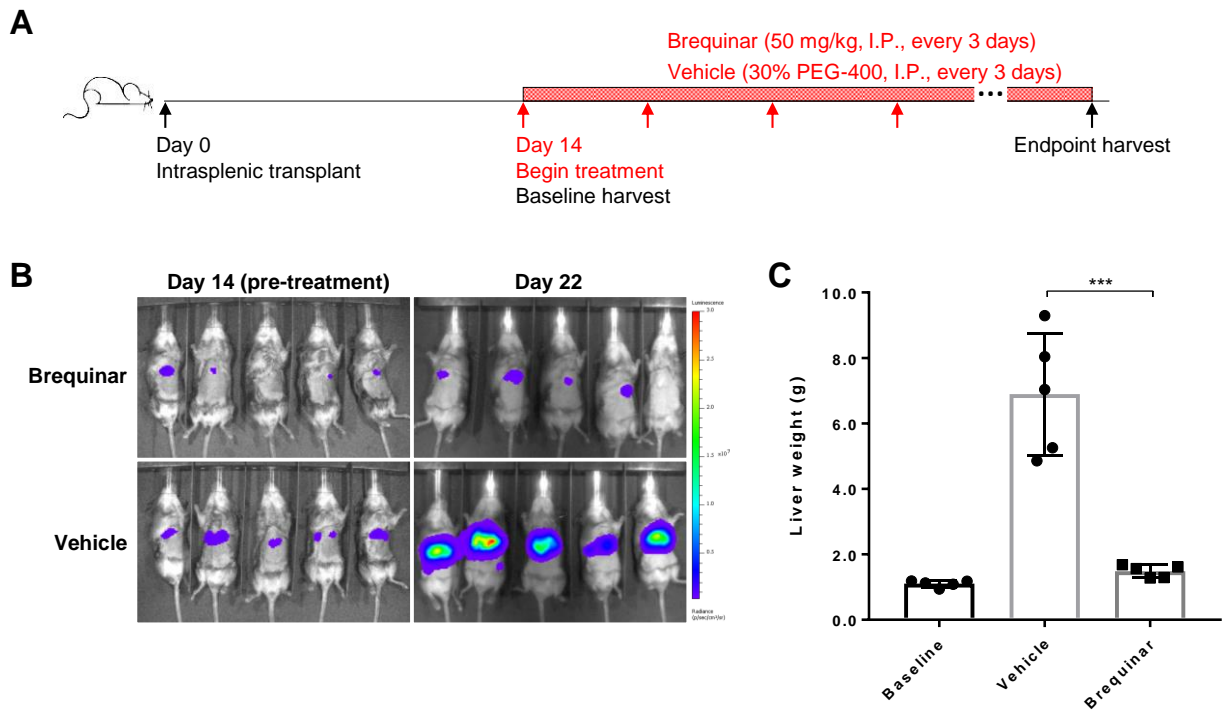
Figure 5: Dhodh inhibition suppresses growth of SCLC tumors *in vivo*.

(A) Dosing strategy for *in vivo* experiments.

(B) Representative images from *in vivo* bioluminescence imaging of tumor-bearing animals before commencement of drug treatment (left) and following 8 days of the indicated treatments (right).

(C) Quantification of liver tumor burden in animals following the indicated treatments, as measured by liver weight following necropsy. Baseline data were obtained from a separate cohort of animals that was sacrificed 2 weeks post-transplantation, prior to commencement of treatment. Treated animals were sacrificed approximately 3 weeks after commencement of treatment. Data are presented as means, with error bars denoting the 95% confidence intervals. *** p<0.001, two-tailed Student's *t*-test.

144

DISCUSSION

Genetic screens enable the large-scale perturbation and functional analysis of genes for unbiased identification of candidate genes involved in specific functions or phenotypes. In cancer, such screens are often performed to identify novel cancer type-specific vulnerabilities that can be exploited for therapeutic targeting, as a complementary approach to large-scale cancer genome sequencing studies. In this study, we have utilized a custom sgRNA library that focuses on therapeutically relevant gene targets to perform genetic screens in a panel of SCLC cell lines, uncovering previously unappreciated metabolic vulnerabilities in this cancer type. Using both genetic and pharmacological approaches, we have shown that SCLC cells are unusually sensitive to disruption of the *de novo* pyrimidine biosynthesis pathway via the inhibition of Dhodh, a key enzyme in this pathway. Finally, we have demonstrated that pharmacological inhibition of Dhodh strongly suppresses SCLC tumor growth *in vivo*, thereby validating DHODH as a novel candidate therapeutic target for the treatment of SCLC.

Large-scale CRISPR-based genetic screens have been performed in a variety of cancer cell lines to identify cancer specific sensitivities (Hart et al., 2015; Tzelepis et al., 2016; Wang et al., 2017). Hart and colleagues utilized data from these screens to perform a pan-screen analysis in an attempt to define reference sets of "core essential genes" and nonessential genes across all cell lines (Hart et al., 2017). However, this data set is heavily skewed towards hematopoietic cancers such as AML and lymphoma, with no lung cancer cell lines represented in the analysis. Meyers and colleagues have performed a much broader set of CRISPR screens across 342 human cancer cell lines

that included 47 lung cancer cell lines, of which only two were SCLC cell lines. To address this shortfall, we have performed CRISPR screens in murine SCLC cell lines derived from a well-established genetically engineered mouse model (GEMM) of SCLC, in which tumors are initiated by the conditional deletion of *Trp53* and *Rb1* in the lung epithelium (Meuwissen et al., 2003). This model has been shown to recapitulate many of the key features of human SCLC, and therefore represents a valid model for therapeutic discovery. Because of the lower background of mutations in murine SCLC compared with human SCLC (McFadden et al., 2014), we reasoned that this would provide a cleaner background for identifying candidate targets from the screen. In addition, we chose to use a focused sgRNA library targeting 5,000 potentially druggable genes rather than a whole-genome library to increase our confidence in identifying true hits from the screen.

Analysis of the results of our screen in parallel with screens in murine LUAD and PDAC cell lines enabled us to identify previously unknown SCLC-specific vulnerabilities, including sensitivity towards disruption of several key metabolic processes. None of these metabolic genes are mutated at a significant frequency in human SCLC tumors. Furthermore, gene expression analysis in these cell lines showed that these genes are expressed at similar levels across SCLC, LUAD and PDAC cell lines (data not shown). This highlights the utility of genetic screens for uncovering functional targets that would otherwise not have been identified by genomic or transcriptomic analyses.

The sensitivity of SCLC cells to disruption of the TCA cycle and oxidative phosphorylation has not been previously described. However, it is known that pRB plays a key role in regulating multiple metabolic processes, including oxidative

146

metabolism and the TCA cycle (Nicolay and Dyson, 2013). Nicolay and colleagues have demonstrated that acute loss of *Rb1* in mice results in a decrease in mitochondrial protein levels in both lung and colon tissue. Furthermore, loss of *RB1* in human hTERT-RPE1 cells leads to decreased mitochondrial mass, reduced mitochondrial function, and enhanced sensitivity to mitochondrial stress (Nicolay et al., 2015). Because our murine SCLC cell lines are derived from tumors initiated by deletion of *Rb1* and *Trp53*, this may explain the increased sensitivity towards disruption of mitochondrial processes that we observe in these cells compared with LUAD and PDAC cells. Given that *RB1* is inactivated in over 90% of human SCLC (George et al., 2015), this is likely to be the case in human SCLC cells as well.

We also found that disruption of the pyrimidine biosynthesis pathway via inhibition of Dhodh strongly suppressed SCLC growth, whereas LUAD and PDAC appeared to be relatively resistant to Dhodh inhibition. Interestingly, DHODH inhibition by brequinar or leflunomide has been studied in other cancers as well. DHODH inhibition by brequinar resulted in reduced tumor burden in human and mouse models of AML by triggering differentiation of leukemic myeloblast cells (Sykes et al., 2016). DHODH inhibition by leflunomide was shown to cooperate with BRAF(V600E) inhibition to reduce tumor growth in a xenograft model of melanoma (White et al., 2011). Similarly, combining both leflunomide and doxorubicin treatments led to significant tumor regression in a xenograft model of triple-negative breast cancer (Brown et al., 2017). In addition, *PTEN*-deficient breast cancer, glioblastoma and prostate cancer cell lines display enhanced sensitivity to DHODH inhibition by leflunomide (Mathur et al., 2017).

Leflunomide is an immunosuppressive drug that has been approved for the treatment of rheumatoid arthritis and psoriatic arthritis. Brequinar has previously been tested in phase II clinical trials involving patients with advanced lung cancer (Maroun et al., 1993). Although the drug failed to advance past phase II trials due to significant toxicity, a small proportion of patients with previously-treated SCLC exhibited partial response to treatment. It is possible that changes to the dosing schedule, which has been shown to significantly affect tolerance to the drug in mice (Sykes et al., 2016), could eliminate some of these toxicities in patients. The striking efficacy of brequinar in treating AML (Sykes et al., 2016), as well as our results in SCLC, justifies further evaluation of the drug in patients.

In summary, we have utilized a CRISPR-mediated genetic screening approach in SCLC cell lines, allowing us to identify vulnerabilities in several key metabolic pathways in SCLC. Through this, we have validated pharmacological inhibition of the pyrimidine biosynthesis pathway as a promising approach for the treatment of SCLC. Further work is currently in progress to establish the efficacy of Dhodh inhibition in both autochthonous murine models and patient-derived xenograft (PDX) models of SCLC. Together, these findings highlight a previous unappreciated link between SCLC and metabolism, which opens the door for developing novel approaches to target this deadly disease.

MATERIALS AND METHODS

Design and cloning of druggable genome sgRNA library

The list of genes included in our library were selected from the Drug-Gene Interaction Database (Griffith et al., 2013; Wagner et al., 2016). Human gene names were first updated using the HUGO Gene Nomenclature Committee (HGNC) multi-symbol checker tool, then converted to the corresponding mouse orthologs using the HGNC Comparison of Orthology Predictions tool (accessed 7/21/2016) (Eyre et al., 2006; Gray et al., 2015; Wright et al., 2005). sgRNA sequences were obtained from the Brie whole-genome sgRNA library, or designed using the Broad Institute sgRNA Designer tool (Doench et al., 2016). Non-targeting control sgRNA sequences were obtained from the Asiago whole-genome sgRNA library (Doench et al., 2016).

The sgRNA library was cloned into lentiGuide-Puro (Addgene #52963; lentiGuide-Puro was a gift from Feng Zhang), using a previously described protocol (Joung et al., 2017). In brief, 5' and 3' flanking adapter sequences, corresponding to the U6 promoter and tracrRNA sequence respectively, were appended to sgRNA sequences. The oligo library was synthesized by Twist Bioscience, PCR amplified with Oligo-Fwd and Oligo-Knockout-Rev primers using PfuUltra II Fusion HS DNA Polymerase (Agilent Technologies #600670), purified using the QIAquick PCR Purification Kit (QIAGEN #28104), then assembled into the BsmBI-digested lentiGuide-Puro vector using the Gibson Assembly Master Mix (New England Biolabs #E2611S). A total of 800 ng of assembled plasmid DNA was electroporated into Endura electrocompetent cells (Lucigen #60242), then plated onto LB-ampicillin plates overnight at 37°C. The total number of colonies was quantified to ensure a

149

representation of >100x (>2 million colonies for a 20,000-sgRNA library). Plasmids were isolated and purified using the QIAGEN Plasmid Plus Maxi Kit (QIAGEN #12963).

To determine sgRNA distribution in the cloned library, the sgRNA target region was amplified with NGS-Lib-Fwd (primers 1-10) and NGS-Lib-KO-Rev (primer 1) primers, size-selected in a 2% agarose gel, purified using the QIAquick Gel Extraction Kit (QIAGEN #28704), then submitted for sequencing on an Illumina HiSeq 2500 system (100-nt single end reads) at the Whitehead Institute Genome Technology Core.

Cell lines and cell culture

HEK293T cells were maintained in DMEM (Corning #10-013-CV) supplemented with 10% fetal bovine serum, 2 mM L-glutamine (Gibco #25030) and 50 µg/mL gentamicin (Gibco #15710). Lentiviral vectors were generated in HEK293T cells. In brief, cells were plated 1 day before transfection, then co-transfected with lentiviral constructs and packaging plasmids psPAX2 and pMD2.G (Addgene #12260 and #12259; both plasmids were gifts from Didier Trono). Viral supernatant was harvested 48 and 72 hours after transfection, then frozen at -80°C.

Murine SCLC cell lines used for the screens were generated from murine SCLC tumors isolated from $Trp53^{flox/flox}$; $Rb1^{flox/flox}$; $Rosa26^{+/+}$ or $Trp53^{flox/flox}$; $Rb1^{flox/flox}$; $Rosa26^{LSL-tdTomato/+}$ mice as previously described (Dooley et al., 2011). Cell lines were maintained on Matrigel-coated plates (Corning #356231, diluted to 50 ug/mL in HBSS) in DMEM/F-12 with HEPES (Gibco #11330), supplemented with 10% fetal bovine serum, 1x non-essential amino acids (Sigma #M7145) and 50 U/mL penicillin / 50 µg/mL streptomycin (Corning #30-002-CI). Cell lines were transduced with lentiCas9-

Blast virus (Addgene #52962; plasmid was a gift from Feng Zhang), then selected with 20 μg/mL Blasticidin S (Gibco #A11139) for 7 days.

For intrasplenic transplants, SCLC cell lines were transduced with a lentivirus expressing firefly luciferase and puromycin resistance, then selected with puromycin for 5 days. The lentiviral vector was generated using constructs generated for the Gibson assembly-based modular assembly platform (GMAP), as described previously (Akama-Garren et al., 2016). In brief, the LV 1–5 lentiviral backbone was assembled with pPGK pA, luciferase gA, pEFS pB, and puroR gB using Gibson assembly.

Infection of cells with sgRNA library

All transductions of SCLC cell lines were performed by spinfection. Cells were plated at a concentration of 2 million cells/well in 12-well plates, together with the appropriate amount of viral supernatant and 10 μg/mL polybrene (Millipore #TR-1003-G) in a total volume of 2 mL per well, then centrifuged at 1,000 RPM (approximately 200 x g) for 2 hours at room temperature. Cells were then incubated at 37°C overnight.

To ensure that most cells harbor single integration events, a multiplicity of infection (M.O.I.) of around 0.3 was used for transduction of cell lines with the sgRNA library (Joung et al., 2017). To determine the viral titer for each cell line, cells were transduced with decreasing volumes of viral supernatant, selected with 1 μg/mL puromycin (Gibco #A11138) for two days, then quantified and normalized to a control, unselected sample.

For the actual screen, a minimum of 20 million cells was maintained at each step to ensure a coverage of >1,000 cells per sgRNA. For each cell line, 200 million cells

were transduced with the sgRNA library in 8 x 12-well plates at an M.O.I. of around 0.3, then pooled into 8 x 10-cm dishes the next day. 48 hours post-transduction, 100 million cells were snap-frozen in liquid nitrogen, then stored at -80ºC (for the pre-selection, PD0 population), while 100 million cells were selected with 1 µg/mL puromycin for two days. Subsequently, cells were passaged every two days, and at least 20 million cells were plated after each passage. After 13 population doublings (PD13), at least 20 million cells were snap-frozen in liquid nitrogen, then stored at -80ºC. Three technical replicates of the screen were performed for each of the four cell lines, for a total of 12 independent replicates.

Genomic DNA isolation

Cell pellets were resuspended in 6 mL NK cell lysis buffer (50 mM Tris HCl, 50 mM EDTA, 1% SDS, pH 8) per 30-50 million cells. 30 µl of 20 mg/mL proteinase K (QIAGEN #19131) was added to the sample, which was then rotated overnight at 55ºC. The next day, 30 µL of 10 mg/mL RNase A (QIAGEN #19101) was added to the sample, which was then inverted 25 times and incubated for 30 minutes at 37ºC. Samples were cooled on ice, then 2 mL of pre-chilled 7.5 M ammonium acetate was added. Samples were vortexed at high speed for 20 seconds, then centrifuged at 4,000 x g for 10 minutes. The supernatant was decanted into a new tube, 6 mL of 100% isopropanol was added to the tube, then the sample was inverted 50 times and centrifuged at 4,000 x g for 10 minutes. The supernatant was discarded, leaving the genomic DNA as a small white pellet. 6 mL of 70% ethanol was added to wash the pellet, then the tube was inverted 10 times and centrifuged at 4,000 x g for 1 minute. The supernatant was discarded, then

the DNA pellet was re-dissolved in 500 uL of TE buffer (Sigma #T9285) and incubated overnight at 55ºC.

Screen deconvolution

Integrated sgRNAs were amplified by PCR to attach sequencing adapters and barcodes, as described (Doench et al., 2016). In brief, a mix of 8 P5_XPR/LKO1 forward primers and a P7 index reverse primer was used to amplify sgRNAs from each genomic DNA sample, using the NEBNext High Fidelity PCR Master Mix (New England Biolabs #M0541L) with 22 cycles of amplification. Assuming that each cell contains approximately 6.6 pg of genomic DNA (Chen et al., 2015), a total of 160 µg of genomic DNA (corresponding to >20 million cells) was used as a template for PD13 samples, while 720 µg of genomic DNA (corresponding to >100 million cells) was used for PD0 samples. Amplified PCR products of around 350 bp in length were size-selected on a 2% agarose gel, then purified using the QIAquick Gel Extraction Kit (QIAGEN #28704). All samples were submitted for sequencing on an Illumina NextSeq system (75-nt single end reads) at the MIT BioMicro Center.

Data analysis

For each time point of each replicate (PD0 and PD13), sequencing read counts for each sgRNA were normalized to the total read count for that sample, followed by a log2 transformation. Next, the log2 fold change (L2FC) score for each sgRNA was calculated as the difference between the PD13 and PD0 log2-transformed counts. Scores for the three technical replicates were averaged to obtain sgRNA scores for each cell line. The

gene score for each gene was then calculated as the median of the sgRNA scores for the four sgRNAs targeting that gene. Finally, the median of the gene scores for all four SCLC cell lines was used as the overall L2FC for SCLC. The same analysis pipeline was used for LUAD and PDAC screens (data not shown in this paper) to ensure that results from all three screens were compatible.

Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) was performed using the default parameters. All analyses were restricted to the canonical pathways sub-collection (CP) under the curated gene sets collection (C2).

In vitro drug treatments

Leflunomide (Sigma #L5025) was dissolved in DMSO at a concentration of 100 mM. Brequinar (Tocris #6196) was dissolved in DMSO at a concentration of 10 mM. For dose response curves, cells were plated in 96-well assay plates (Corning #3903) (3,000 cells per well) with decreasing concentrations of either leflunomide or brequinar. For uridine rescue experiments, cells were plated as above, with 30 µM of leflunomide or 1 µM of brequinar, and 125 µM of uridine. After three days of respective treatments, the number of viable cells per well was measured using the CellTiter-Glo Luminescent Cell Viability Assay (Promega #G7570). Results were normalized to control, untreated samples. All assays were performed in four technical replicates.

In vivo drug dosing

For *in vivo* administration, brequinar was dissolved in 30% PEG-400 (Sigma #06855) in PBS, pH 8 at a concentration of 12.5 mg/mL. Animals were dosed with 50 mg/kg of

brequinar (4 μL/g) every three days. Control animals were dosed with the equivalent volume of 30% PEG-400 in PBS every three days.

Intrasplenic transplantation in mice

All animal studies were approved by the MIT Institutional Animal Care and Use Committee. Intrasplenic transplants were performed in B6129SF1/J hybrid mice (The Jackson Laboratory #101043). 100,000 cells in 50 μL of HBSS were transplanted per animal. For *in vivo* bioluminescence imaging, mice were anesthetized by isoflurane inhalation, administered with 150 mg/kg of D-luciferin (PerkinElmer #122799) by intraperitoneal injection, then imaged 10 minutes post-administration using the IVIS Spectrum In Vivo Imaging System (PerkinElmer). Visualization and quantification of bioluminescence signal was performed using Living Image (PerkinElmer).

Animals were dosed with brequinar beginning two weeks after transplantation. At the end of the experiment, animals were sacrificed, and livers were harvested and weighed to determine liver tumor burden.

ACKNOWLEDGEMENTS

SUPPLEMENTARY TABLES AND FIGURES

Supplementary Table 1: List of primary sources for drugged and druggable genes

All sources were obtained from Wagner et al., 2016

Sources for drugged genes:	Sources for druggable genes:
CancerCommons	Bader Lab
ChEMBL	Caris Molecular Intelligence
CIViC	dGENE
Clearity Foundation: Biomarkers	Guide To Pharmacology: Genes
Clearity Foundation: Clinical Trials	Foundation One Genes
DoCM	GO
DrugBank	Hopkins & Groom
Guide To Pharmacology: Interactions	MSK IMPACT
My Cancer Genome	Russ & Lampel
My Cancer Genome: Clinical Trials	
PharmGKB	
Targeted Agents in Lung Cancer	
TDG Clinical Trials	
Trends in the Exploitation of Novel Drug Targets (TEND)	
Therapeutic Target Database (TTD)	

Supplementary Table 2: Top 20 negatively enriched gene sets for SCLC

Name	Normalized enrichment score	Nominal p-val	FDR q-val
REACTOME_TRNA_AMINOACYLATION	-1.5596	0.0000	0.0000
KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	-1.5254	0.0000	0.0000
REACTOME_CYTOSOLIC_TRNA_AMINOACYLATION	-1.5237	0.0000	0.0000
REACTOME_APC_C_CDH1_MEDIATED_DEGRADATION_OF_CDC20_AND_OTHER_APC_C_CDH1_TARGETED_PROTEINS_IN_LATE_MITOSIS_EARLY_G1	-1.5224	0.0000	0.0000
REACTOME_MITOTIC_G1_G1_S_PHASES	-1.5176	0.0000	0.0000
REACTOME_CDT1_ASSOCIATION_WITH_THE_CDC6_ORC_ORIGIN_COMPLEX	-1.5168	0.0000	0.0000
REACTOME_3_UTR_MEDIATED_TRANSLATIONAL_REGULATION	-1.5144	0.0000	0.0000
KEGG_PROTEASOME	-1.5142	0.0000	0.0000
REACTOME_ASSEMBLY_OF_THE_PRE_REPLICATIVE_COMPLEX	-1.5114	0.0000	0.0000
REACTOME_AUTODEGRADATION_OF_CDH1_BY_CDH1_APC_C	-1.5112	0.0000	0.0000
REACTOME_G1_S_TRANSITION	-1.5110	0.0000	0.0000
REACTOME_CYCLIN_E_ASSOCIATED_EVENTS_DURING_G1_S_TRANSITION_	-1.5096	0.0000	0.0000
REACTOME_M_G1_TRANSITION	-1.5070	0.0000	0.0000
BIOCARTA_PROTEASOME_PATHWAY	-1.5067	0.0000	0.0000
REACTOME_SCFSKP2_MEDIATED_DEGRADATION_OF_P27_P21	-1.5050	0.0000	0.0000
REACTOME_S_PHASE	-1.5041	0.0000	0.0000
REACTOME_CROSS_PRESENTATION_OF_SOLUBLE_EXOGENOUS_ANTIGENS_ENDOSOMES	-1.5032	0.0000	0.0000
REACTOME_SCF_BETA_TRCP_MEDIATED_DEGRADATION_OF_EMI1	-1.4993	0.0000	0.0000
REACTOME_ORC1_REMOVAL_FROM_CHROMATIN	-1.4989	0.0000	0.0000
REACTOME_MITOCHONDRIAL_TRNA_AMINOACYLATION	-1.4977	0.0000	0.0000

Supplementary Table 3: Top 20 negatively enriched gene sets for LUAD

Name	Normalized enrichment score	Nominal p-val	FDR q-val
REACTOME_M_G1_TRANSITION	-1.6852	0.0000	0.0000
KEGG_PROTEASOME	-1.6823	0.0000	0.0000
REACTOME_VIF_MEDIATED_DEGRADATION_OF_APOBEC3G	-1.6789	0.0000	0.0000
REACTOME_APC_C_CDH1_MEDIATED_DEGRADATION_OF_CDC20_AND_OTHER_APC_C_CDH1_TARGETED_PROTEINS_IN_LATE_MITOSIS_EARLY_G1	-1.6771	0.0000	0.0000
REACTOME_CDT1_ASSOCIATION_WITH_THE_CDC6_ORC_ORIGIN_COMPLEX	-1.6767	0.0000	0.0000
REACTOME_CROSS_PRESENTATION_OF_SOLUBLE_EXOGENOUS_ANTIGENS_ENDOSOMES	-1.6753	0.0000	0.0000
REACTOME_AUTODEGRADATION_OF_CDH1_BY_CDH1_APC_C	-1.6735	0.0000	0.0000
REACTOME_P53_INDEPENDENT_G1_S_DNA_DAMAGE_CHECKPOINT	-1.6719	0.0000	0.0000
BIOCARTA_PROTEASOME_PATHWAY	-1.6688	0.0000	0.0000
REACTOME_APC_C_CDC20_MEDIATED_DEGRADATION_OF_MITOTIC_PROTEINS	-1.6633	0.0000	0.0000
REACTOME_REGULATION_OF_APOPTOSIS	-1.6631	0.0000	0.0000
REACTOME_REGULATION_OF_MRNA_STABILITY_BY_PROTEINS_THAT_BIND_AU_RICH_ELEMENTS	-1.6608	0.0000	0.0000
REACTOME_ASSEMBLY_OF_THE_PRE_REPLICATIVE_COMPLEX	-1.6605	0.0000	0.0000
REACTOME_AUTODEGRADATION_OF_THE_E3_UBIQUITIN_LIGASE_COP1	-1.6590	0.0000	0.0000
REACTOME_CDK_MEDIATED_PHOSPHORYLATION_AND_REMOVAL_OF_CDC6	-1.6569	0.0000	0.0000
REACTOME_METABOLISM_OF_MRNA	-1.6533	0.0000	0.0000
REACTOME_ER_PHAGOSOME_PATHWAY	-1.6525	0.0000	0.0000
REACTOME_DESTABILIZATION_OF_MRNA_BY_AUF1_HNRNP_D0	-1.6524	0.0000	0.0000
REACTOME_SCF_BETA_TRCP_MEDIATED_DEGRADATION_OF_EMI1	-1.6522	0.0000	0.0000
REACTOME_ANTIGEN_PROCESSING_CROSS_PRESENTATION	-1.6520	0.0000	0.0000

Supplementary Table 4: Top 20 negatively enriched gene sets for PDAC

Name	Normalized enrichment score	Nominal p-val	FDR q-val
REACTOME_ASSEMBLY_OF_THE_PRE_REPLICATIVE_COMPLEX	-1.6370	0.0000	0.0000
REACTOME_CDK_MEDIATED_PHOSPHORYLATION_AND_REMOVAL_OF_CDC6	-1.6317	0.0000	0.0000
REACTOME_APC_C_CDH1_MEDIATED_DEGRADATION_OF_CDC20_AND_OTHER_APC_C_CDH1_TARGETED_PROTEINS_IN_LATE_MITOSIS_EARLY_G1	-1.6303	0.0000	0.0000
REACTOME_VIF_MEDIATED_DEGRADATION_OF_APOBEC3G	-1.6292	0.0000	0.0000
REACTOME_CDT1_ASSOCIATION_WITH_THE_CDC6_ORC_ORIGIN_COMPLEX	-1.6287	0.0000	0.0000
REACTOME_P53_INDEPENDENT_G1_S_DNA_DAMAGE_CHECKPOINT	-1.6225	0.0000	0.0000
BIOCARTA_PROTEASOME_PATHWAY	-1.6208	0.0000	0.0000
REACTOME_M_G1_TRANSITION	-1.6152	0.0000	0.0000
REACTOME_SIGNALING_BY_WNT	-1.6148	0.0000	0.0000
REACTOME_REGULATION_OF_APOPTOSIS	-1.6135	0.0000	0.0000
KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	-1.6129	0.0000	0.0000
KEGG_PROTEASOME	-1.6123	0.0000	0.0000
REACTOME_CROSS_PRESENTATION_OF_SOLUBLE_EXOGENOUS_ANTIGENS_ENDOSOMES	-1.6123	0.0000	0.0000
REACTOME_SCF_BETA_TRCP_MEDIATED_DEGRADATION_OF_EMI1	-1.6096	0.0000	0.0000
REACTOME_DESTABILIZATION_OF_MRNA_BY_AUF1_HNRNP_D0	-1.6069	0.0000	0.0000
REACTOME_CELL_CYCLE_CHECKPOINTS	-1.6046	0.0000	0.0000
REACTOME_AUTODEGRADATION_OF_THE_E3_UBIQUITIN_LIGASE_COP1	-1.6044	0.0000	0.0000
REACTOME_APC_C_CDC20_MEDIATED_DEGRADATION_OF_MITOTIC_PROTEINS	-1.6015	0.0000	0.0000
REACTOME_REGULATION_OF_MITOTIC_CELL_CYCLE	-1.6007	0.0000	0.0000
REACTOME_METABOLISM_OF_RNA	-1.6000	0.0000	0.0000

Supplementary Table 5:
SCLC-specific negatively enriched gene sets with FDR q-val < 0.01

Name	Normalized enrichment score	Nominal p-val	FDR q-val
REACTOME_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT	-2.1636	0.0000	0.0000
REACTOME_PYRUVATE_METABOLISM_AND_CITRIC_ACID_TCA_CYCLE	-2.1344	0.0000	0.0000
KEGG_CITRATE_CYCLE_TCA_CYCLE	-2.0809	0.0000	0.0000
REACTOME_CITRIC_ACID_CYCLE_TCA_CYCLE	-2.0494	0.0000	0.0000
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT	-2.0196	0.0000	0.0000
KEGG_OXIDATIVE_PHOSPHORYLATION	-2.0118	0.0000	0.0000
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS_	-1.9861	0.0000	0.0000
KEGG_PARKINSONS_DISEASE	-1.9812	0.0000	0.0000
KEGG_HUNTINGTONS_DISEASE	-1.9536	0.0000	0.0000
REACTOME_METABOLISM_OF_VITAMINS_AND_COFACTORS	-1.9426	0.0000	0.0000
KEGG_BUTANOATE_METABOLISM	-1.8720	0.0013	0.0010
KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION	-1.8625	0.0000	0.0011
KEGG_GLUTATHIONE_METABOLISM	-1.8416	0.0000	0.0019
KEGG_ALZHEIMERS_DISEASE	-1.8323	0.0000	0.0023
KEGG_PROPANOATE_METABOLISM	-1.7961	0.0000	0.0062
KEGG_PYRIMIDINE_METABOLISM	-1.7827	0.0000	0.0086
REACTOME_DNA_REPAIR	-1.7781	0.0000	0.0091

**Supplementary Table 6: List of primers used for sgRNA library cloning**

Primer	Sequence	Source
Oligo-Fwd	GTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATA TCTTGTGGAAAGGACGAAACACC	Joung, Konermann et al., 2017
Oligo-Knockout-Rev	ACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAA CTTGCTATTTCTAGCTCTAAAAC	
NGS-Lib-Fwd-1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTTAAGTAGAGGCTTTAT ATATCTTGTGGAAAGGACGAAACACC	
NGS-Lib-Fwd-2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTATCATGCTTAGCTTTA TATATCTTGTGGAAAGGACGAAACACC	
NGS-Lib-Fwd-3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTGATGCACATCTGCTTT ATATATCTTGTGGAAAGGACGAAACACC	
NGS-Lib-Fwd-4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTCGATTGCTCGACGCTT TATATATCTTGTGGAAAGGACGAAACACC	
NGS-Lib-Fwd-5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTTCGATAGCAATTCGCT TTATATATCTTGTGGAAAGGACGAAACACC	
NGS-Lib-Fwd-6	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTATCGATAGTTGCTTGC TTTATATATCTTGTGGAAAGGACGAAACACC	
NGS-Lib-Fwd-7	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTGATCGATCCAGTTAGG CTTTATATATCTTGTGGAAAGGACGAAACACC	
NGS-Lib-Fwd-8	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTCGATCGATTTGAGCCT GCTTTATATATCTTGTGGAAAGGACGAAACACC	
NGS-Lib-Fwd-9	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTACGATCGATACACGAT CGCTTTATATATCTTGTGGAAAGGACGAAACACC	
NGS-Lib-Fwd-10	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTTACGATCGATGGTCCA GAGCTTTATATATCTTGTGGAAAGGACGAAACACC	
NGS-Lib-KO-Rev-1	CAAGCAGAAGACGGCATACGAGAT**TCGCCTTG**GTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCTCCGACTCG GTGCCACTTTTTCAA	
P5_XPR/LKO1_1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTTTGTGGAAAGGACGAA ACACCG	Broad Institute GPP Web

P5_XPR/LKO1_2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTCTTGTGGAAAGGACGA AACACCG	Portal (May 2015)
P5_XPR/LKO1_3	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTGCTTGTGGAAAGGACG AAACACCG	
P5_XPR/LKO1_4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTAGCTTGTGGAAAGGAC GAAACACCG	
P5_XPR/LKO1_5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTCAACTTGTGGAAAGGA CGAAACACCG	
P5_XPR/LKO1_6	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTTGCACCTTGTGGAAAG GACGAAACACCG	
P5_XPR/LKO1_7	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTACGCAACTTGTGGAAA GGACGAAACACCG	
P5_XPR/LKO1_8	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTGAAGACCCTTGTGGAA AGGACGAAACACCG	
P7 index 2	CAAGCAGAAGACGGCATACGAGAT**NNNNNNNN**GTGAC TGGAGTTCAGACGTGTGCTCTTCCGATCTTCTACTAT TCTTTCCCCTGCACTGT	

NNNNNNNN: P7 index sequence

Supplementary Table 7: List of P7 index sequences used in this study

Name	Sequence to include in primer	Name	Sequence to include in primer
A01	CGGTTCAA	A10	TACAGAGG
A02	GCTGGATT	B01	ATTGGATT
A03	TAACTCGG	B02	ATACTCGG
A04	TAACAGTT	B03	TATGAGAA
A05	ATACTCAA	B04	GCACAGTT
A06	GCTGAGAA	B05	CGTGGATT
A07	ATTGGAGG	B06	TAGTAGAA
A08	TAGTCTAA	B07	GCACGATT
A09	CGGTGACC	B08	CGGTAGCC
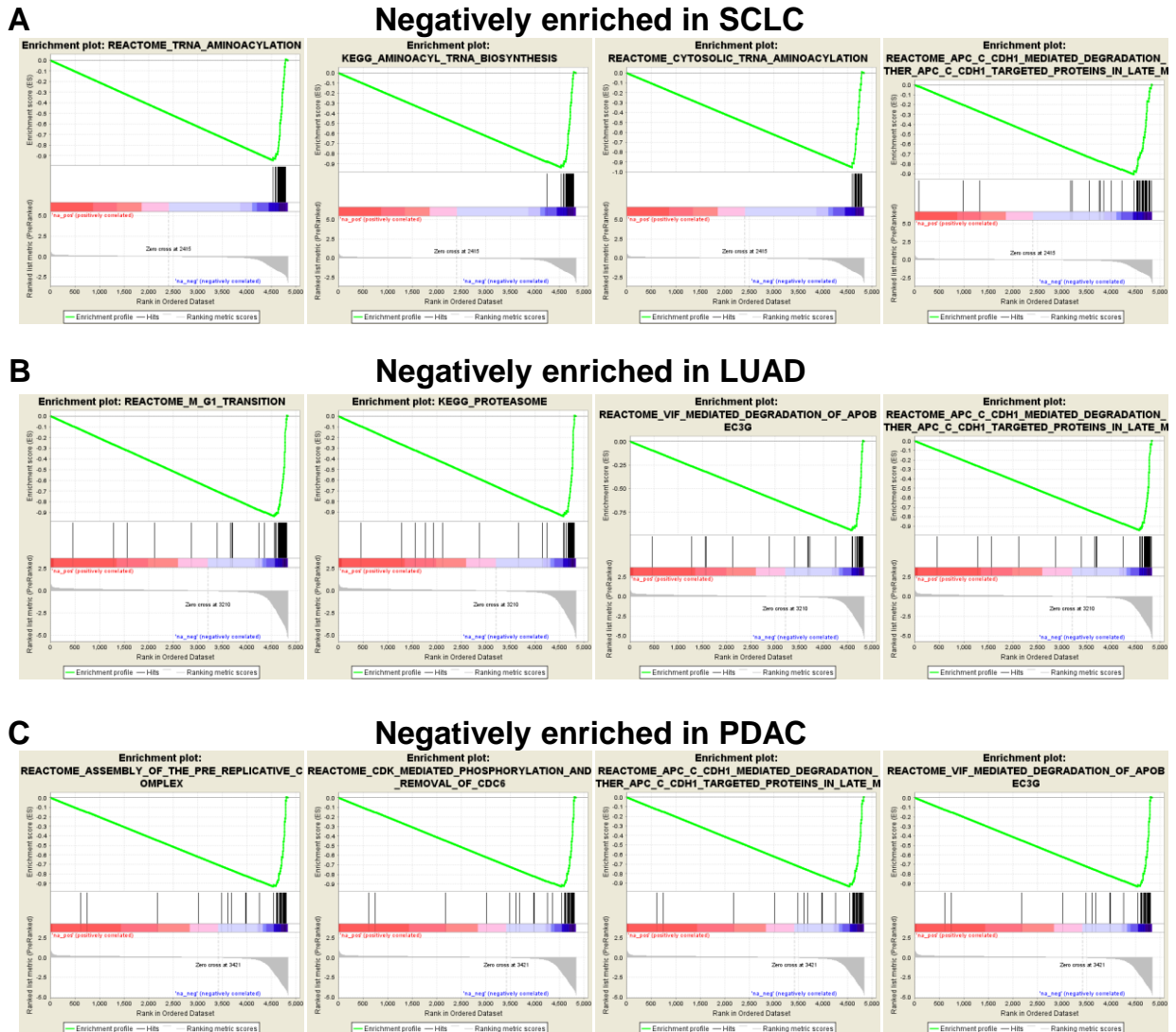
Supplementary Figure 1: Gene sets for essential cellular processes are negatively enriched in screens.
Gene set enrichment analysis (GSEA) of all genes represented in the sgRNA library, ranked by gene scores for SCLC **(A)**, LUAD **(B)**, or PDAC **(C)**, before exclusion of broadly essential genes.

REFERENCES

Akama-Garren, E.H., Joshi, N.S., Tammela, T., Chang, G.P., Wagner, B.L., Lee, D.-Y., Rideout III, W.M., Papagiannakopoulos, T., Xue, W., and Jacks, T. (2016). A Modular Assembly Platform for Rapid Generation of DNA Constructs. Sci. Rep. *6*, 16836.

Brown, K.K., Spinelli, J.B., Asara, J.M., and Toker, A. (2017). Adaptive reprogramming of De novo pyrimidine synthesis is a metabolic vulnerability in triple-negative breast cancer. Cancer Discov. *7*, 391–399.

Byers, L.A., and Rudin, C.M. (2015). Small cell lung cancer: Where do we go from here? Cancer *121*, 664–672.

Califano, R., Abidin, A.Z., Peck, R., Faivre-Finn, C., and Lorigan, P. (2012). Management of Small Cell Lung Cancer. Drugs *72*, 471–490.

Chen, S., Sanjana, N.E., Zheng, K., Shalem, O., Lee, K., Shi, X., Scott, D.A., Song, J., Pan, J.Q., Weissleder, R., et al. (2015). Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis. Cell *160*, 1246–1260.

Chen, S.F., Ruben, R.L., and Dexter, D.L. (1986). Mechanism of action of the novel anticancer agent 6-fluoro-2-(2'-fluoro-1,1'-biphenyl-4-yl)-3-methyl-4-quinolinecarbo xylic acid sodium salt (NSC 368390): inhibition of de novo pyrimidine nucleotide biosynthesis. Cancer Res. *46*, 5014–5019.

Davis, J.P., Cain, G.A., Pitts, W.J., Magolda, R.L., and Copeland, R.A. (1996). The immunosuppressive metabolite of leflunomide is a potent inhibitor of human dihydroorotate dehydrogenase. Biochemistry *35*, 1270–1273.

Demedts, I.K., Vermaelen, K.Y., and Van Meerbeeck, J.P. (2010). Treatment of extensive-stage small cell lung carcinoma: Current status and future prospects. Eur. Respir. J. *35*, 202–215.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

Dooley, A.L., Winslow, M.M., Chiang, D.Y., Banerji, S., Stransky, N., Dayton, T.L., Snyder, E.L., Senna, S., Whittaker, C.A., Bronson, R.T., et al. (2011). Nuclear factor I/B is an oncogene in small cell lung cancer. Genes Dev. *25*, 1470–1475.

Echeverri, C.J., Beachy, P.A., Baum, B., Boutros, M., Buchholz, F., Chanda, S.K., Downward, J., Ellenberg, J., Fraser, A.G., Hacohen, N., et al. (2006). Minimizing the risk of reporting false positives in large-scale RNAi screens. Nat. Methods *3*, 777–779.

Eyre, T.A., Wright, M.W., Lush, M.J., and Bruford, E.A. (2006). HCOP: a searchable database of human orthology predictions. Brief. Bioinform. *8*, 2–5.

George, J., Lim, J.S., Jang, S.J., Cun, Y., Ozretić, L., Kong, G., Leenders, F., Lu, X., Fernández-Cuesta, L., Bosco, G., et al. (2015). Comprehensive genomic profiles of small cell lung cancer. Nature *524*, 47–53.

Gray, K.A., Yates, B., Seal, R.L., Wright, M.W., and Bruford, E.A. (2015). Genenames.org: the HGNC resources in 2015. Nucleic Acids Res. *43*, D1079–D1085.

Griffith, M., Griffith, O.L., Coffman, A.C., Weible, J. V., McMichael, J.F., Spies, N.C., Koval, J., Das, I., Callaway, M.B., Eldred, J.M., et al. (2013). DGIdb: mining the druggable genome. Nat. Methods *10*, 1209–1210.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell *163*, 1515–1526.

Hart, T., Tong, A.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A., et al. (2017). Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. G3 (Bethesda). *7*, 2719–2727.

Hopkins, A.L., and Groom, C.R. (2002). The druggable genome. Nat. Rev. Drug Discov. *1*, 727–730.

Joung, J., Konermann, S., Gootenberg, J.S., Abudayyeh, O.O., Platt, R.J., Brigham, M.D., Sanjana, N.E., and Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nat. Protoc. *12*, 828–863.

Khanna, C., and Hunter, K. (2005). Modeling metastasis in vivo. Carcinogenesis *26*, 513–523.

Maroun, J., Ruckdeschel, J., Natale, R., Morgan, R., Dallaire, B., Sisk, R., and Gyves, J. (1993). Multicenter phase II study of brequinar sodium in patients with advanced lung cancer. Cancer Chemother. Pharmacol. *32*, 64–66.

Mathur, D., Stratikopoulos, E., Ozturk, S., Steinbach, N., Pegno, S., Schoenfeld, S., Yong, R., Murty, V. V., Asara, J.M., Cantley, L.C., et al. (2017). PTEN Regulates Glutamine Flux to Pyrimidine Synthesis and Sensitivity to Dihydroorotate Dehydrogenase Inhibition. Cancer Discov. *7*, 380–390.

McFadden, D.G., Papagiannakopoulos, T., Taylor-Weiner, A., Stewart, C., Carter, S.L., Cibulskis, K., Bhutkar, A., McKenna, A., Dooley, A., Vernon, A., et al. (2014). Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. Cell *156*, 1298–1311.

Meuwissen, R., Linn, S.C., Linnoila, R.I., Zevenhoven, J., Mooi, W.J., and Berns, A. (2003). Induction of small cell lung cancer by somatic inactivation of both Trp53 and Rb1 in a conditional mouse model. Cancer Cell *4*, 181–189.

Nicolay, B.N., and Dyson, N.J. (2013). The multiple connections between pRB and cell metabolism. Curr. Opin. Cell Biol. *25*, 735–740.

Nicolay, B.N., Danielian, P.S., Kottakis, F., Lapek, J.D., Sanidas, I., Miles, W.O., Dehnad, M., Tschöp, K., Gierut, J.J., Manning, A.L., et al. (2015). Proteomic analysis of pRb loss highlights a signature of decreased mitochondrial oxidative phosphorylation. Genes Dev. *29*, 1875–1889.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M. a, Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. *102*, 15545–15550.

Sykes, D.B., Kfoury, Y.S., Mercier, F.E., Wawer, M.J., Law, J.M., Haynes, M.K., Lewis, T.A., Schajnovitz, A., Jain, E., Lee, D., et al. (2016). Inhibition of Dihydroorotate Dehydrogenase Overcomes Differentiation Blockade in Acute Myeloid Leukemia. Cell *167*, 171–186.e15.

Tzelepis, K., Koike-Yusa, H., De Braekeleer, E., Li, Y., Metzakopian, E., Dovey, O.M., Mupo, A., Grinkevich, V., Li, M., Mazan, M., et al. (2016). A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. Cell Rep. *17*, 1193–1205.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558.

Wagner, A.H., Coffman, A.C., Ainscough, B.J., Spies, N.C., Skidmore, Z.L., Campbell, K.M., Krysiak, K., Pan, D., McMichael, J.F., Eldred, J.M., et al. (2016). DGIdb 2.0: mining clinically relevant drug-gene interactions. Nucleic Acids Res *44*, D1036-44.

Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. Science *350*, 1096–1101.

Wang, T., Yu, H., Hughes, N.W., Liu, B., Kendirli, A., Klein, K., Chen, W.W., Lander, E.S., and Sabatini, D.M. (2017). Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. Cell *168*, 890–903.e15.

White, R.M., Cech, J., Ratanasirintrawoot, S., Lin, C.Y., Rahl, P.B., Burke, C.J., Langdon, E., Tomlinson, M.L., Mosher, J., Kaufman, C., et al. (2011). DHODH modulates transcriptional elongation in the neural crest and melanoma. Nature *471*, 518–522.

Wright, M.W., Eyre, T.A., Lush, M.J., Povey, S., and Bruford, E.A. (2005). HCOP: The HGNC comparison of orthology predictions search tool. Mamm. Genome *16*, 827–828.

CHAPTER 4

DISCUSSION AND FUTURE DIRECTIONS

In this thesis, I have utilized the CRISPR-Cas9 system in two complementary ways to advance our understanding of the molecular mechanisms involved in SCLC initiation, progression and maintenance. In Chapter 2, I have successfully adapted the CRISPR-Cas9 system for use in GEMMs of SCLC, demonstrating its use for carrying out functional validation and modeling of candidate tumor suppressor genes in SCLC, as well as for assessing the phenotypic consequences of loss of gene function in these models. Recent large-scale genomic sequencing of human SCLC tumors has identified multiple genes that are mutated at significant frequencies (George et al., 2015), but which have not been functionally investigated in the context of SCLC. The *in vivo* CRISPR-based system that we have established enables more rapid modeling of these candidate genes of interest to identify functionally important genetic alterations. Given the lack of mutations in common oncogenic drivers in SCLC, such as alterations in the RTK/RAS/RAF pathway that frequently occur in other cancers, such studies should aid in the identification of important mediators of tumor progression in SCLC, which could uncover novel approaches for treating the disease.

In Chapter 3, we have taken a CRISPR-based genetic screening approach in murine SCLC cell lines to identify SCLC-specific genetic vulnerabilities. This allowed us to uncover several novel metabolic sensitivities in these cells, as well as to validate the importance of one particular pathway, the *de novo* pyrimidine synthesis pathway, as a promising therapeutic target in SCLC. Additional experiments involving pharmacological inhibition of this pathway are currently ongoing, both in autochthonous murine models and in patient-derived xenograft (PDX) models of SCLC. These studies will allow us to further validate the efficacy of targeting this pathway in SCLC. Furthermore, several

other metabolic pathways, such as the TCA cycle and oxidative phosphorylation, have been identified from the results of our screen, but have yet to be functionally validated. Future experiments will enable us to validate these preliminary results.

Rapid functional validation of candidate genes

As discussed in Chapter 1, both cancer genome sequencing studies and functional genetic screens in cancer cells have generated long lists of candidate genes of interest in various cancers, but functional validation of these genes in relevant preclinical models remains a huge bottleneck. The traditional approach of generating germline targeted alleles in mice for each gene of interest is not suited for high-throughput interrogation of large numbers of genes. Somatic genome editing approaches, such as those involving the use of CRISPR-Cas9 as described in Chapter 2 and elsewhere (Chiou et al., 2015; Dow et al., 2015; Roper et al., 2017; Sánchez-Rivera et al., 2014), are much more suited for this purpose, since one only needs to generate a new viral vector, rather than a new mouse, for each gene of interest, significantly reducing both time and cost.

However, one downside of this approach is that it is primarily useful for generating loss-of-function mutations in genes. This makes it especially useful for interrogating candidate tumor suppressor genes, as described in Chapter 2, but less useful for studying gain-of-function oncogenic events. In addition, because not all editing events generate loss-of-function, frameshift mutations, phenotypes resulting from CRISPR-based approaches, such as changes in tumor burden or rate of tumor progression, may be confounded by the presence of tumors that retain wild-type gene

171

function. This is not usually an issue for strong phenotypes that are positively selected

for, but may present a problem for subtler phenotypes. Therefore, I expect that such

approaches will complement, rather than replace, the traditional approach of using

conditional targeted alleles. For example, one can first use CRISPR-Cas9-based

approaches to quickly narrow down a long list of genes into a shorter list of candidates

that display functional phenotypes *in vivo*, then generate new targeted alleles for this

more manageable list of strong candidates.

In addition, variants of Cas9 that mediate transcriptional activation or repression

rather than gene editing have begun to be used *in vivo* (Liao et al., 2017; Wangensteen

et al., 2017). These approaches overcome some of the limitations discussed earlier. For

example, the use of CRISPRi bypasses the issue of non-frameshift mutations, while the

use of CRISPRa enables gain-of-function phenotypes to be studied. Further

development of these tools for more efficient *in vivo* activity will expand the utility of

CRISPR-based somatic approaches for modeling cancer.

Significance of low-frequency genetic alterations in cancer

Analyses of cancer genome sequencing studies have confirmed the key roles of

major oncogenes and tumor suppressor genes across different cancers, including well-

known oncogenes such as *KRAS*, *EGFR*, *ERBB2* and *PIK3CA*, as well as tumor

suppressor genes such as *TP53*, *RB1*, *APC* and *NF1* (Bailey et al., 2018). What has

also been increasingly appreciated, however, is the importance of genes that are

altered at lower frequencies within each cancer type, which have been described as

"hills" in the mutational landscapes of tumors, in contrast to the "mountains" that

represent frequently mutated genes (Wood et al., 2007). In certain cases, low-frequency

mutations can confer significantly different phenotypes to tumors. For example,

activation of MYC in a murine model of SCLC results in the development of the variant

subtype of SCLC, which confers sensitivity of these tumors to Aurora kinase inhibitors

(Mollaoglu et al., 2017). In another example in lung adenocarcinoma (LUAD), mutations

in *KEAP1*, which is a key regulator of the antioxidant response in cells that is mutated in

approximately 20% of *KRAS*-mutant LUAD tumors, were found to cause these tumors

to be dependent on increased glutaminolysis, which could be targeted by

pharmacological inhibition of glutaminase (Romero et al., 2017).

Clinically, the profiling of tumors from individual patients forms the basis of

precision medicine. This has already been utilized to identify patients that may benefit

from particular targeted therapies, based on the presence of specific driver mutations in

their tumors. For example, LUAD patients that harbor *EGFR* mutations in their tumors

are treated with EGFR inhibitors such as erlotinib and gefitinib; these drugs do not show

clinical benefit in LUAD patients with tumors driven by other oncogenic drivers such as

KRAS. Likewise, patients with *EML4-ALK*-driven LUAD are treated with ALK inhibitors

such as crizotinib and ceritinib.

Collectively, these examples validate the importance of studying low-frequency

genetic events, since they may reveal previously-unappreciated fitness defects in these

subsets of tumors. The use of CRISPR-Cas9 with *in vivo* models of cancers, such as

the approach described in Chapter 2, significantly increases the feasibility of

interrogating the ever-growing list of mutations found in human tumors.

Alternatives to GEMMs for modeling cancer

The work in this thesis has focused almost exclusively on the use of advanced GEMMs for dissecting mechanisms of cancer progression. However, it would be remiss not to mention ongoing efforts to develop models that more faithfully recapitulate certain aspects of human cancer biology that GEMMs are unable to replicate. Such differences between mice and humans, which can be difficult to overcome, may confound certain types of studies such as modeling responses to therapy. These include cell-intrinsic factors, such as differences in telomere lengths between mouse and human cells, as well as systemic factors, such as differences in metabolic enzymes that can affect the metabolism of and response to drugs (reviewed in Frese and Tuveson, 2007).

In an effort to overcome some of these problems, patient-derived xenograft (PDX) models have been increasingly used to complement GEMMs. Such models involve the implantation of tumor tissue obtained from patients directly into immunocompromised mice, without an intervening *in vitro* step such as cell culture. Several groups have also successfully derived xenograft models using circulating tumor cells (CTCs) as the starting material (CTC-derived xenografts, or CDXs) (Drapkin et al., 2018; Hodgkinson et al., 2014). In addition to the advantage of using human tumor cells instead of murine cells, another advantage of PDXs over GEMMs is the potential preservation of aspects of the human tumor microenvironment in the engrafted tumor, at least during early passages in host mice. However, one major disadvantage of such models is the need for immunocompromised mice to prevent immune rejection of the xenografted tumor; this means that such models have limited utility for the study of cancer immunology and immunotherapy. Ongoing efforts to generate humanized mice

with components of their immune system replaced with human versions may provide a potential solution to this problem (reviewed in Morton et al., 2016). In addition, PDX models are much less amenable to genetic manipulation compared with GEMMs. Because of the complementary strengths and weaknesses of GEMMs and PDXs, the use of both types of models should be encouraged whenever possible, especially for preclinical testing of cancer therapies.

Functional profiling of cancer

As discussed in the Chapter 1, genetic perturbation screens can be a powerful approach to complement cancer genome sequencing studies for identifying new therapeutic targets. Because genetic screens utilize functional readouts such as changes in cell proliferation, they can reveal candidates that may not have been identified through other approaches such as genomic or transcriptomic profiling. This is demonstrated in Chapter 3, where *Dhodh* was identified as a potential target in SCLC cells, despite not being significantly mutated in human SCLC tumors, nor being differentially expressed in SCLC compared with LUAD or PDAC. Interestingly, a recent study that performed quantitative proteomic analysis of the consequences of *Rb1* loss demonstrated that changes at the RNA and protein levels were not always correlated (Nicolay et al., 2015). Furthermore, the changes observed at the protein level correlated better with the phenotypes observed in *Rb1*-null tissues compared with the changes at the RNA level. This suggests that proteomic profiles of tumor cells may also correlate more closely with results from genetic perturbation screens. The development of cheaper methods for quantitative proteomic profiling could enable this to become a

useful approach to complement genetic screens for functional profiling of cancer cells in the future.

Targeting metabolism in cancer

Although the work described in Chapter 3 revealed several metabolic vulnerabilities that were novel in SCLC, the idea of targeting metabolism in cancer is not novel (reviewed in Luengo et al., 2017). Cancer cells have long been known to be metabolically distinct from normal cells, as first observed by Otto Warburg nearly 100 years ago. Drugs that target metabolism in cancer were first used in 1947, when Sidney Farber pioneered the use of aminopterin, an inhibitor of *de novo* nucleotide synthesis that blocks the activity of dihydrofolate reductase, in the treatment of acute lymphoblastic leukemia in children (Farber et al., 1948).

Many chemotherapies, including those that target various aspects of metabolism, are thought to be effective at targeting cancer cells due to their effects on highly proliferative cells in general, which explains many of the common side effects experienced by patients, such as gastrointestinal toxicities and hair loss. Because most metabolic pathways are essential for cell survival, it is challenging to identify metabolic drugs that specifically target cancer cells and spare normal proliferative cells in the body, although some exceptions do exist, such as therapies that target tumors with mutations in *IDH1* or *IDH2* (Dang et al., 2016).

Using a genetic screening approach, we were able to identify vulnerabilities in several metabolic pathways that appear to be specific to SCLC and not to two other epithelial cancers, LUAD and PDAC. Because both LUAD and PDAC cell lines

proliferate significantly more quickly than SCLC cell lines *in vitro*, this suggests that

these metabolic vulnerabilities are not simply a result of differences in cell proliferation

rates. In the case of the *de novo* pyrimidine biosynthesis pathway, we have observed

that inhibition of this pathway with the Dhodh inhibitor brequinar significantly suppresses

growth of SCLC tumors *in vivo*, but does not result in toxicities in these animals.

Likewise, others have demonstrated that brequinar is effective in specifically targeting

leukemic cells, while sparing normal blood cells, in various murine models of acute

myeloid leukemia (AML) (Sykes et al., 2016). Together, these indicate that most normal

cells are unaffected by inhibition of this pathway, at least at the doses of brequinar used

in these studies, suggesting the existence of a therapeutic window for the treatment of

SCLC and AML using this approach. The use of focused sgRNA libraries that

specifically target metabolic genes could be used in the future to systematically identify

other metabolic vulnerabilities in different cancer types.

Final perspective

The declaration of SCLC as a recalcitrant cancer has ignited renewed interest in

a disease that was relatively neglected for a period of time (Gazdar et al., 2017). This

has led to significant advances in our understanding of the molecular mechanisms that

drive SCLC progression. This thesis presents tools that have facilitated functional

interrogation of SCLC, both in advanced GEMMs of SCLC and in cell lines derived from

these models. With some of the earlier discoveries starting to make their way into

clinical trials, there is reason to be optimistic that the growing body of knowledge on

SCLC will eventually lead to breakthroughs in the treatment of this deadly disease.

References

Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371–385.e18.

Chiou, S., Winters, I.P., Wang, J., Naranjo, S., Dudgeon, C., Tamburini, F.B., Brady, J.J., Yang, D., Grüner, B.M., Chuang, C., et al. (2015). Pancreatic cancer modeling using retrograde viral vector delivery and in vivo CRISPR/Cas9-mediated somatic genome editing. Genes Dev. *29*, 1576–1585.

Dang, L., Yen, K., and Attar, E.C. (2016). IDH mutations in cancer and progress toward development of targeted therapeutics. Ann. Oncol. *27*, 599–608.

Dow, L.E., Fisher, J., O'Rourke, K.P., Muley, A., Kastenhuber, E.R., Livshits, G., Tschaharganeh, D.F., Socci, N.D., and Lowe, S.W. (2015). Inducible in vivo genome editing with CRISPR-Cas9. Nat. Biotechnol. *33*, 390–394.

Drapkin, B.J., George, J., Christensen, C.L., Mino-Kenudson, M., Dries, R., Sundaresan, T., Phat, S., Myers, D.T., Zhong, J., Igo, P., et al. (2018). Genomic and Functional Fidelity of Small Cell Lung Cancer Patient-Derived Xenografts. Cancer Discov. *8*, 600–615.

Farber, S., Diamond, L.K., Mercer, R.D., Sylvester, R.F., and Wolff, J.A. (1948). Temporary Remissions in Acute Leukemia in Children Produced by Folic Acid Antagonist, 4-Aminopteroyl-Glutamic Acid (Aminopterin). N. Engl. J. Med. *238*, 787–793.

Frese, K.K., and Tuveson, D. a (2007). Maximizing mouse cancer models. Nat. Rev. Cancer *7*, 654–658.

Gazdar, A.F., Bunn, P.A., and Minna, J.D. (2017). Small-cell lung cancer: what we know, what we need to know and the path forward. Nat. Rev. Cancer *17*, 725–737.

George, J., Lim, J.S., Jang, S.J., Cun, Y., Ozretić, L., Kong, G., Leenders, F., Lu, X., Fernández-Cuesta, L., Bosco, G., et al. (2015). Comprehensive genomic profiles of small cell lung cancer. Nature *524*, 47–53.

Hodgkinson, C.L., Morrow, C.J., Li, Y., Metcalf, R.L., Rothwell, D.G., Trapani, F., Polanski, R., Burt, D.J., Simpson, K.L., Morris, K., et al. (2014). Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. Nat. Med. *20*, 897–903.

Liao, H.-K., Hatanaka, F., Araoka, T., Reddy, P., Wu, M.-Z., Sui, Y., Yamauchi, T., Sakurai, M., O'Keefe, D.D., Núñez-Delicado, E., et al. (2017). In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans -epigenetic Modulation. Cell 1–13.

Luengo, A., Gui, D.Y., and Vander Heiden, M.G. (2017). Targeting Metabolism for Cancer Therapy. Cell Chem. Biol. *24*, 1161–1180.

Mollaoglu, G., Guthrie, M.R., Böhm, S., Brägelmann, J., Can, I., Ballieu, P.M., Marx, A., George, J., Heinen, C., Chalishazar, M.D., et al. (2017). MYC Drives Progression of Small Cell Lung Cancer to a Variant Neuroendocrine Subtype with Vulnerability to Aurora Kinase Inhibition. Cancer Cell 1–16.

Morton, J.J., Bird, G., Refaeli, Y., and Jimeno, A. (2016). Humanized Mouse Xenograft Models: Narrowing the Tumor–Microenvironment Gap. Cancer Res. *76*, 6153–6158.

Nicolay, B.N., Danielian, P.S., Kottakis, F., Lapek, J.D., Sanidas, I., Miles, W.O., Dehnad, M., Tschöp, K., Gierut, J.J., Manning, A.L., et al. (2015). Proteomic analysis of pRb loss highlights a signature of decreased mitochondrial oxidative phosphorylation. Genes Dev. *29*, 1875–1889.

Romero, R., Sayin, V.I., Davidson, S.M., Bauer, M.R., Singh, S.X., LeBoeuf, S.E., Karakousi, T.R., Ellis, D.C., Bhutkar, A., Sánchez-Rivera, F.J., et al. (2017). Keap1 loss promotes Kras-driven lung cancer and results in dependence on glutaminolysis. Nat. Med. *23*, 1362–1368.

Roper, J., Tammela, T., Cetinbas, N.M., Akkad, A., Roghanian, A., Rickelt, S., Almeqdadi, M., Wu, K., Oberli, M.A., Sánchez-Rivera, F., et al. (2017). In vivo genome editing and organoid transplantation models of colorectal cancer and metastasis. Nat. Biotechnol. *35*, 569–576.

Sánchez-Rivera, F.J., Papagiannakopoulos, T., Romero, R., Tammela, T., Bauer, M.R., Bhutkar, A., Joshi, N.S., Subbaraj, L., Bronson, R.T., Xue, W., et al. (2014). Rapid modelling of cooperating genetic events in cancer through somatic genome editing. Nature *516*, 428–431.

Sykes, D.B., Kfoury, Y.S., Mercier, F.E., Wawer, M.J., Law, J.M., Haynes, M.K., Lewis, T.A., Schajnovitz, A., Jain, E., Lee, D., et al. (2016). Inhibition of Dihydroorotate Dehydrogenase Overcomes Differentiation Blockade in Acute Myeloid Leukemia. Cell *167*, 171–186.e15.

Wangensteen, K.J., Wang, Y.J., Dou, Z., Wang, A.W., Mosleh-Shirazi, E., Horlbeck, M.A., Gilbert, L.A., Weissman, J.S., Berger, S.L., and Kaestner, K.H. (2017). Combinatorial genetics in liver repopulation and carcinogenesis with a novel in vivo CRISPR activation platform. Hepatology . Accepted Author Manuscript. doi:10.1002/hep.2962.

Wood, L.D., Parsons, D.W., Jones, S., Lin, J., Sjöblom, T., Leary, R.J., Shen, D., Boca, S.M., Barber, T., Ptak, J., et al. (2007). The genomic landscapes of human breast and colorectal cancers. Science *318*, 1108–1113.

APPENDIX 1

An optofluidic real-time cell sorter for longitudinal CTC studies in mouse models of cancer

Bashar Hamza[1,2] *, **Sheng Rong Ng**[1,3] *, Sanjay M. Prakadan[1,4,5,10,11] *, Francisco Feijó Delgado[1,6], Christopher R. Chin[1], Emily M. King[1], Lucy F. Yang[1], Shawn M. Davidson[1,3], Kelsey L. DeGouveia[1], Nathan Cermak[1,7], Andrew W. Navia[1,4,5,10,11], Peter S. Winter[1,4,5,10,11], Tuomas Tammela[1], Carman Man-Chung Li[1,3], Thales Papagiannakopoulos[1], Alejandro J. Gupta[1,4,5,10,11], Josephine Shaw Bagnall[1,6], Scott M. Knudsen[1], Matthew G. Vander Heiden[1,3,10], Steven C. Wasserman[6], Tyler Jacks[1,3,9,10] ‡, Alex K. Shalek[1,4,5,10,11,12,13] ‡, and Scott R. Manalis[1,6,8] ‡

[1] David H. Koch Institute for Integrative Cancer Research; [2] Electrical Engineering and Computer Science; [3] Department of Biology; [4] Department of Chemistry; [5] Institute for Medical Engineering and Science; [6] Department of Biological Engineering; [7] Program in Computational and Systems Biology; [8] Department of Mechanical Engineering; [9] Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139
[10] Broad Institute of MIT and Harvard, Cambridge, MA 02142
[11] Ragon Institute of MGH, MIT and Harvard University, Cambridge, MA 02139
[12] Division of Health Sciences and Technology, Harvard Medical School, Massachusetts Institute of Technology, Cambridge, MA 02139
[13] Department of Immunology, Massachusetts General Hospital, Boston, MA 02114
* These authors contributed equally to this work.
‡ Co-corresponding authors.

ABSTRACT

Circulating tumor cells (CTCs) play a fundamental role in cancer progression. However, the combination of limited blood volume in mice and the rarity of CTCs in the bloodstream precludes longitudinal, in-depth studies of these cells using existing liquid biopsy techniques. Here, we present an optofluidic system that continuously collects fluorescently labeled CTCs from a genetically engineered mouse model for several hours per day over multiple days or weeks. The system is based on a microfluidic cell-sorting chip connected serially to an un-anesthetized mouse via an implanted arteriovenous shunt. Pneumatically-controlled microfluidic valves capture CTCs as they flow through the device and CTC-depleted blood is returned to the mouse via the shunt. To demonstrate the utility of our system, we profile CTCs isolated longitudinally from animals over a four-day treatment with the BET inhibitor JQ1 using single-cell RNA-Seq (scRNA-Seq). We show that our approach eliminates potential biases driven by inter-mouse heterogeneity that can occur when CTCs are collected across different mice. The CTC isolation and sorting technology presented here provides a research tool to help reveal details of how CTCs change over time, enabling studies to validate the use of CTCs as biomarkers of drug response and facilitating future studies to understand the role of CTCs in metastasis.

INTRODUCTION

Circulating tumor cells (CTCs) are an intermediate in the hematogenous spread of tumors during metastasis (Pantel and Speicher, 2016). Given their accessibility and potential prognostic and diagnostic value, CTCs have been the focus of significant clinical research efforts monitoring response to therapy and predicting risk of relapse (Miyamoto et al., 2015; Ozkumur et al., 2013; Sarioglu et al., 2015; Ting et al., 2014; Yu et al., 2011). Over the last decade, novel microfluidic liquid biopsy tools have been developed to detect and collect CTCs from blood samples (Miyamoto et al., 2015; Ozkumur et al., 2013; Sarioglu et al., 2015; Ting et al., 2014; Yu et al., 2011). Combined with recently developed single-cell profiling methods, such as single-cell RNA-Seq (scRNA-Seq) (Macosko et al., 2015; Satija et al., 2015; Shalek et al., 2013; Zeisel et al., 2015), in-depth examination of CTCs is now possible. Such studies can provide new insights into the genomic properties of CTCs, as well as their relationship to matched primary and metastatic tumors (Aceto et al., 2014; Alix-Panabieres and Pantel, 2016; Lohr et al., 2014; Miyamoto et al., 2015; Sarioglu et al., 2015; Ting et al., 2014; Vishnoi et al., 2015).

Genetically engineered mouse models (GEMMs) of cancer, which mimic the natural multistage evolution of their human counterparts, facilitate characterization of both acute perturbations (e.g. drug treatment) and long-term phenotypic changes (e.g. tumor evolution) not possible in human subjects. However, despite the utility of GEMMs in cancer research, the combination of the small total murine blood volume (~1.5 mL) and the rarity of CTCs in circulating blood (fewer than 100 cells per mL) (Miyamoto et al., 2015; Rhim et al., 2012) precludes the use of existing liquid biopsy techniques for

longitudinal CTC studies in mice. When repeated blood samples are required at short intervals, a maximum of 1.0% of an animal's total blood volume can be removed every 24 hours (~16.5µL for a 25g mouse) (Parasuraman et al., 2010) – a miniscule volume that does not yield a sufficient sample of CTCs for analysis.

GEMMs have been developed that combine genetic perturbations (manipulation of oncogenes and tumor suppressor genes) with genetically encoded fluorescent makers, enabling the unbiased detection and isolation of CTCs from the bloodstream. *In vivo* flow cytometry (IVFC) techniques have been used to enumerate CTCs in ear capillaries or tail veins longitudinally without sacrificing animals (Georgakoudi et al., 2004; Nedosekin et al., 2014; Zettergren et al., 2012), but these techniques do not permit isolation and downstream molecular characterization of CTCs.

To enable longitudinal, in-depth studies of CTC biology in GEMMs and other murine cancer models, we have developed an optofluidic system capable of detecting and capturing fluorescent CTCs in living mice over several hours, days, or weeks. We combine the use of this system with scRNA-Seq to profile gene expression changes in CTCs in response to drug treatment. We demonstrate that this approach eliminates potential biases driven by inter-mouse heterogeneity that can occur when CTCs are collected across different mice.

RESULTS

Key components of our optofluidic system include a polydimethylsiloxane

(PDMS)-based microfluidic CTC sorter chip, a fluorescence detector, and computer-

controlled pneumatic valves (**Fig. 1A, B**). A cannulated mouse with two permanent

catheters easily accessible on its back allows continuous blood withdrawal from the left

carotid artery and return through the right jugular vein. Blood flows at a rate of 30

µL/min into the CTC sorter chip. Two closely-spaced laser beam lines illuminate the

main flow channel of the chip. As such, each fluorescent cell that passes through the

device emits two pulses of light, which are detected by a photomultiplier tube (**Fig. 1C;**

Supplementary Fig. 1). The second laser line allows the controller to compute the

velocity of the cells, which is essential to ensure reliable CTC capture. After detecting a

fluorescent cell and calculating its speed, the controller instantly operates pneumatic

valves (Unger et al., 2000) to redirect a small blood volume that includes the CTC

toward a collection tube (mean ± s.d. = 127 ± 47 nL/sort event, **Fig. 1D;**

Supplementary Fig. 2). Blood from the collection tube can then be further enriched for

CTCs and run through a secondary single-CTC sorting chip for downstream

characterization using techniques such as scRNA-Seq (**Fig. 1E**).

To ascertain the detection limit of our CTC sorter, we passed a sample of healthy

mouse blood spiked with flow-cytometry calibration beads through the system. The

reference beads comprised 5 fluorescence intensity groups, including one with zero

fluorescence. The system consistently detected the two brightest fluorescence levels

(peaks 4 and 5) and approximately the brightest 30% of level 3 (peak 3, **Fig. 2A-C**).

This sensitivity was sufficient to detect nearly the entire population of tdTomato-
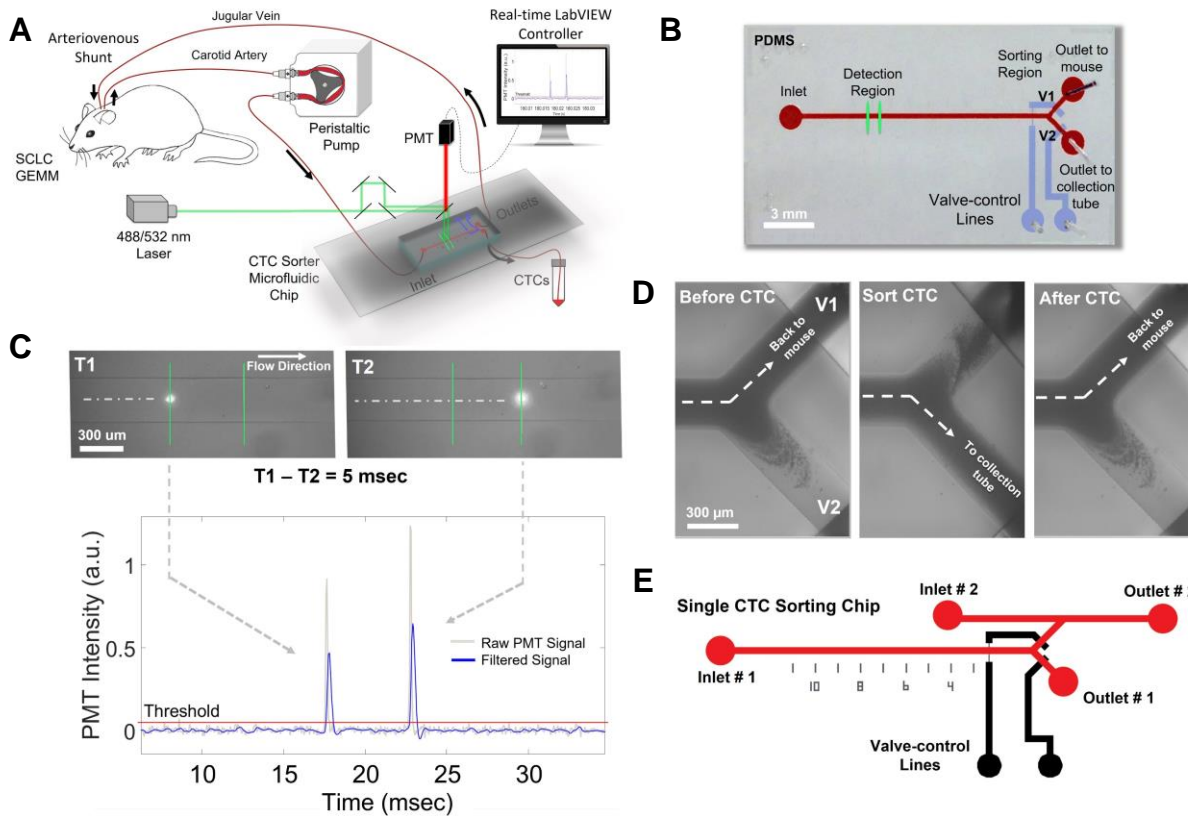
Figure 1: A microfluidic sorter for longitudinal CTC studies in GEMMs.

(**A**) A peristaltic pump withdraws blood from a surgically-implanted cannula in the carotid artery of a mouse at a flow rate of 30 µL/min. The blood is directed into the main flow channel of the CTC sorter chip. For tdTomato-positive cells, a green (532 nm) laser illuminates two points along the main flow channel of the CTC chip separated by a known distance. Thus, fluorescent CTCs emit two red-shifted pulses of light, which are detected by a photomultiplier tube (PMT). Based on the timing of the pulses, a LabVIEW program computes the velocity of the cells and operates computer-controlled pneumatic valves to redirect fluorescent CTCs toward a collection tube. After exiting the chip, CTC-depleted blood returns to the jugular vein of the mouse via a second surgically-implanted cannula.

(**B**) Top-view image of the CTC sorter microfluidic chip showing the inlet, outlets, and the valve actuation lines (V1 and V2, scale bar = 3 mm).

(**C**) Illustration of the CTC detection mechanism using the two excitation laser lines. A low-pass filter is applied to the raw data for determining true peaks.

(**D**) The outlet by which blood is returned to the mouse is briefly sealed while the opposite outlet is opened to allow for CTC isolation in real-time.

(**E**) After collection, CTCs are further enriched by a secondary CTC-sorting chip designed with a parallel channel to flush CTCs into wells containing cell-lysis buffer for downstream scRNA-Seq.

expressing murine small cell lung cancer (SCLC) cells spiked in healthy mouse blood. We also tested blood isolated from autochthonous SCLC tumor-bearing mice that exhibit metastasis to distant organs in a pattern similar to metastatic spread in human patients (McFadden et al., 2014). Tumors in these mice were initiated by Cre-mediated deletion of tumor suppressor genes *Trp53*, *Rb1* and *Pten* in the murine lung epithelium (McFadden et al., 2014). This GEMM also includes a Cre-activated tdTomato allele (Madisen et al., 2010) that fluorescently labels all tumor cells after tumor initiation, including CTCs. The majority of CTCs from the blood of SCLC tumor-bearing mice displayed fluorescence above the detection threshold of our system **(Fig. 2D)**.

After establishing that the sensitivity of the optical detector was sufficient, we characterized and optimized the sorting efficacy using blood samples from healthy mice spiked with low concentrations of tdTomato-expressing murine SCLC cells. In samples containing 100 cells/mL or more, over 80% of detected tdTomato-positive cells were successfully captured. For samples with only 10 cells spiked into 500 µL of healthy mouse blood, the sorted sample contained 6.0 ± 0.7 cells (mean ± s.d., N = 3 repeats). Applying a slight delay in actuating the pneumatic valves until the cell has moved closer to the sorting region decreased the collected blood volume per CTC to 76 ± 28 nL (mean ± s.d.) without compromising capture efficiency. At this volume, approximately 700 neighboring white blood cells (WBCs) and over 700,000 red blood cells (RBCs) and platelets in the bloodstream are collected in addition to the target CTC on each valve actuation. These experiments demonstrate that the CTC sorter is capable of isolating fluorescent CTCs from blood even at very low concentrations **(Supplementary Fig. 3)**.
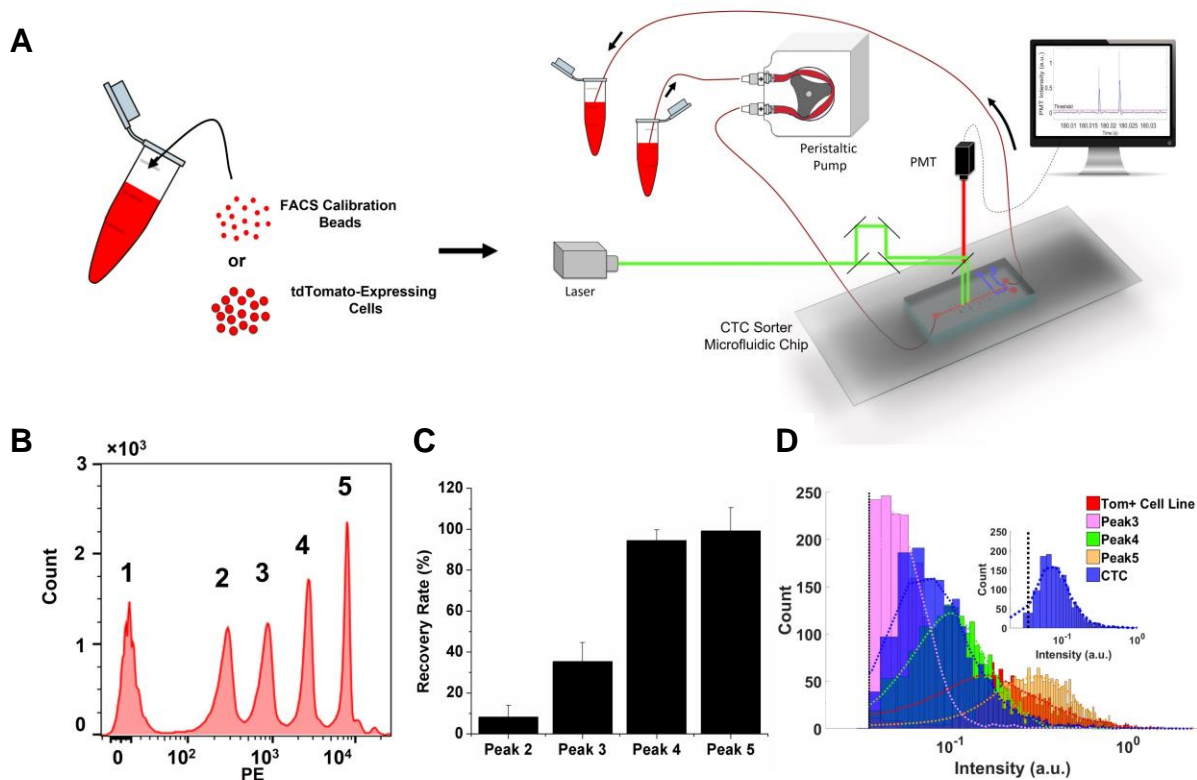
Figure 2: CTC sorter system characterization.

(A) Healthy mouse blood spiked with either fluorescent beads or tdTomato-positive cells is used to determine the detection limits of the system. Each sample is flowed through the sorter at a flow rate of 30 µL/min**.**

(B) Five separate peaks representing the five different intensity levels of the SpheroTech PE Calibration beads as detected by FACS.

(C) Beads representing fluorescent peaks 2 through 5 were spiked separately into four 500 µL samples of healthy mouse blood. Each sample was run through the CTC sorter to measure the recovery rate (i.e. Recovery Rate for Peak 2 sample = total detected/1,000×100). N = 3, results demonstrated as mean ± s.d.

(D) Comprehensive histogram plot with Kernel smoothing function fit (dotted envelope) demonstrating the result of all spiked-blood experiments using beads, a tdTomato-positive fluorescent cell line, or terminal bleed samples from SCLC tumor-bearing mice containing tdTomato-positive CTCs (N = 1,900 events per population. For CTC population, six terminal bleed samples accounted for the 2,000 events). Inset represents the CTC population that lies above the detection threshold (vertical dotted line).

Next, we conducted a longitudinal study of CTCs collected from autochthonous

SCLC tumor-bearing mice treated with the BET bromodomain inhibitor JQ1, which has

been demonstrated to have anti-proliferative effects in SCLC (Jahchan et al., 2016;

Kato et al., 2016; Lenhart et al., 2015). CTCs were isolated from mice over a two-hour

period prior to treatment (0-hour) and at 24-hour intervals following treatment initiation,

continuing over 96 hours (**Fig. 3A, B)**. Isolated samples were enriched for CTCs by

RBC lysis, followed by WBC depletion using magnetic-activated cell sorting (MACS)

(**Fig. 3C**), and finally by passing through a secondary, single-cell CTC sorting chip (**Fig.

3D**). Purified CTCs were then processed using Smart-Seq2 (Picelli et al., 2014). Cells

with insufficient gene complexity for downstream analysis post-scRNA-Seq were

eliminated computationally, in addition to cells with high expression of immune and

platelet signature genes (Miyamoto et al., 2015; Satija et al., 2015; Ting et al., 2014).

The overall yield of the process (from blood to successful scRNA-Seq library) was

11.5% and 5.3% for samples from treated and untreated mice, respectively (median

values with a range of 7.4-31% for treated samples and 3.3-6.7% for untreated

samples, **Supplementary Fig. 4**).

We then examined our data to determine how the information collected

longitudinally from the same mouse with our system compared to the common

approach of capturing CTCs across different mice using asynchronous terminal bleeds

(Aceto et al., 2014; Rhim et al., 2012). To analyze our longitudinal CTC data, we pooled

our collected CTC transcriptomes across all mice, performed a principal component

analysis (PCA), visualized by t-distributed stochastic neighbor embedding (tSNE) (Patel

et al., 2014; Shalek et al., 2013; Tirosh et al., 2016), and identified clusters (using k-
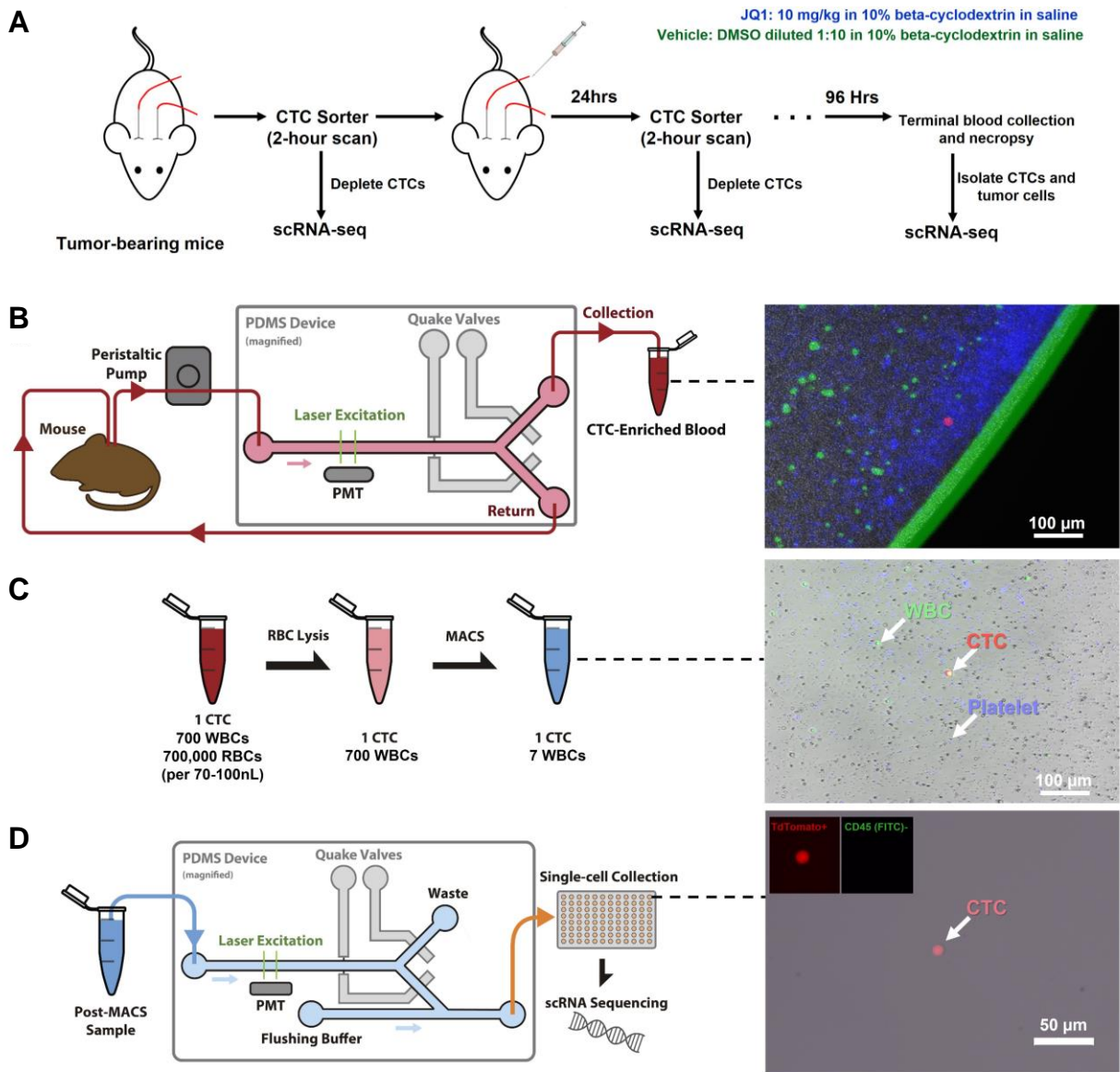
Figure 3: Longitudinal JQ1 treatment and CTC isolation.
(A) Workflow for longitudinal CTC experiments with tumor-bearing mice. Mice are treated with either JQ1 or vehicle daily for four days. Mice undergo two-hour scans to isolate CTC samples, once before treatment and subsequently daily throughout the duration of treatment.
**(B)** Schematic of initial CTC collection from a tumor-bearing mouse. Right panel shows an isolated CTC (red) together with WBCs (green) and platelets (blue).
(C) Isolated CTCs undergo RBC lysis and MACS to enrich for CTCs. Right panel shows the product of this enrichment step.
(D) Enriched samples are run through a secondary single-cell sorting chip to sort single CTCs into wells containing cell-lysis buffer for downstream scRNA-Seq. Right panel shows a purified single CTC within a well of a 96-well plate (insets represent tdTomato-positive (TRITC) and CD45-positive (FITC) channels).

means clustering) over the significant principal components (PCs) (Chung and Storey, 2015; Satija et al., 2015) (**Fig. 4A-C**). This unsupervised analysis revealed that mouse of origin contributed significantly to the variation observed in our dataset, with cluster representation driven primarily by individual mice (Miyamoto et al., 2015; Sarioglu et al., 2015; Ting et al., 2014) (**Fig. 4C**). We next performed PCA on CTCs collected from each mouse individually. Here, we found that PC1 significantly correlated (Spearman correlation) with time since treatment ($p < 0.05$; Student's t-test following a Lilliefors test for normality) when independently calculated for each of the treated mice but not for either control (**Fig. 4D-E**; **Supplementary Fig. 5**). This suggests that by isolating CTCs from the same animal longitudinally, we are able to eliminate potentially confounding differences between animals that could otherwise mask biologically relevant gene expression changes that occur over time.

In comparison, the conventional approach for performing a longitudinal CTC analysis would be to begin the experiment with a cohort of mice and obtain terminal bleeds from a subset of mice at each time point. We simulated this approach from our measurements by selecting a different treated mouse to represent each of the 0, 48, and 96hr time points (**Supplementary Fig. 6**); here, regardless of which mouse was chosen to represent which time point, we found that the mean PC1 coordinate of treated mouse 1 existed outside the interquartile range of the other mice, suggesting a consistent mouse-specific effect that dominates the first PC (**Fig 4F**; **Supplementary Fig. 7**). As such, conclusions drawn from analysis of CTCs from terminal bleeds at different time points across mice would be confounded by organism-specific features from the different mice.
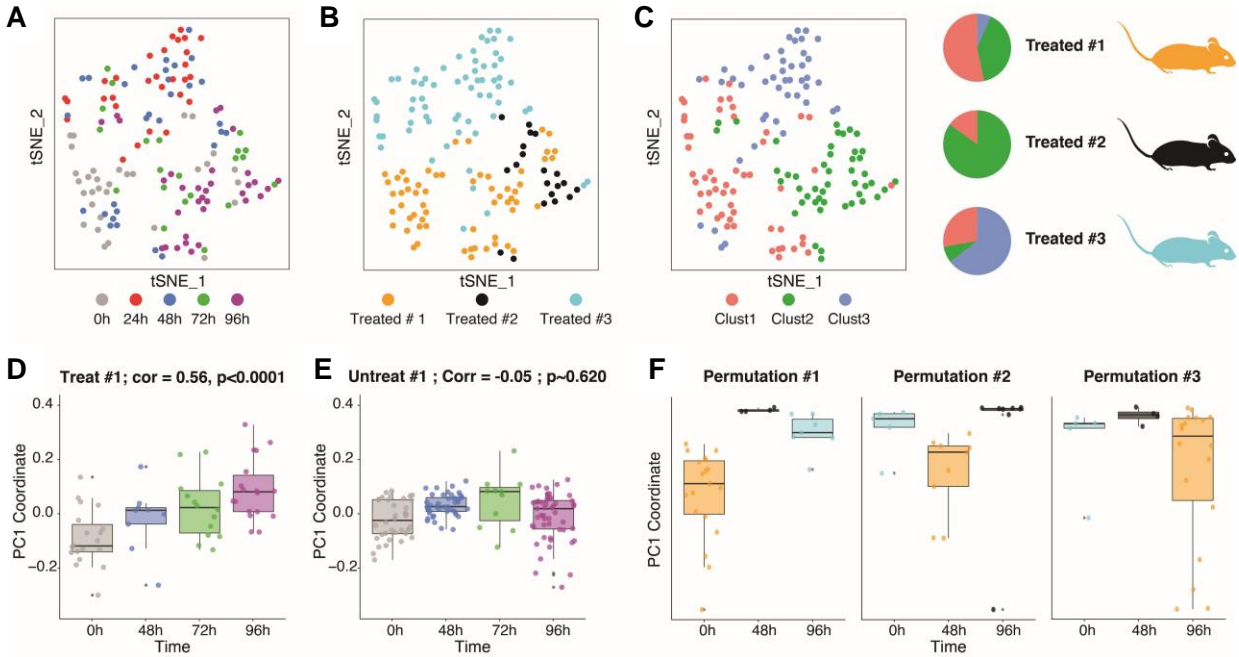
Figure 4: scRNA-Seq of captured CTCs demonstrates utility of intra-mouse CTC profiling.
(A-C) tSNE of all CTCs collected across three JQ1-treated mice colored by time point post-treatment **(A)**, mouse **(B)** and cluster of assignment based on kNN clustering (**C**). Pie charts on the top right show the fractional representation of each cluster in each treated mouse.
(D-E) Boxplots of the first principal component of CTC transcriptomes from PCAs obtained from longitudinally following the same treated mouse (Corr = 0.56) **(D)** or untreated mouse (Corr = -0.05) **(E)**. Each point represents a CTC.
(F) Boxplots of the first principal component from three different "mock terminal bleed" permutations across three treated mice (see **Supplementary Fig. 6**).

To more formally examine treatment-induced shifts in gene expression, we calculated differential expression across all pairs of time points within each longitudinally-profiled mouse (McDavid et al., 2013; Satija et al., 2015) and once again simulated terminal bleed data (**Supplementary Fig. 6**). Our analyses showed that the majority of the differentially expressed genes within each mouse (per-mouse) were unique (**Fig. 5A**). Furthermore, each per-mouse differentially expressed gene set shared little overlap ($p<0.05$; Hypergeometric test) with those calculated from mock terminal bleed datasets (**Fig. 5B**), regardless of which mouse was selected to represent which time point. At each time point, differentially expressed genes in the mock terminal bleed data were enriched for several functional processes, such as mitochondrial function, cellular organization, and metabolism (Huang et al., 2009; Subramanian et al., 2005); however, upon further inspection of the different mock terminal bleed permutations, we found that these enrichments were linked primarily to mouse rather than to time point (**Fig. 4F, 5B; Supplementary Fig. 7;** see, for example, "Housekeeping genes" or "Ribosome"), suggesting confounding mouse-to-mouse heterogeneity. This is evocative of the marked inter- and intra-patient heterogeneity observed in CTCs collected longitudinally from human patients (Miyamoto et al., 2015; Sarioglu et al., 2015; Ting et al., 2014), and suggests the importance of examining the same mouse over time.
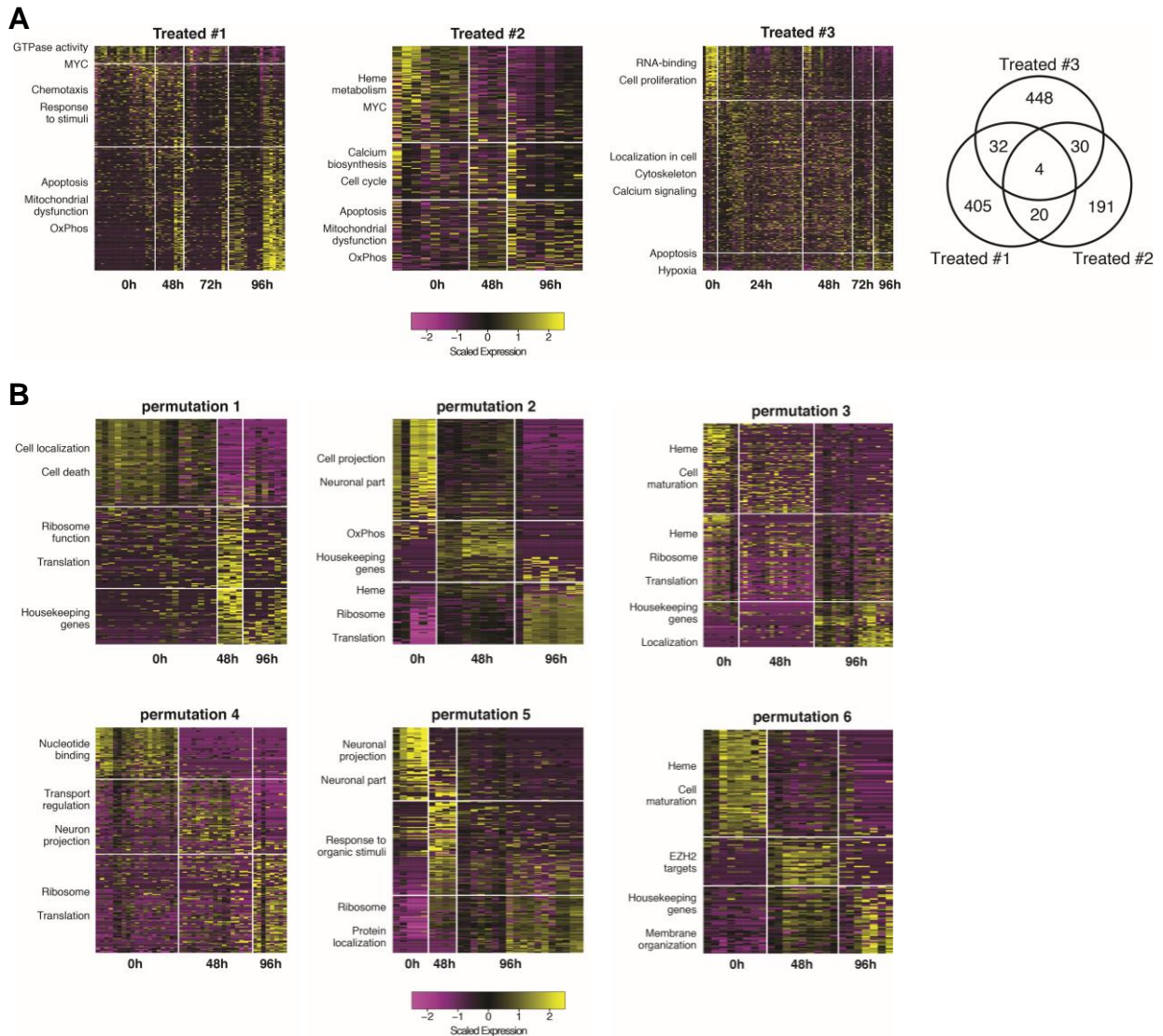
Figure 5: Differential expression analysis of CTCs between different time points.
(A) Left: Heatmaps of differentially expressed genes between all pairs of time points within each treated mouse. Rows represent genes, columns represent cells. Time points sampled are provided on the x-axis and selected gene set enrichments for each differentially expressed gene module (identified by hierarchical clustering) are displayed to the left of the relevant cluster**.**
Right: Venn diagram shows overlap of differentially expressed genes across all treated mice.
(B) Heatmaps of differentially expressed genes between all pairs of time points within each possible permutation for a mock treatment time course using the common 0, 48 and 96h timepoints (See **Supplementary Fig. 6 and 7**).

DISCUSSION

The platform outlined here represents an important advance in the detection and continuous capture of single CTCs from an individual mouse over time. Our method allows CTCs to be isolated in low blood volumes and prepares them for downstream characterization. Here, we used scRNA-Seq to show that continuous CTC capture from the same mouse eliminates biases driven by inter-mouse heterogeneity that can occur when CTCs are collected across different mice. Although future work will be needed to elucidate the underlying drivers of this variability, given the baseline genetic homogeneity of the animals used to generate our GEMM, one potential explanation could be underlying differences in the cellular composition of the primary tumors across different animals. scRNA-Seq results from the primary tumor samples harvested from each animal after terminal CTC collection (96hr) are consistent with this hypothesis (**Supplementary Fig. 8A, B**). These data suggest that primary tumors from each mouse may contain multiple malignant gene expression states (Patel et al., 2014; Tirosh et al., 2016), which appear to be shared across mice. Although some of these differences could be attributed to the presence of multiple primary tumors within each mouse that evolved independently, the fact that each gene expression state is comprised of cells from multiple mice suggests that these states may be a shared feature of this GEMM of SCLC, though future experiments will be needed to robustly validate this finding. Intriguingly, when we computationally assigned terminally collected CTCs (96hr) to these shared states based on gene expression similarity (**Supplementary Fig. 8C**), we observed comparable state frequencies among the CTCs and primary tumor samples of each mouse. Interestingly, similar scoring of CTCs from

earlier time points demonstrated a divergence from these final proportions. This suggests that one reason for the inter-mouse response heterogeneity we observe in our data may be initial differences in the fractional abundances of these tumor gene expression states across mice, as well as dissimilarities in how each gene expression state responds to treatment (**Supplementary Fig. 9**). This highlights the potential importance of having information on primary tumor samples as a reference, though further experiments will be needed to corroborate these preliminary findings.

Future analyses of this kind may shed new light on the relationship between CTCs, primary tumors, and metastases, enabling the exploration of their utility as biomarkers and facilitating examination of how individual CTCs contribute to metastasis. Moreover, they may help elucidate the features that inform shifts observed upon perturbation, such as drug treatment. Ultimately, these data show that our platform opens the door for novel CTC experimentation, such as examining longitudinal drug responses and comparing CTCs to primary tumors (shown here), characterizing their relationship to metastases, and measuring the rate of CTC production in an acute window. With additional development, our device can enable longitudinal studies in mice to find associations between individual CTCs and clusters of CTCs, profile rare immune cells (for example, using genetic reporters or based on tetramer-staining), monitor mesenchymal cells in a variety of contexts (including wound healing and tumor formation), and measure induction rates of drugs or nanoparticles in circulating mononuclear cells.

MATERIALS AND METHODS

Mouse model and cell lines

The *Trp53^{fl/fl}; Rb1^{fl/fl}; Pten^{fl/fl}; Rosa26^{LSL-Luciferase/LSL-Luciferase}* (PRP-L/L) mouse model of SCLC has been described previously (McFadden et al., 2014). *Rosa26^{LSL-tdTomato/LSL-tdTomato}* mice were obtained from Jackson Laboratories (*Gt(ROSA)26Sor^{tm14(CAG-tdTomato)Hze}*) and crossed into the PRP-L/L model to obtain *Trp53^{fl/fl}; Rb1^{fl/fl}; Pten^{fl/fl}; Rosa26^{LSL-tdTomato/LSL-Luciferase}* mice. Tumors were initiated by intratracheal delivery of 2×10^8 plaque-forming units (p.f.u.) of adenovirus expressing Cre recombinase under the control of a CGRP promoter (Ad5-CGRP-Cre) (Sutherland et al., 2011), as previously described (DuPage et al., 2009). Adenoviral stocks were purchased from the Viral Vector Core Facility at the University of Iowa Carver College of Medicine. Candidates for arteriovenous shunt surgery were identified by *in vivo* bioluminescence imaging using the IVIS Spectrum In Vivo Imaging System (PerkinElmer).

Murine SCLC cell lines (AF1281-M1, AF3291LN) were generated from mSCLC tumors isolated from *Trp53^{fl/fl}; Rb1^{fl/fl}; Rosa26^{LSL-tdTomato/+}* or *Trp53^{fl/fl}; Rb1^{fl/fl}; Pten^{fl/fl}; Rosa26^{LSL-tdTomato/LSL-Luciferase}* mice as previously described (Dooley et al., 2011).

Shunt surgery

All animal-based procedures were approved by the Massachusetts Institute of Technology Committee on Animal Care (CAC), Division of Comparative Medicine (DCM). The arteriovenous shunt method was validated as an approach for real-time manipulation of nutrient levels in the serum of anesthetized mice (Ayala et al., 2006, 2010, 2011). We adopted the same technique for continuous sampling of mouse blood

for CTCs. Briefly, catheters are inserted into the right jugular vein and the left carotid artery and are externalized using standard cannulation surgical techniques. A peristaltic pump (Instech Laboratories Inc., Plymoth Meeting, PA, USA) is then connected to the catheters for blood sampling and return through the carotid artery and jugular vein, respectively, in the conscious mouse (**Fig. 1A**). During the four-day longitudinal studies, the total collected blood volume is monitored. If depleted blood volume exceeds 1% of the animal's body weight (for example 260 µL for a 25g mouse), per MIT DCM guidelines, healthy-mouse donor blood (of same strain and sex) is infused directly into the mouse using the jugular vein catheter, equivalent to the amount removed.

Optical detection platform

The optical system is comprised of two optical trains, making up two compact vertical microscopes. The top optical train divides the laser beam (OBIS 532 LS, Coherent Inc) into two separate beams that are focused along one axis to produce two illumination lines projected at the sample focal plane (perpendicular to the blood flow channel) for precise velocity measurements of the flowing CTCs (**Supplementary Fig. 1A, B**). The laser passes through a line filter and polarizing beam splitters to generate the two beams with minimal losses. Next, the two laser lines pass through a cylindrical lens to focus the two beams into lines. The focused lines are then projected onto the microfluidic channel with a 4F optical system. The dichroic mirror and longpass filters, placed directly above the detection region, pass a filtered fluorescence signal to the PMT (Hamamatsu H10722-20) by blocking the 532 nm laser line signal with a notch filter (532 nm StopLine single-notch filter, Semrock). A 90:10 beam splitter is added

before the PMT to allow imaging of the illumination region for device alignment purposes.

The bottom optical train is similar in configuration to the top train and uses a second green laser to illuminate a circular region covering the valves, enabling fluorescent and bright field imaging for confirming successful CTC sorting and valve functionality throughout the experiment.

Microfluidic device design and fabrication

The microfluidic device for sorting minute amounts of blood containing single CTCs was designed to have one inlet to a 1000μm-long microfluidic channel that bifurcates into two channel outlets (90º apart, **Fig. 1B, C**); one for returning the blood to the mouse and the other for collecting the sorted CTC-containing blood sample. The fabrication was performed using standard soft lithographic techniques on two four-inch wafers. A single layer of photoresist (SU8 2050, Microchem, Newton, MA) was patterned to create the pneumatic channels on the valve control wafer. For the blood flow channel, AZ9260 positive resist (MicroChemicals) was exposed, developed, and then reflowed at 120˚C for 10 minutes to create the half ellipsoid channel profile necessary for a complete valve seal (**Supplementary Fig. 2**) (Unger et al., 2000).

A mixture of PDMS (Polydimethylsiloxane) and its curing agent (SYLGARD 184 A/B, Dowcorning, Midland, MI, USA) at a 10:1 ratio was spun on top of the actuation wafer to a thickness of 50 μm and baked in an oven set to 65 ˚C for at least 3 hours. For the flow channel layer, the mixture was poured to a thickness of ~5 mm and cured at 65 ˚C for 3 hours. Afterwards, the flow channel layer was peeled off and punched with a

0.75 mm puncher (Harris Uni-Core, Ted Pella Inc., Reading, CA) to define the inlet and outlets to and from the channel, respectively, and diced to prepare for bonding. The flow channel devices and the actuation layers were then treated with oxygen plasma (100 watt, 1 ccm, 140 torr, 10 sec). Next, the flow layer was aligned to the actuation layer and transferred to a hot plate set to 85 ˚C. After 15 minutes, the assembled PDMS layers were peeled off and punched with a 0.75 mm puncher to define inlets to the actuation channels. The assembled PDMS layers were treated with oxygen plasma (100 watt, 1 ccm, 140 torr, 10 sec) for irreversible bonding to a glass slide (Fisherbrand 1x3", Fisher Scientific, Pittsburgh, PA).

Prior to flow experiments, the device was aligned to project the two laser lines across the flow channel approximately 8 mm away from the valve actuation region. The device was then primed with Heparinized-Saline (diluted to 100 USP units per mL, NDC 25021-400-30) to prevent any clotting within the microfluidic channel.

Real-time data processing and system control

The PMT module generates an output voltage that is sampled by a NI USB-6009 (National Instruments) 14 bit analog-to-digital converter. This analog-to-digital converter is also used to output the control voltage signals for the peristaltic pump and the PMT gain. The PMT voltage is acquired and displayed in real time in LabVIEW (National Instruments) at 20,000 samples per second. The raw data is filtered with a low pass filter to remove spurious noise and the output of a running median filter is subtracted to remove low frequency drift or any DC offset. Upon transit of a fluorescent cell or particle, the PMT output signal passes a specific threshold, triggering the NI USB-9472

device to actuate the SMC Solenoid Valves (S070A-6DC, 8 ports) that are connected to an external 6-volt power supply.

During non-sorting operation of the system (i.e. cell counting experiments), the first valve (V1) region is kept sealed to force the blood to flow directly through to the opposite outlet and back into the jugular vein catheter of the mouse (Top-view image, **Fig. 1B**). When the sorting functionality is activated, and upon detection of a cell, its velocity is calculated (**Supplementary Fig. 3A**) and used to estimate the time of its arrival at the valve region of the channel, all using a virtual instrument (VI) program in LabVIEW (**Supplementary Fig. 3B**). V2 then closes immediately and V1 opens for an amount of time equal to twice the estimated time of arrival of the cell at the valve to deflect a 127 ± 47 nL (or 76 ± 28 nL with a valve actuation delay) blood bolus containing the cell (**Supplementary Fig. 3C**). This amount of blood per single sort provides an upper limit for the number of events we can collect from a mouse over the four-day study. For a 25g mouse, a maximum number of ~2,600 total events can be collected without the need for fluid replacement, per MIT DCM guidelines.

After CTC enrichment with RBC lysis and MACS (described below), the cells were sorted using our secondary single-CTC sorting chip. Upon detection of a single fluorescent CTC in the secondary single-cell sorting chip (**Fig. 1E**), micro-valves actuate to push the cell into the parallel channel. The micro-valves then seal the primary channel and sample flow is stopped to introduce fresh buffer into the parallel channel, releasing the cell out of the chip into the tubing and then into a collection well containing TCL buffer (QIAGEN) with 1% 2-mercaptoethanol for downstream scRNA-seq (**Fig. 3D**).

CTC enrichment by RBC lysis and MACS

Following CTC isolation, the collected blood sample is first treated with ACK Lysing

Buffer (Gibco A10492-01) to remove RBCs, then rinsed, filtered using a 30µm Pre-

Separation Filter (Miltenyi Biotec, #130-041-407), and processed through magnetic

activated cell sorting (MACS) using mouse CD45 MicroBeads (Miltenyi #130-052-301)

and MS columns (Miltenyi #130-042-201) to remove CD45-positive cells, according to

the manufacturers' protocols. The final product is then diluted to a total volume of 2 mL

of MACS buffer and flowed through the secondary single-CTC sorting chip.

Characterization with cell line and beads

To validate the sensitivity of the system to detect fluorescent cells in blood, 5-peak

FACS calibration beads (Sphero PE 5-peak, Spherotech Inc) were spiked into mouse

blood and flowed through the device at 30 µL/min (**Fig. 2**). To validate the sorting

functionality of the system, a tumor cell line (AF1281-M1) was established from a

mouse with autochthonous small cell lung cancer expressing a bright and stable

cytoplasmic tdTomato fluorescent protein. Cells were then counted using Coulter

Counter (Multisizer 4, Beckman-Coulter) and re-suspended in 500 µL of mouse blood at

different densities (10, 100, 500, and 1000 cells, **Supplementary Fig. 3**).

JQ1 treatment of tumor-bearing mice

Tumor-bearing SCLC mice were treated with 10 mg/kg JQ1 (Cayman Chemical) by

intravenous injection daily for the duration of the study. JQ1 was dissolved in DMSO to

make a 20 mg/mL stock, then diluted 1:10 with 10% beta-cyclodextrin in 0.9% saline to

obtain a working concentration of 2 mg/mL. Vehicle-treated control mice received an equivalent dose of DMSO diluted 1:10 with 10% beta-cyclodextrin in 0.9% saline (**Fig. 3A**).

Dissociation of tumor samples for single-cell RNA-sequencing analysis

Primary tumors from tumor-bearing animals were dissected, dissociated into single cells using a lung dissociation kit according to the manufacturer's protocol (Miltenyi Biotec #130-095-927), then stained with APC-conjugated antibodies against CD11b (eBioscience #17-0112-82), CD31 (BioLegend #102510), CD45 (eBioscience #17-0451-83) and TER-119 (BD Biosciences #557909). tdTomato-positive, APC-negative cells were single-cell sorted by FACS into TCL buffer (QIAGEN) containing 1% 2-mercaptoethanol, then frozen at -80ºC for downstream processing for scRNA-Seq.

Single-cell RNA-sequencing sample preparation

Both CTC samples and primary tumor samples collected in TCL buffer were processed through Smart-Seq2 as follows. Total nucleotide material from lysed single cells was extracted with RNA-clean AMPure nucleotide extraction beads (Beckman-Coulter) and washed with 80% ethanol before undergoing reverse transcription with Maxima enzyme (Thermo Scientific), followed by PCR with a KAPA Hotstart Readymix 2x kit (KAPA biosystems). Following quantification and quality control analysis by Qubit DNA quantification (Thermo Scientific) and tape station (Agilent), whole transcriptome amplifications (WTAs) of each single cell were transformed into sequencing libraries with a Nextera XT kit (Illumina) and barcoded with unique 8-bp DNA barcodes. cDNA

libraries were pooled, quantified, and sequenced on an Illumina NextSeq 500 to an average depth of 1.2M reads/CTC.

Analysis of raw sequencing data

Following sequencing, BCL files were converted to merged, demultiplexed FASTQs. Paired-end reads were mapped to mm10 mouse transcriptome (UCSC) with Bowtie. Expression levels of genes were log-transformed transcript-per-million ($TPM[i,j]$) for gene i in sample j, estimated by RSEM in paired-end mode. For each cell, we enumerated genes for which at least one read was mapped, and the average expression level of a curated list of housekeeping genes. We excluded from analysis profiles with fewer than 500 detected genes or an average housekeeping expression below 0.5 $\log_2$(TPM).

Identification of leukocytes and correction of platelet effect

A matrix of TPM estimates for all genes across all cells resulted from preprocessing of RNA-Seq data. A raw principal components analysis was run for each sample, and the first PC separated cells of immune lineage from cells expressing epithelial genes. We selected from overlaps between this first PC and a curated set of immune-related genes to score CTCs according to their immune score. Cells with an average immune score below 0.5 $\log_2$(TPM) were taken forward for analysis of their transcriptomes as CTCs.

Likewise, a curated set of platelet-related genes was used to develop a platelet signature for each CTC transcriptome. The signal of this platelet signature was then regressed out using the RegressOut function in Seurat.

Principal component analysis of CTCs

We performed a gene selection based on a binning strategy across expression using Seurat to define a unique set of variable genes for each treated and untreated mouse, as well as combined sets of treated and untreated mice. Principal component analysis was performed over these variable gene sets for each case (per-mouse or combined). PCs were correlated to time using Spearman correlation. To test for significance of PC1 correlation with time, we permuted assignments of time (N = 1000) and calculated statistics for each set of CTCs per mouse. The sets of correlations for each mouse were tested with Lilliefors test for normality and used to determine PCs with significant correlations with time.

Generation of mock data

To simulate current methods of terminal bleed assays for CTC collection across an acute time scale, we generated a series of mock datasets using our true, continuously collected data. At each time point, the CTCs from a single treated mouse were selected without replacement. This process was exhaustedly repeated five times to generate all unique treated mock data over which further differential expression analysis was performed.

Differential expression analysis

For each of the treated mice, as well as the mock data, we performed differential expression using Seurat's built-in single-cell differential expression tool, with a bimodal distribution model. Differential expression was performed between all pairs of time

points available for each dataset. Genes with avg_diff > 1 and p-value < 0.01 were selected and visualized using the DoHeatmap function in Seurat. Furthermore, these genes were enriched for upstream regulators using gene set enrichment analysis (GSEA) through the Broad Institute.

Analysis of primary tumor data

Variable genes across primary tumor cells of all mice were calculated and principal component analysis was performed as described above. Relevant PCs were determined by visualizing percent variance explained in an elbow plot, significance by Jackstraw (Zheng et al., 2017), and manual inspection of loadings and coordinates. Following inspection, PCs 1-5 were selected for downstream visualization of the primary tumor cells by tSNE in Seurat, with perplexity set to 15, and 2500 iterations run.

Assignment of CTCs to primary tumor cluster

Differential expression, as reported above, was used to identify marker genes that describe the resultant clusters in the primary tumor. These marker genes were used to develop a signature score for each cluster for each CTC, using weighted averaging of the genes for each signature. Next, we permuted random sets of genes with similarly binned expression distribution and size to create cluster-independent background scores for each CTC. CTCs with cluster-specific signatures above their cluster-independent background were "assigned" to that cluster. CTCs were assigned to all clusters which were above background – if a CTC scored above background for more than one cluster, it was assigned to all those clusters; if a CTC scored above

background for no clusters, it was classified as "unassigned" (**Supplementary Fig. 8**).

The total number of CTCs per mouse, per time point for each cluster was visualized by

a stacked bar plot, generated through ggplot2 in R.

ACKNOWLEDGEMENTS

SUPPLEMENTARY FIGURES

Supplementary Figure 1: Optical configuration of the CTC sorter.
(A) Side-view image showing the optical components of the enclosed CTC sorter system. Inset indicates the location of the microfluidic device that is sandwiched between two vertical optical trains.
(B) Top optical train splits the excitation source and focuses it into two lines across the microfluidic channel near the inlet. Emitted light is then filtered and passed to a PMT for real-time CTC detection. Bottom optical train combines a diffused excitation source at the bifurcation region for continuous imaging of the valve region (Key – Sph: Spherical lens, Cyl: Cylindrical lens, PBS: Polarizing Beamsplitter, R: Reflection, T: Transmission, NF: Notch filter, and NA: Numerical Aperture).
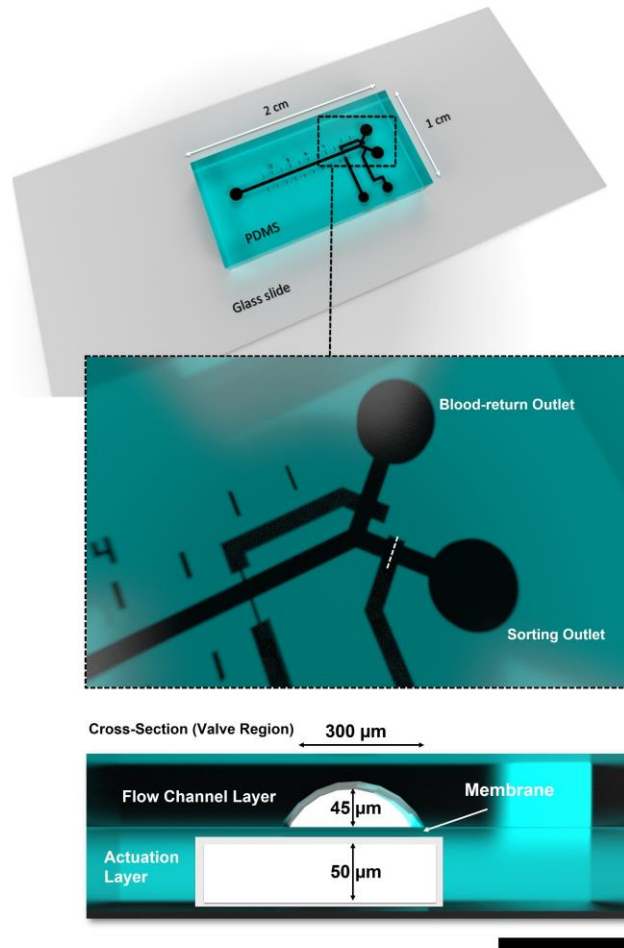
Supplementary Figure 2: Three-dimensional rendering of the microfluidic CTC sorting chip.
The device is fabricated using standard photo- and soft-lithographic techniques to create a two-layer PDMS device bonded to a 1×3-inch glass slide. The microfluidic design contains a single inlet and two outlets. Actuation air channels (50 µm in height) lying underneath the outlet regions of the blood flow channels (half ellipsoid cross section, height = 45 µm, width = 300 µm) deflect a thin PDMS membrane upward or downward to stop or resume blood flow above, respectively.
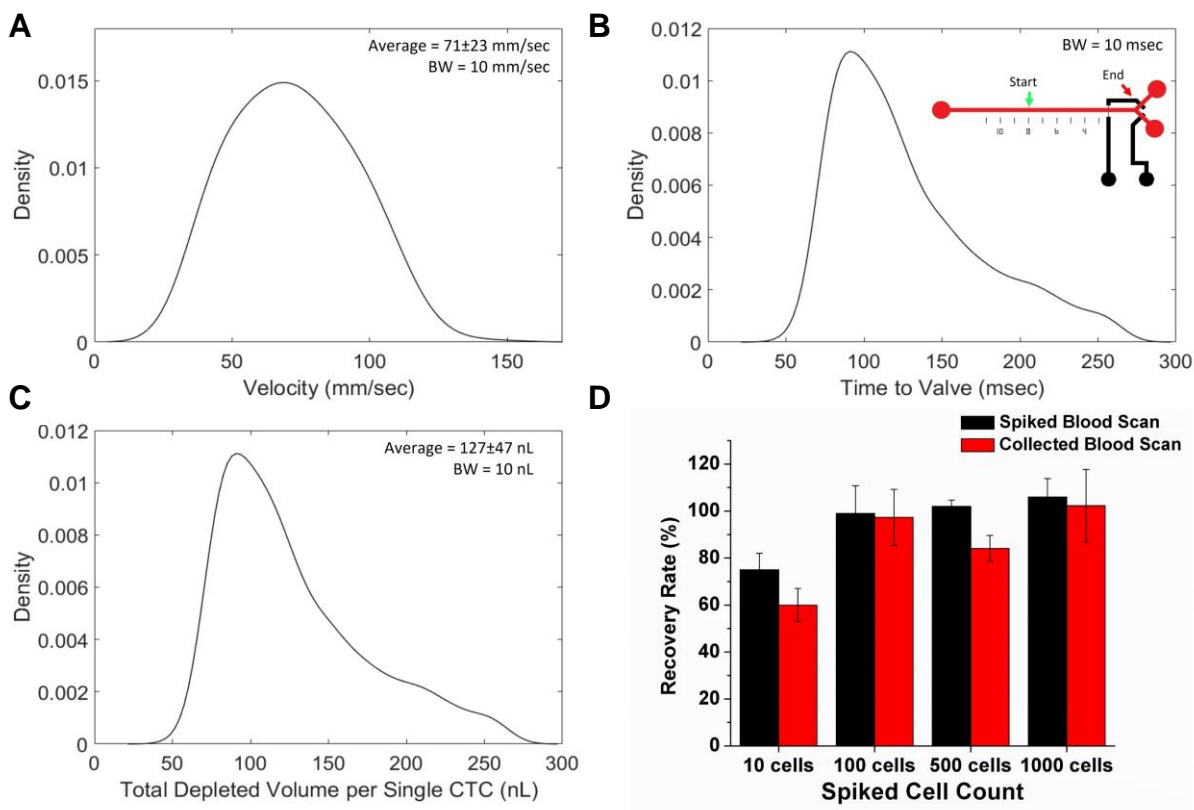
Supplementary Figure 3: Characterization of sorting reliability.
(A) Velocity distribution of cells passing through the dual-excitation detection region of the channel.
(B) Estimated time of arrival at the outlet distribution. Inset demonstrates the distance between the detection region (Start) and the valve region (End) which equals 8 mm.
**(C)** Distribution of the total depleted blood volume per single sort (N = 2,531 cells, Kernel smoothing function with bandwidth (BW) = 10 mm/sec, msec, or nL for parts a through c, respectively).
(D) Healthy mouse blood samples spiked with tdTomato-expressing cells at physiologically-relevant densities (10, 100, 500, and 1,000 counted using a Coulter Counter) were run through the sorter to first identify the actual number of cells in each sample (black columns) with the sorting functionality activated. Sorted samples were then diluted and flowed through the sorter again (red columns) to measure the recovery yield of the sorted samples (N = 3 repeats, mean ± s.d.).
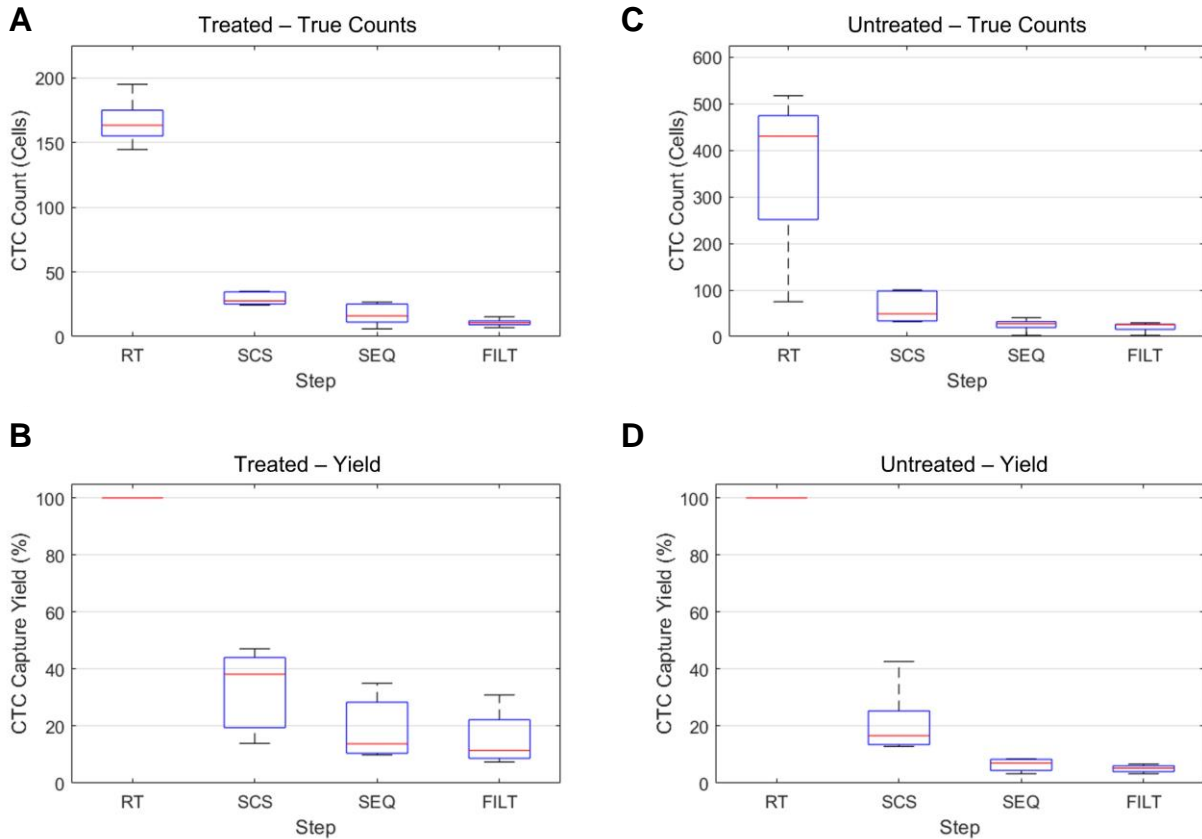
Supplementary Figure 4: Analysis of overall CTC yield.
Boxplots of CTC counts (**A, C**) and overall yields (**B, D**, normalized to RT count) of CTCs undergoing the enrichment process of: 1. real-time CTC collection during the two-hour mouse-blood scan (RT); 2. single-cell sorting (SCS) post-leukocyte depletion using the secondary single-cell sorting chip; 3. technical filtering of cells with sufficient gene complexity post-RNA sequencing (SEQ); and 4. final biological selection (filtering = FILT) of single CTCs that lack immune and platelet signature genes, for both the JQ1-treated samples **(A, B)** and untreated samples **(C, D)**, respectively (the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively).
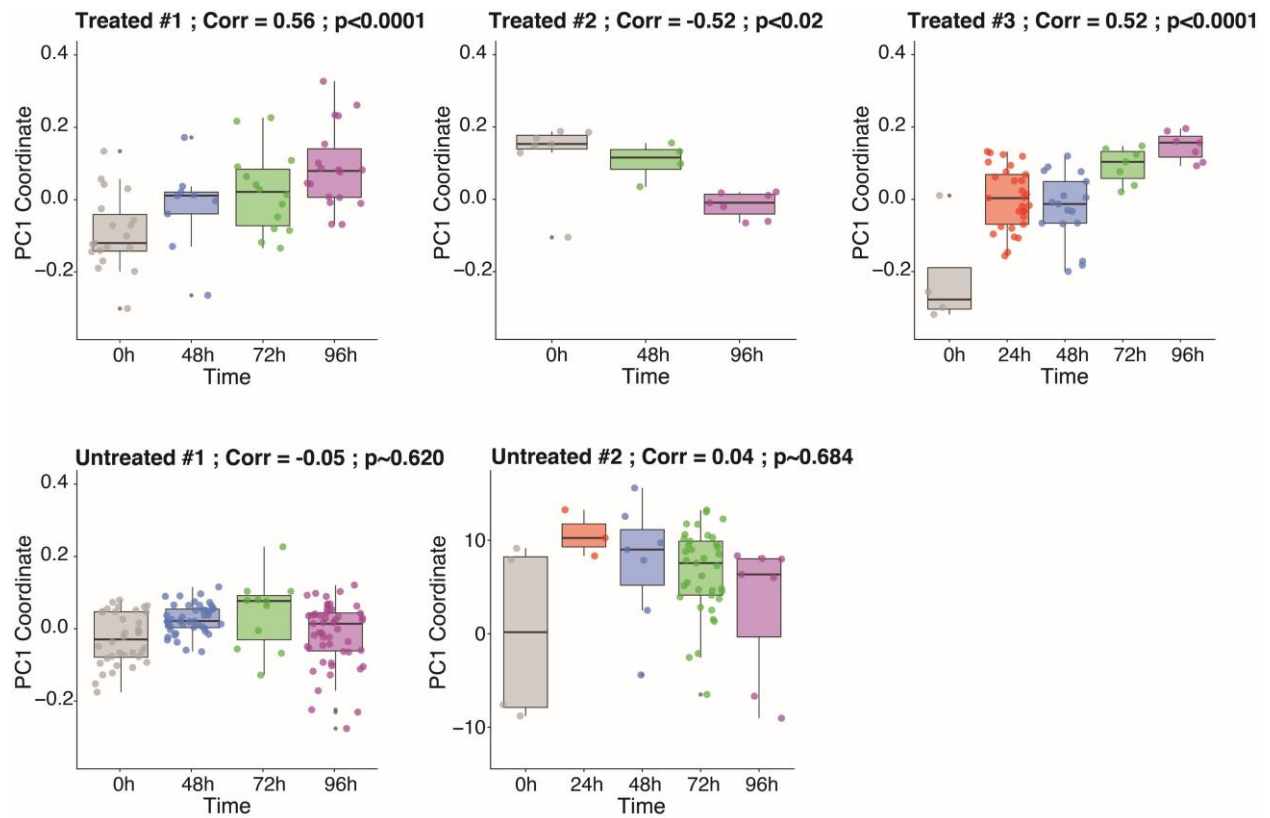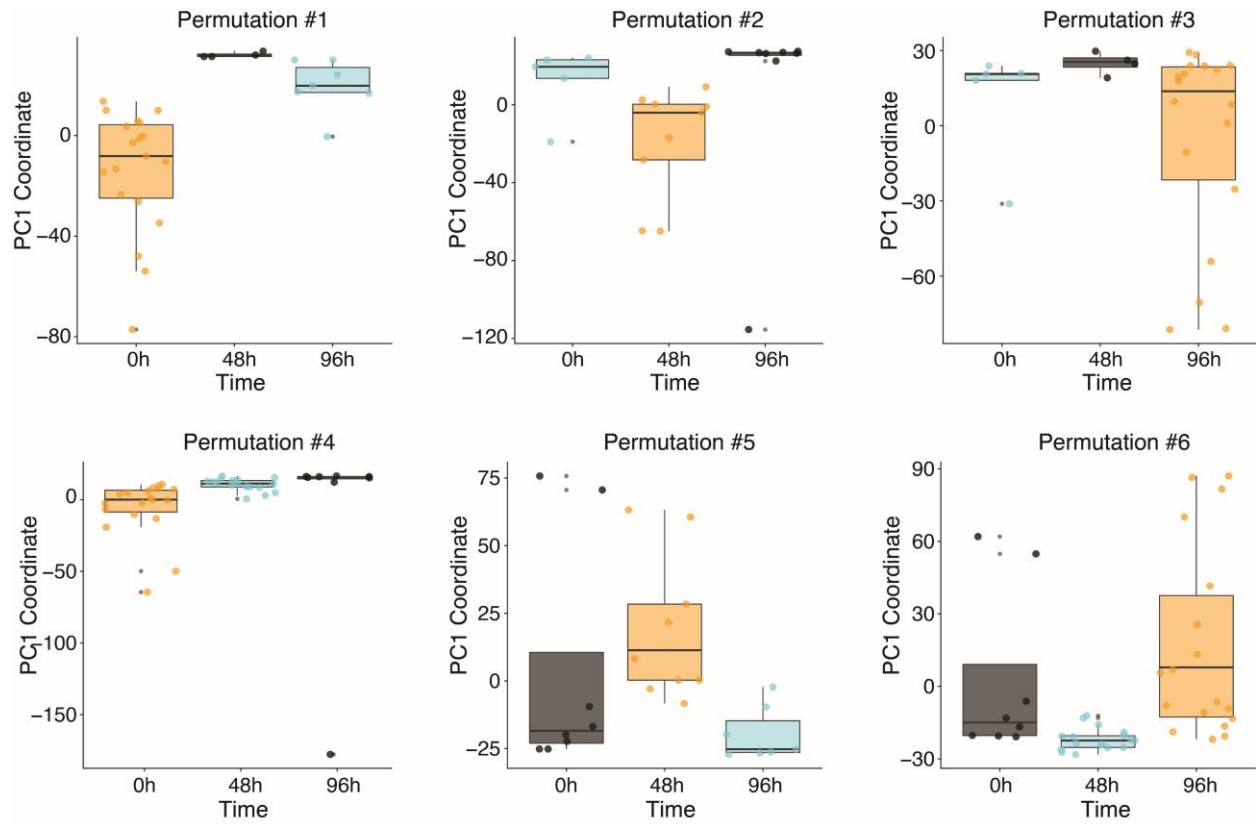
Supplementary Figure 5. Boxplots of the PC1 coordinate for all CTCs from treated mice (top row) and untreated mice (bottom row) over variable genes with correlations and p-values of correlations as determined by Student's t-test of significance, with all correlations of principal components following a Normality test (Lillefors). Data are delineated by time point of collection for each mouse presented. Treated #1 and Untreated #1 are also presented in **Fig. 4**; Treated #'s 2, 3 and Untreated #2 are replicate experiments.
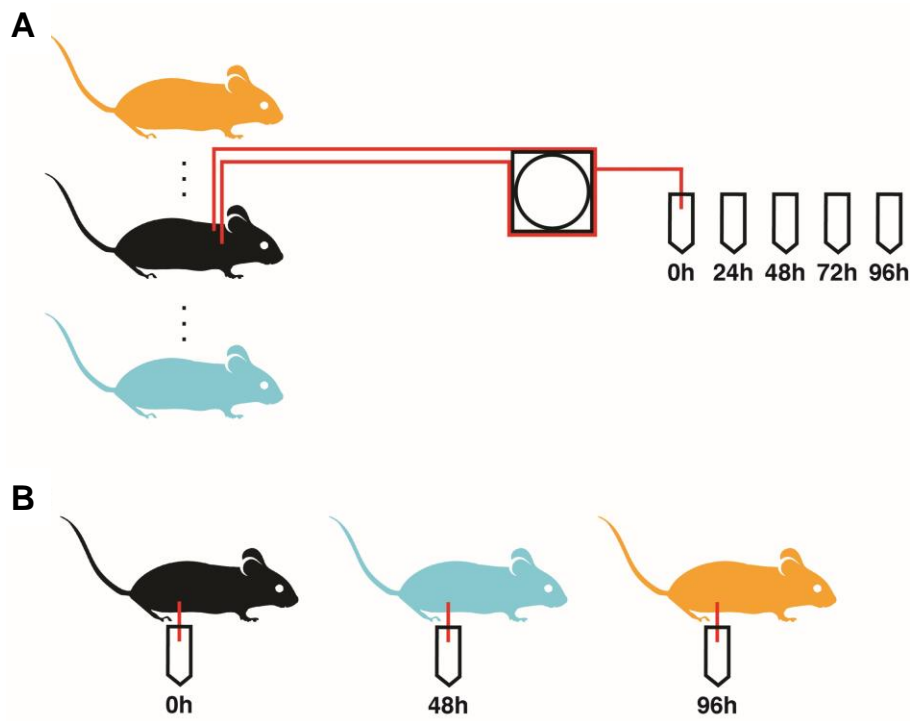
Supplementary Figure 6: Schematic of true and mock data for simulations of experiments.
(A) Collection of single CTCs from individual mice using our CTC sorter.
(B) Simulation of a terminal bleed CTC experiment showing one of the combinations of mice that could lead to a time course collection of CTCs. All six possible permutations were analyzed.

Supplementary Figure 7. PC1 boxplots of all combinations of treated mouse simulated data (see **Supplementary Fig. 6B**). The mean PC1 coordinate of the time point from which the CTCs of Treated #1 were selected (yellow box) was outside the interquartile range of the other two treated mice in all possible combinations. Permutations 1-3 are also shown in **Fig. 4**.
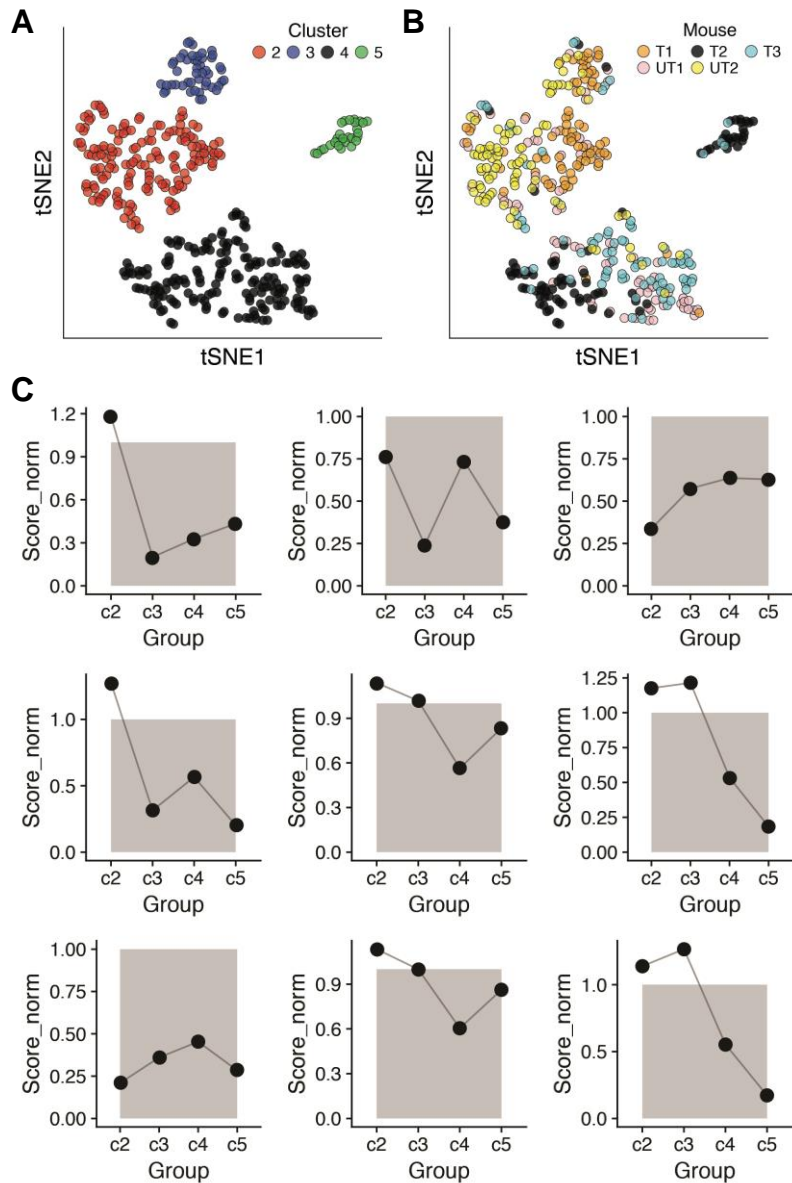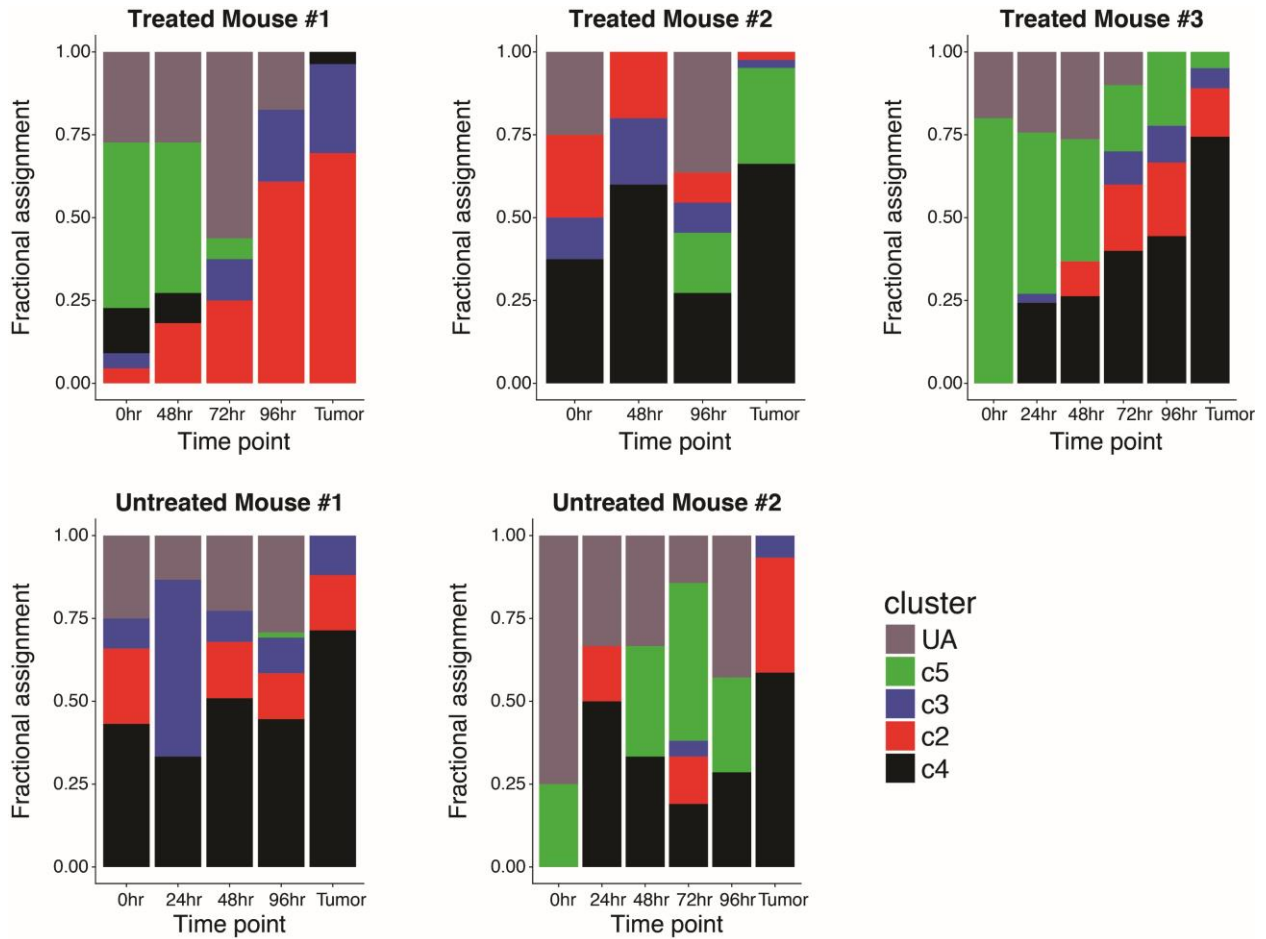
Supplementary Figure 8: scRNA-Seq of primary tumor samples and CTC assignment.
(A) tSNE of primary tumor cells across treated and untreated mice, colored by clusters called from kNN clustering.
(B) tSNE of primary tumor cells across treated and untreated mice, colored by mouse.
(C) Schematic representation of scoring algorithm used to assign CTCs to specific clusters of phenotypes from primary tumor data. CTC signatures were calculated for each cluster of primary tumor cells in **(A)** and compared to a background generated by permuting random sets of genes with similar mean bins of expression (gray background). CTCs were scored to clusters above background (e.g. top left) or not scored if no clusters were above background (e.g. top middle).

Supplementary Figure 9. Assignment plots of CTCs to primary tumor clusters for each mouse from each time point. CTCs were assigned (or not assigned) to clusters based on the highest signature score above permuted background and the results were visualized using stacked bar plots of assignment. UA: unassigned.

REFERENCES

Aceto, N., Bardia, A., Miyamoto, D.T., Donaldson, M.C., Wittner, B.S., Spencer, J.A., Yu, M., Pely, A., Engstrom, A., Zhu, H., et al. (2014). Circulating Tumor Cell Clusters Are Oligoclonal Precursors of Breast Cancer Metastasis. Cell *158*, 1110–1122.

Alix-Panabieres, C., and Pantel, K. (2016). Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy. Cancer Discov. *6*, 479–491.

Ayala, J.E., Bracy, D.P., McGuinness, O.P., and Wasserman, D.H. (2006). Considerations in the Design of Hyperinsulinemic-Euglycemic Clamps in the Conscious Mouse. Diabetes *55*, 390–397.

Ayala, J.E., Samuel, V.T., Morton, G.J., Obici, S., Croniger, C.M., Shulman, G.I., Wasserman, D.H., and McGuinness, O.P. (2010). Standard operating procedures for describing and performing metabolic tests of glucose homeostasis in mice. Dis. Model. Mech. *3*, 525–534.

Ayala, J.E., Bracy, D.P., Malabanan, C., James, F.D., Ansari, T., Fueger, P.T., McGuinness, O.P., and Wasserman, D.H. (2011). Hyperinsulinemic-euglycemic Clamps in Conscious, Unrestrained Mice. J. Vis. Exp. 1–8.

Chung, N.C., and Storey, J.D. (2015). Statistical significance of variables driving systematic variation in high-dimensional data. Bioinformatics *31*, 545–554.

Dooley, A.L., Winslow, M.M., Chiang, D.Y., Banerji, S., Stransky, N., Dayton, T.L., Snyder, E.L., Senna, S., Whittaker, C.A., Bronson, R.T., et al. (2011). Nuclear factor I/B is an oncogene in small cell lung cancer. Genes Dev. *25*, 1470–1475.

DuPage, M., Dooley, A.L., and Jacks, T. (2009). Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. Nat. Protoc. *4*, 1064–1072.

Georgakoudi, I., Solban, N., Novak, J., Rice, W.L., Wei, X., Hasan, T., and Lin, C.P. (2004). In vivo flow cytometry: a new method for enumerating circulating cancer cells. Cancer Res. *64*, 5044–5047.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Jahchan, N.S., Lim, J.S., Bola, B., Morris, K., Seitz, G., Tran, K.Q., Xu, L., Trapani, F., Morrow, C.J., Cristea, S., et al. (2016). Identification and Targeting of Long-Term Tumor-Propagating Cells in Small Cell Lung Cancer. Cell Rep. *16*, 644–656.

Kato, F., Fiorentino, F.P., Alibés, A., Perucho, M., Sánchez-Céspedes, M., Kohno, T., and Yokota, J. (2016). *MYCL* is a target of a BET bromodomain inhibitor, JQ1, on growth suppression efficacy in small cell lung cancer cells. Oncotarget *7*, 77378–77388.

Lenhart, R., Kirov, S., Desilva, H., Cao, J., Lei, M., Johnston, K., Peterson, R., Schweizer, L., Purandare, A., Ross-Macdonald, P., et al. (2015). Sensitivity of Small Cell Lung Cancer to BET Inhibition Is Mediated by Regulation of ASCL1 Gene Expression. Mol. Cancer Ther. *14*, 2167–2174.

Lohr, J.G., Adalsteinsson, V.A., Cibulskis, K., Choudhury, A.D., Rosenberg, M., Cruz-Gordillo, P., Francis, J.M., Zhang, C.Z., Shalek, A.K., Satija, R., et al. (2014). Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. Nat Biotechnol *32*, 479–484.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat. Neurosci. *13*, 133–140.

McDavid, A., Finak, G., Chattopadyay, P.K., Dominguez, M., Lamoreaux, L., Ma, S.S., Roederer, M., and Gottardo, R. (2013). Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. Bioinformatics *29*, 461–467.

McFadden, D.G., Papagiannakopoulos, T., Taylor-Weiner, A., Stewart, C., Carter, S.L., Cibulskis, K., Bhutkar, A., McKenna, A., Dooley, A., Vernon, A., et al. (2014). Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. Cell *156*, 1298–1311.

Miyamoto, D.T., Zheng, Y., Wittner, B.S., Lee, R.J., Zhu, H., Broderick, K.T., Desai, R., Fox, D.B., Brannigan, B.W., Trautwein, J., et al. (2015). RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. Science *349*, 1351–1356.

Nedosekin, D.A., Verkhusha, V. V., Melerzanov, A. V., Zharov, V.P., and Galanzha, E.I. (2014). In vivo photoswitchable flow cytometry for direct tracking of single circulating Tumor Cells. Chem. Biol. *21*, 792–801.

Ozkumur, E., Shah, A.M., Ciciliano, J.C., Emmink, B.L., Miyamoto, D.T., Brachtel, E., Yu, M., Chen, P. -i., Morgan, B., Trautwein, J., et al. (2013). Inertial Focusing for Tumor Antigen-Dependent and -Independent Sorting of Rare Circulating Tumor Cells. Sci. Transl. Med. *5*, 179ra47-179ra47.

Pantel, K., and Speicher, M.R. (2016). The biology of circulating tumor cells. Oncogene *35*, 1216–1224.

Parasuraman, S., Raveendran, R., and Kesavan, R. (2010). Blood sample collection in small laboratory animals. J. Pharmacol. Pharmacother. *1*, 87.

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B. V, Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

Rhim, A.D., Mirek, E.T., Aiello, N.M., Maitra, A., Bailey, J.M., McAllister, F., Reichert, M., Beatty, G.L., Rustgi, A.K., Vonderheide, R.H., et al. (2012). EMT and dissemination precede pancreatic tumor formation. Cell *148*, 349–361.

Sarioglu, a F., Aceto, N., Kojic, N., Donaldson, M.C., Zeinali, M., Hamza, B., Engstrom, A., Zhu, H., Sundaresan, T.K., Miyamoto, D.T., et al. (2015). A microfluidic device for label-free, physical capture of circulating tumor cell clusters. Nat. Methods *12*, 1–10.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236–240.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M. a, Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. *102*, 15545–15550.

Sutherland, K.D., Proost, N., Brouns, I., Adriaensen, D., Song, J.Y., and Berns, A. (2011). Cell of origin of small cell lung cancer: Inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. Cancer Cell *19*, 754–764.

Ting, D.T., Wittner, B.S., Ligorio, M., Vincent Jordan, N., Shah, A.M., Miyamoto, D.T., Aceto, N., Bersani, F., Brannigan, B.W., Xega, K., et al. (2014). Single-Cell RNA Sequencing Identifies Extracellular Matrix Gene Expression by Pancreatic Circulating Tumor Cells. Cell Rep. *8*, 1905–1918.

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189–196.

Unger, M.A., Chou, H.P., Thorsen, T., Scherer, A., and Quake, S.R. (2000). Monolithic microfabricated valves and pumps by multilayer soft lithography. Science *288*, 113–116.

Vishnoi, M., Peddibhotla, S., Yin, W., T. Scamardo, A., George, G.C., Hong, D.S., and Marchetti, D. (2015). The isolation and characterization of CTC subsets related to breast cancer dormancy. Sci. Rep. *5*, 17533.

Yu, M., Stott, S., Toner, M., Maheswaran, S., and Haber, D.A. (2011). Circulating tumor cells: approaches to isolation and characterization. J. Cell Biol. *192*, 373–382.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.

Zettergren, E., Vickers, D., Runnels, J., Murthy, S.K., Lin, C.P., and Niedre, M. (2012). Instrument for fluorescence sensing of circulating cells with diffuse light in mice in vivo. J. Biomed. Opt. *17*, 037001.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049.