Expanding cancer therapy options through genome-scale identification of synthetic lethal paralogs

Phoebe C. R. Parrish

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Alice H. Berger, Chair

Robert N. Eisenman

Robert K. Bradley

Program Authorized to Offer Degree:

Genome Sciences

University of Washington

Abstract

Expanding cancer therapy options through genome-scale identification of synthetic lethal paralogs

Phoebe C. R. Parrish

Chair of the Supervisory Committee:
Associate Professor Alice H. Berger
Human Biology Division, Fred Hutchinson Cancer Center

Synthetic lethal therapies have the potential to expand cancer treatment options. Synthetic lethality is a form of context-dependent gene essentiality in which dual inactivation of a gene pair leads to cell death but single-gene inactivation does not affect viability. When one synthetic lethal gene is inactivated in cancer, targeting the remaining gene results in tumor cell death while normal cells remain unaffected. Therefore, synthetic lethal therapies may show substantially reduced toxicity compared to chemo- or radiotherapy and oncogene-targeted therapies. The successful use of PARP inhibitors to treat *BRCA1* or *BRCA2*-mutant breast, ovarian, and prostate cancers shows that synthetic lethal therapies are a viable approach. Yet *PARP-BRCA* remains the only synthetic lethal target with a clinically-approved inhibitor to date. There thus is an urgent need to find more targetable synthetic lethal interactions in cancer.

Given the Berger lab's focus on lung cancer, we developed a dual knockout CRISPR method to systematically identify synthetic lethal human gene pairs in lung cells. Previous studies showed that less than 3% of unrelated human gene pairs are synthetic lethal, while duplicated yeast genes showed a 25% synthetic lethal hit rate. We thus chose to identify new synthetic lethal lung cancer drug targets by developing a pooled dual-targeting CRISPR-Cas9 library called paired guide RNAs for paralog genetic interaction mapping (pgPEN). pgPEN targets over 2,000 human paralogs, or duplicated genes. We applied pgPEN to lung and cervical cancer cell lines and found that 12% (n = 122) of paralog pairs exhibited synthetic lethality in at least one context. These synthetic lethal paralogs represent new potential cancer therapeutic targets.

We also developed two computational methods to ensure that pgPEN can be applied by other researchers and that targeting paralogs is a viable synthetic lethal therapy approach. The first tool, paired guide mapper (pgMAP), is a user-friendly software pipeline that reproducibly maps sequencing reads from large, dual-targeting CRISPR sequencing datasets. The second method leverages existing gene expression and drug sensitivity datasets to identify 68 cases where loss of one paralog led to significantly decreased cancer cell viability upon treatment with a drug targeting that gene family. These analysis methods enable other groups to find synthetic lethal paralog pairs in any cancer type and support targeting paralogs as a potentially viable therapeutic approach.

This work reveals over 100 novel, potentially targetable synthetic lethal interactions in human cancer cells that can be further tested and translated to the clinic. Additionally, the pgPEN CRISPR library and the pgMAP pipeline will enable other researchers to identify targetable synthetic lethal interactions in other cancer types. Synthetic lethal therapy has the potential to provide a relatively low-toxicity treatment approach that can expand cancer therapy options, help address the challenge of acquired drug resistance, and improve patient outcomes.

Non-Technical Summary

How can we slow or stop tumor growth, or even cure cancer, without harming the healthy cells in a patient's body? This is the central question I have tackled in my dissertation research in cancer biology. I developed a new technology that will accelerate the identification of drug targets that kill cancer cells without harming healthy tissue. Current commonly-used cancer therapies like chemotherapy and radiation target tumor cells as well as fast-growing healthy cells in a patient's body, which is why they cause negative side effects like hair loss, gastrointestinal problems, dizziness, and fatigue. My PhD research has helped to advance a therapeutic approach that uses drugs to specifically target and kill cancer cells, so it is less toxic to patients and more effective at shrinking or completely eliminating tumors. I uncovered hundreds of new cancer targets by developing an approach to simultaneously inactivate thousands of human genes and by analyzing large datasets of cancer cell growth in the presence of targeted drugs. I have also shared my experimental methods and analysis pipelines freely so other researchers can identify drug targets in any cancer type.

To find more cancer drug targets, I first had to develop a new method to inactivate two genes at once in human cancer cells. This is because the therapeutic approach I study is based on synthetic lethality, or an interaction between two genes where inactivating either gene alone does not affect cell growth but inactivating both genes at once kills the cell. Synthetic lethal therapies exploit the built-in backup systems for key cellular processes like growth and division to kill cancer cells. These cellular backup systems are like a car's headlights: if one goes out, you can still drive safely but you are now very dependent on the function of the remaining headlight. If both

headlights go out at once, it creates a dangerous situation that all drivers want to avoid. Given that cancer cells usually have many more mutations than healthy cells, we will likely find cases where one gene in a synthetic lethal pair is inactivated by a mutation in a patient's tumor but not in their healthy tissue. By targeting the remaining gene with a drug, we can kill tumor cells without harming healthy cells. In 2005, scientists discovered the first synthetic lethal therapy that has been used to successfully treat patients with breast, ovarian, and prostate cancer. But since then, researchers have been unable to find any other synthetic lethal therapies that can be used to treat cancer in the clinic.

I decided to look for new synthetic lethal drug targets so that more cancer patients could benefit from this low-toxicity therapeutic approach. When I started my PhD in 2018, biologists did not have a way to inactivate, or "knock out", two human genes at once on the scale of thousands of gene pairs. I therefore developed a new, large-scale method to simultaneously inactivate thousands of gene pairs in human cancer cells. My approach uses CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) to do genetic loss-of-function experiments in which 2,060 genes are inactivated alone and in pairs. I targeted paralogs, or duplicated genes that arose when one ancestral gene was accidentally copied to produce two genes that then diverged. CRISPR screens of non-duplicated human gene pairs showed a synthetic lethal hit rate of less than 0.1% but 25% of paralogs are synthetic lethal in yeast, a genetically tractable organism that biologists use to study human genetics. We named our method pgPEN (paired guide RNAs for Paralog gENetic interaction mapping), and we refer to the set of CRISPR guide RNAs that we designed to target our duplicated genes of interest as a "library". pgPEN was the largest paralog dual-targeting CRISPR library ever developed.

Though our new method can be applied to find synthetic lethal drug targets in any cancer type, we first used pgPEN in lung and cervical cancer cells. We chose these cancer types because lung and cervical tumors are highly mutated relative to other cancers, so we would be more likely to find cases where one synthetic lethal gene was mutated in the tumor but not in healthy tissue. Using the pgPEN approach, I found 122 new, high-confidence synthetic lethal gene pairs. This translates to a hit rate of 12%, which is much higher than the synthetic lethal hit rates seen in previous studies. Moreover, we found that over a quarter of our hits were classified as drug targets. I published this work in 2021 (Parrish et al., 2021) and my paper has since been cited over 35 times and was highlighted by multiple science news outlets[1]. I also shared the pgPEN library via an online repository and it has been adopted by 10 labs around the world to find synthetic lethal targets in cancer and other diseases.

It has been very rewarding to dive into the field of synthetic lethal drug discovery to find new cancer drug targets during my PhD research. To enable other researchers to find synthetic lethal targets, I am developing software packages to share the code I used for my pgPEN studies[2]. To ensure that targeting duplicated genes with drugs is a viable approach, I also developed a computational approach to analyze large-scale datasets generated by treating cells derived from many different cancer types with paralog-targeting drugs. Altogether, my PhD work has advanced the goal of identifying synthetic lethal therapy targets and has had a global impact on the cancer research field. After further study, synthetic lethal therapies could be implemented in the clinic to treat or even cure cancer while improving quality-of-life for cancer patients.

---

[1] See articles from: Fred Hutch Science Spotlight, Fred Hutch News, and Science in Seattle.

[2] See: https://github.com/FredHutch/pgMAP_pipeline and https://github.com/FredHutch/GI_mapping.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I want to thank the entire Berger lab, particularly Dr. Alice Berger, for helping me think through, conduct, and interpret all the scientific experiments and analyses in this dissertation. Alice inspired me with her creative thinking and excitement about science from our very first meeting where we discussed a potential rotation in the lab and she pitched the project that became pgPEN to me. Her encouragement, support, and assistance with troubleshooting experiments was invaluable throughout my PhD research. I also want to thank many past and present members of the Berger lab who helped me get my project off the ground and supported me throughout the highs and lows of scientific research. In roughly chronological order, this includes Dr. Athea Vichas, Dr. April Lo, Dr. Sitapriya Moorthi, Amanda Riley, Shriya Kamlapurkar, Amanda Bradley, Heather Borror, Callie Rominger, and Daniel Groso. I also want to thank all my mentees for helping me stay motivated and reminding me that science is exciting.

Thanks to many past and present members of Dr. Rob Bradley's lab, who provided me a second scientific home and served as sounding boards and key collaborators for my experiments throughout graduate school. Rob himself provided extremely helpful advice to me and Alice throughout the pgPEN and pgMAP development and publication processes. I also want to acknowledge Dr. James Thomas, Dr. Austin Gabel, Dr. Emma De Neef, and Dr. Emma Hoppe.

My thesis committee-members Dr. Bob Eisenman, Dr. Maitreya Dunham, Dr. Rob Bradley, and Dr. Jesse Bloom have served as an incredibly useful sounding board and helped guide me through my PhD research and beyond. I particularly want to thank Rob and Bob for serving on my reading committee and Maitreya for helping me identify key papers on yeast genetic interaction

mapping and for helpful advice on preparing for my defense, applying for graduate school awards, and post-PhD career options.

None of my work would have been possible without the support of amazing staff and administrators in both the Fred Hutch Human Biology Division and the UW Genome Sciences Department. I would particularly like to thank the Berger lab's amazing admin James Adams and the Genome Sciences graduate program admin Brian Giebel for helping me navigate the often-confusing red tape of academia across two institutions. Thanks also to the Fred Hutch Offices of Graduate Education and Scientific Career Development for advice on navigating grad school and post-PhD career options. Finally, none of the computational analyses included in this thesis would have been possible without the support of the amazing Information Technology and Scientific Computing teams at UW and Fred Hutch. Thank you to Skylar Thompson, Charles Winston, and Justin Reichel at UW and Luna Yu, Pat Heath, and David Chambers at Fred Hutch.

I have also been involved in a number of teaching and service opportunities throughout graduate school. I would like to thank all past and present members of the Genome Sciences Association for the Inclusion of Marginalized Students (GSAIMS) for their fellowship, support, and collaboration in promoting diversity, equity, and inclusion in science. Dr. Atom Lesiak provided invaluable advice on promoting change, applying for funding, and organizing outreach events. I also want to thank Maureen Larsen, the GS Department Director, for helping us coordinate funding for GSAIMS events and for her help when I served as the graduate student representative during GS faculty meetings. Dr. Hannah Jordt and Dr. M. K. Raghuraman provided extremely helpful advice and support during my teaching assistantships. I aspire to be half as good at teaching and science communication as they are!

The Brewer-Raghuraman lab was the first UW lab I rotated in, and I want to thank all of the lab-members for helping me find my equilibrium during my first quarter. They helped me feel welcome in the department and have continued to support me by giving extremely valuable feedback during practice talks and providing a sounding board for my extracurricular activities and career plans. Thank you to Dr. Bonny Brewer, Dr. M. K. Raghuraman, Dr. Liz Kwan, Dr. Rebecca Martin, and Gina Alvino.

I never would have chosen a career in science or applied to graduate school without the support of past professors and mentors. Dr. A. Malcolm Campbell and Dr. Kevin G. Smith at Davidson College helped encourage me to choose Biology as a major and to think through the broader ethical, legal, and social implications of science research. Dr. Patricio Boyer, Dr. Patricia Tilburg, and many other humanities professors at Davidson taught me to think critically about gender, history, colonialism, linguistics, philosophy, and so much more. Additionally, my research experiences in Dr. Campbell's lab and in Dr. Alan Collmer's lab in the Cornell University School of Integrative Life Sciences helped me realize I could pursue science research as a career. Particular thanks to Dr. Collmer, Dr. Suma Chakravarthy, and Dr. Hai-Lei Wei for their support during my undergraduate Honors thesis research in plant-microbe biology.

My post-baccalaureate fellowship in Dr. Beth Kozel's lab at the National Institutes of Health taught me what it would be like to do science full time. And the experience was so positive, I decided to go to graduate school! Thank you to Beth for taking a chance on me and giving me a dataset of whole-exome sequences to analyze even though I had extremely limited coding experience. Thanks also to Russ Knutsen and the other members of the Kozel lab for their support during my postbac and beyond. You taught me what working full-time in a healthy, supportive, and fun lab environment looks like.

Though my mentors and lab-mates have supported me throughout my career, I truly wouldn't have been able to get to graduate school—let alone successfully defend my PhD—without the support of an incredible network of friends and family both within and outside the world of science. I want to thank my grad school friends especially for helping me navigate the complexities of academic science all the way from the first day of my first year to submitting my dissertation. Shoutout to Robin Aguilar, Thomas Pulliam, Katharine Chen, Emma De Neef, Zorian Thornton, Alberto Rivera, Ken Jean-Baptiste, Maddy Duran, Laila Shehata, Andie Doak, and so many other UW graduate students in both the Genome Sciences and Molecular and Cellular Biology PhD programs for advising, commiserating, and celebrating with me throughout this process. I especially want to thank Robin for role as a friend and co-leader in the DEI-focused work we did through GSAIMS. Even when my science wasn't working, the events we ran and the space we created within the department made my time here feel useful and important.

Additionally, my friends from outside UW provided invaluable support and perspective throughout my time in graduate school. Thank you to Jenna Goff, Anna Cox, Teta Alim, Amelia Kaufman, Rachel Wiltshire, Grace Shuk Hang Li, Jose Balcazar, An-Phuong Ha, Zoë McAlear, Gabe Nussbaum, and too many other people to list here. Also thanks to all the baristas and restaurant workers, particularly those at Big Mario's in Lower Queen Anne and Café An' Clair, who kept me caffeinated and fed while I was writing my thesis and who checked in on me after my defense to see how it went. It's been wonderful having a community in Seattle and beyond, and for those of you who couldn't come to my defense in person I can't wait to celebrate together soon!

Finally, I would like to thank my family. My extended family has cheered me on throughout my entire 21 years of education—I promise I'm done now! My grandmother, Sally

Nero, has always been just a phone call away when I need her and always welcomes me to her house in Maine, my favorite place in the world to relax and eat blueberry muffins. My grandfather Russell Parrish tells the best stories and gives excellent advice, and also makes delicious scones for his grandchildren when we visit. Thank you for helping me get to know Grandma better even though I couldn't make it to New Zealand during the pandemic. To my cousins, aunts, and uncles in the U.S. and New Zealand, thank you for giving me places to stay on my travels and for your support throughout this process! To my sister, Tessa Parrish, you are the smart one for not getting a PhD. Thank you for always being there when I need you, for your generally snarky but always helpful advice, and for the cute cat pictures. To my parents, Dr. Colin Parrish and Dr. Debra Nero, thank you for showing me what a being a smart, thoughtful, kind scientist, teacher, and mentor looks like. Whenever I need help or advice, you have both always been there to help me navigate the highs and lows of school, adulthood, career planning, and life. And thank you for being excellent dog grandparents to my dog, Jack. I guess now I'll have to get a real job so I can support him!

DEDICATION

This dissertation is dedicated to my mom, Dr. Debra Nero. From helping me understand Punnett squares in high school biology to answering my panicked phone calls and helping me problem-solve when I'm stressed in graduate school, you've always been there for me. Thank you for being my role model as a strong, opinionated, kind, intelligent female scientist. I couldn't have done any of this without you.

Chapter 1. INTRODUCTION

1.1 CURRENT STATE OF THE FIELD OF SYNTHETIC LETHAL CANCER TARGET DISCOVERY

Finding cancer therapies that can kill tumor cells without harming a patient's healthy cells is a major challenge for biomedical research. Current widely used therapeutic approaches like chemotherapy and radiation target cancer as well as other fast-growing cells in a patient's body. Though these approaches kill tumor cells effectively, doctors must carefully calibrate the dosage of chemo- and radiotherapy because of negative off-target effects including anemia, hair loss, and gastrointestinal disease.

To develop therapies that specifically target tumor cells, cancer biologists have leveraged advances in sequencing technology to better understand the genetic alterations that drive tumor formation. This research has resulted in the development of targeted therapies, in which patients are treated with drugs that target their tumor's specific mutation. Targeted therapies can quickly and potently kill tumor cells, but these mutation-specific drugs often still have serious side effects because many of the targeted genes are essential, or required for growth, in both tumor and normal cells. To more effectively treat, or even cure, cancer while improving patient quality-of-life, we must identify new therapeutic targets that are essential in tumor cells but not required for healthy cell growth.

In 1997, Dr. Leland Hartwell and colleagues proposed a solution to this problem (Hartwell et al., 1997). Rather than targeting genes that are required in both cancerous and healthy cells, we can leverage the fact that cancer cells are highly mutated to find cases where a context-specific essential gene is inactivated only in the tumor. Specifically, we can treat cancer by exploiting

synthetic lethality, a type of genetic interaction in which inactivating either gene in a pair alone does not affect cell growth but dual inactivation leads to cell death (**Figure 1.1A**). If one member of a synthetic lethal gene pair is inactivated via mutation only in the tumor, targeting the remaining gene with a drug will result in cancer cell death while healthy cell viability remains unaffected (**Figure 1.1B**).

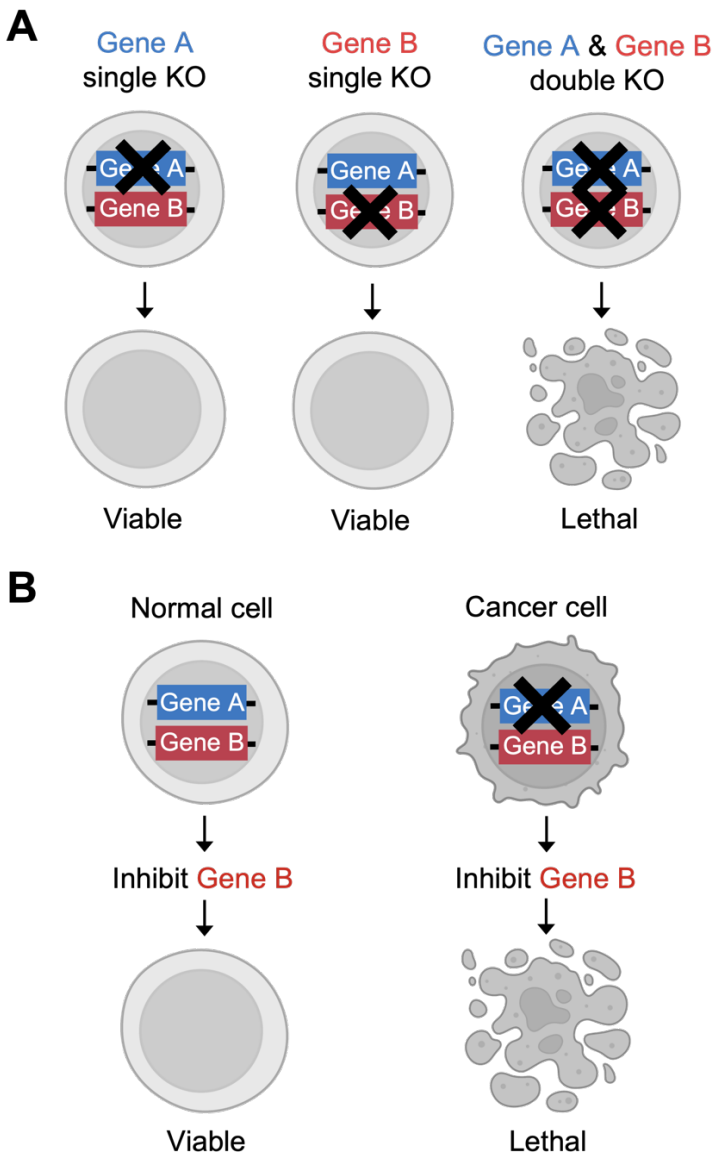


Figure 1.1: Synthetic lethality and its applications for cancer therapy.

A) Illustration of synthetic lethality in healthy human cells. KO, knockout.

B) Illustration of the synthetic lethal therapy approach for targeting cancer cells. Note that the loss of Gene A is tumor-specific. Figure created using BioRender.com.

When Dr. Hartwell et al. proposed this therapeutic approach, the cancer research community lacked the tools required to identify synthetic lethal targets in human cells. Given that there are over 200 million possible pairwise interactions between human genes (20,000 x 20,000 genes), mapping all possible genetic interactions in human cells would require a massive effort. At the time, individual human cancer genes such as *KRAS* and *TP53* had been identified. Yet the entire human genome had not yet been sequenced and mapped, which prevented the large-scale identification of single genes essential for growth—let alone gene pairs required for cancer growth.

The completion of the Human Genome Project in 2001 (Lander et al., 2001; Venter et al., 2001) provided a map for researchers to individually target and functionally profile every human gene. To identify essential genes, scientists have applied genetic loss-of-function screening approaches to simultaneously target and inactivate thousands of human genes in a pooled format. Typically, the goal is to inactivate, or "knock out", one gene per cell within the pooled population. Thus any cells containing essential gene knockouts will be out-competed by fitter cells whose gene knockouts do not affect cell growth.

The first method developed to inactivate individual genes at scale was RNA interference (RNAi), which uses small noncoding RNAs to disrupt gene expression in a targeted fashion (Fire et al., 1998; Harland & Weintraub, 1985; Izant & Weintraub, 1984; Nellen & Lichtenstein, 1993). Once researchers developed methods to introduce these noncoding RNAs into human cells (Caplen et al., 2001; Elbashir et al., 2001), they leveraged RNAi screening methods to perturb each of the 20,000 genes in the human genome individually with the goal of determining which genes were required for cancer cell growth (Baldwin et al., 2010, 2008; Bommi-Reddy et al., 2008; Grueneberg, Degot, et al., 2008; Grueneberg, Li, et al., 2008).

But because RNAi screens were difficult to validate due to off-target effects (X. Lin et al., 2007; Smith et al., 2017), the rigorous identification of essential cancer genes had to wait until the development of the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) system (O'Neil et al., 2017). The genome editing capabilities of CRISPR were first discovered in bacteria in 2012 (Gasiunas et al., 2012; Jinek et al., 2012), and subsequent work enabled the application of CRISPR for genome-scale loss-of-function screens in human cells (Cong et al., 2013; Jinek et al., 2013). CRISPR is thus the current state-of-the-art technique for identifying essential genes because it can both precisely target and effectively inactivate target genes. CRISPR screens enable cancer biologists to mimic the action of targeted therapies to determine which genes are required for cancer cell growth and should be prioritized for drug development.

Large-scale RNAi- and CRISPR-based efforts to identify individual genes required for cancer cell growth have since been applied to thousands of cancer cell lines (Pacini et al., 2021; Tsherniak et al., 2017). Additional studies have sequenced thousands of human tumors to identify somatic mutations driving cancer growth as well as vulnerabilities that could be targeted to prevent cancer growth (Bailey et al., 2018; Cancer Genome Atlas Research Network et al., 2013; Cheng et al., 2015; Hoadley et al., 2018). Finally, many groups have also carried out large-scale screens using existing drugs to determine whether these drugs could be applied or repurposed for cancer treatment (Corsello et al., 2017, 2020; Garnett et al., 2012; Yang et al., 2013). Together these datasets are a rich resource for identifying individual genes required for cancer cell growth (often referred to as "dependencies"), identifying tumors and cancer subtypes with potentially targetable mutations in essential cancer genes, and determining which existing drugs can be used for cancer treatment. Yet though these sequencing and single-gene loss-of-function screening datasets have improved our understanding of cancer biology and human gene essentiality they cannot be used to

efficiently identify which genes are cancer-specific versus genes that are also needed by normal healthy cells on a genome scale.

1.1.1 *Overview of cancer therapeutic approaches*

Cancer is one of the leading causes of death worldwide, and cancer incidence is growing rapidly due to aging and population growth (Sung et al., 2021). In 2020, approximately 20 million people were diagnosed with cancer and another 10 million people died of cancer around the world (Sung et al., 2021). Finding better ways to detect, treat, and ultimately cure cancer is therefore an urgent priority for the global biomedical research community.

1.1.1.1 Conventional treatment approaches

First-line treatment for most cancer patients involves some combination of surgery, chemotherapy, and radiation. These approaches were first discovered in the 19th and 20th centuries, and studies have demonstrated that each approach has some effectiveness at shrinking or ablating tumors and prolonging patient lifespans (Schirrmacher, 2019). However, each approach also has severe and often disfiguring or debilitating side effects (Chargari et al., 2016; Schirrmacher, 2019; Tilsed et al., 2022). Below, I summarize the benefits and drawbacks of each approach.

Before physicians and scientists began to elucidate the molecular mechanisms driving tumor formation and metastasis, surgery was the only effective treatment available to treat cancer patients. Indeed, early removal of the primary tumor can extend patient lifespans (Debela et al., 2021). But once the tumor has metastasized, or spread beyond the primary site, relapse will occur (Schirrmacher, 2019). Moreover, depending on the tissue type in which tumors develop, many cancer types such as lung cancer may be difficult to detect and to treat surgically.

Chemotherapy, or the use of chemical drugs to kill tumor cells, can overcome the challenge of metastasis because it is a systemic approach and can thus kill cancer cells regardless of their

location in the body. The two main types of chemotherapy are cytostatic and cytotoxic drugs: cytostatic drugs prevent cell division from occurring by targeting the cell cycle (Schirrmacher, 2019) and cytotoxic drugs kill existing cells by disrupting DNA synthesis or replication in order to trigger a cell death response (Bracci et al., 2014). Chemotherapy can be more effective than surgery at treating late-stage cancer (Debela et al., 2021) and can be used in combination with surgery or other therapeutic approaches to improve treatment outcomes at all stages of cancer progression (Schirrmacher, 2019; Tilsed et al., 2022). But chemotherapy also has major drawbacks: it targets all fast-growing cells in the body (Debela et al., 2021) and thus can have extremely short- and long-term severe side effects including gastrointestinal disease, infertility, and neurotoxicity leading to paralysis or coma (Schirrmacher, 2019). Chemotherapy drugs can also cause off-target toxicity in tissues like the heart and lung where cells are not rapidly dividing (Kaelin, 2005). Moreover, chemotherapy contributes relatively little to reducing cancer survival rates (one study estimated an impact of ~2%) and resistance to chemotherapeutic drugs frequently develops (Debela et al., 2021; Schirrmacher, 2019).

Radiation therapy, or radiotherapy, is a third commonly used cancer treatment approach in which X-rays or other types of radiation are used to kill cancer cells. Radiation therapy is highly effective in treating many cancer types (Baskar et al., 2012; Thariat et al., 2013) and modern radiotherapy approaches in particular can have limited off-target effects in normal tissue compared to chemotherapy (Baskar et al., 2012; Chandra et al., 2021; Thariat et al., 2013). Yet since radiation therapy affects both healthy and cancer cells, it can severely damage tissues and organs near the tumor site (Debela et al., 2021). Additionally, radiotherapy is still a localized form of treatment and cannot target cancer cells that have metastasized or tumors that have not yet been detected.

1.1.1.2 Oncogene-targeted therapy

Like chemotherapy, oncogene-targeted therapy uses chemical drugs, often called "small molecule inhibitors", to kill cancer cells by targeting tumor-specific genetic alterations that drive cancer growth. These mutated cancer driver genes are known as oncogenes. The major benefit of oncogene-targeted therapy over chemotherapy is that, in principle, oncogene-targeted drugs should specifically kill cancer cells. Chemotherapy drugs have a very narrow therapeutic index, which is a measure of the toxic dose of a drug relative to its therapeutic dose. On the other hand, since oncogene-targeted drugs are designed to target tumor-specific mutations they should theoretically be able to kill cancer cells across a broader range of doses without harming normal cells (Kaelin, 1999).

The first FDA-approved oncogene-targeted therapy, imatinib, targets the oncogenic BCR-ABL fusion mutation in chronic myeloid leukemia (CML). The first clinical trials for imatinib took place in 1998, and subsequent studies showed that oncogene-targeted drug was highly effective at treating leukemia and had fewer side effects relative to chemotherapy (Druker et al., 2001; Iqbal & Iqbal, 2014). The discoverers of imatinib were awarded a clinical research Lasker Award for "'converting a fatal cancer into a manageable condition'" (Iqbal & Iqbal, 2014), highlighting the impact of oncogene-targeted therapy on cancer treatment.

Another oncogene-targeted therapy, the EGFR inhibitor gefitinib, was discovered in 2004. The advent of more advanced sequencing approaches enabled physicians to group lung cancer patients treated with EGFR inhibitors based on the presence or absence of *EGFR* activating mutations in their tumors. Researchers thus discovered that gefitinib was specifically effective in patients with *EGFR* mutations (Lynch et al., 2004; Paez et al., 2004). Another group found that in addition to gefitinib, a second EGFR inhibitor, erlotinib, showed similar effects (Pao et al., 2004).

Dozens of targeted therapies have since been developed and used to prolong patient survival across many cancer types (Bedard et al., 2020; Debela et al., 2021). To date, targeted therapies have mainly been used to treat chemotherapy-resistant cancers, but they show fewer side effects relative to cytotoxic chemotherapy (Bedard et al., 2020) and some targeted therapeutics are used in combination with chemotherapy to increase treatment effectiveness (Bedard et al., 2020; Debela et al., 2021). In non-small cell lung cancer (NSCLC), more recent studies have shown that targeted therapies induce stronger responses to treatment and prolong survival relative to chemotherapy. Overall, NSCLC patients with targetable mutations that are treated with oncogene-targeted therapy live up to 2-3 years longer than those without (Yoda et al., 2019).

Yet there are major drawbacks to oncogene-targeted therapy. One limitation is the severity of negative side effects such as gastrointestinal and skin toxicity, particularly in cases where the targeted gene is highly expressed in both cancer cells and in normal tissue (Bedard et al., 2020; Schirrmacher, 2019). Drugs that target oncogenes with a single nucleotide variant (SNV) driving cancer may have a narrower therapeutic window, since it can be more difficult to design drugs that selectively inhibit the oncogenic protein (Bedard et al., 2020) compared to drugs that target a fusion oncogene or the amplification of a gene that is lowly-expressed in normal tissue. Another drawback to oncogene-targeted therapy is that the selective pressure exerted on tumor cells by the drug often drives the emergence of resistant sub-clones, requiring the development of next-generation inhibitors to treat patients whose tumors have acquired resistance to first-line targeted drugs (Bedard et al., 2020; Yoda et al., 2019). Finally, rather than being driven solely by oncogene hyperactivation many tumors are driven by the loss of a tumor suppressor gene or by dual activation of an oncogene and inactivation of a tumor suppressor. Thus, only patients with oncogene-driven cancers can benefit from oncogene-targeted therapies.

1.1.1.3 Synthetic lethal therapy

By 1997, it was clear that cancer was a disease driven by genetic mutations and the technology was developed to profile these mutations at scale. Biologists including Leland Hartwell at the Fred Hutchinson Cancer Center proposed using genetic methods to identify new, tumor-specific cancer drug targets (Hartwell et al., 1997). These researchers proposed using genetic screens to mimic the action of drugs, thus identifying new cancer targets. At the time, it was impossible to carry out genome-scale screens in human cells—thus, Dr. Hartwell and colleagues argued that genetic studies and drug screens in model organisms could be used to identify key genes required for cancer growth. Specifically, they proposed using synthetic lethal screens in the model organism *Saccharomyces cerevisiae*, a species of yeast, to expand drug target identification beyond the focus on inhibiting oncogene products to include tumor suppressor-driven cancers (Hartwell et al., 1997). This group and others also noted that synthetic lethal approaches could widen the therapeutic window at which drugs could kill tumor cells without harming normal tissue, which would be a major benefit compared to chemotherapeutic approaches (Hartwell et al., 1997; Kaelin, 1999, 2005).

After the human genome was sequenced and large-scale genetic screening methods like RNAi were developed that could be applied in human cells, the identification of synthetic lethal cancer drug targets became a more realistic goal (Kaelin, 2005). In fact, the first targetable synthetic lethal interaction was discovered in 2005—though not through genetic screens. Instead, two research groups studying DNA repair in cancer found that PARP inhibitors can be used *in vitro* and *in vivo* to selectively kill human and mouse cells containing inactivating germline mutations in *BRCA1* or *BRCA2* (Bryant et al., 2005; Farmer et al., 2005). It took 10 years for this finding to be translated to the clinic—but PARP inhibitors are now used to treat a range of cancers

with mutations in the *BRCA1/BRCA2* gene family including breast, ovarian, and prostate cancer (Ashworth & Lord, 2018; Lord & Ashworth, 2017).

Since the development of genome-wide CRISPR screening approaches, more synthetic lethal therapy targets are being identified and carried forward to clinical trials (Huang et al., 2020; Mullard, 2022; Setton et al., 2021). Synthetic lethal approaches targeting other DNA damage repair pathway-members include drugs targeting POLQ in *BRCA1/BRCA2*-mutant tumors that have developed resistance to PARP inhibitors (Ceccaldi et al., 2015; Mateos-Gomez et al., 2015) and the use of PKMYT1 inhibitors in *CCNE1*-amplified solid tumors (Gallo et al., 2022). Targets outside the DNA repair pathway have also been identified. These include inhibiting PRMT5 in *MTAP*-deleted cancers, since *MTAP* is located near the frequently-deleted tumor suppressor *CDKN2A* (Kryukov et al., 2016; Mavrakis et al., 2016), and targeting the DNA helicase WRN in microsatellite unstable cancers (Chan et al., 2019). Inhibitors targeting at least six synthetic lethal interactions in cancer are currently in development at a range of pharmaceutical companies (Mullard, 2022).

Though synthetic lethal therapies are overall in an earlier stage of development and clinical translation compared to other cancer therapeutic approaches, some drawbacks to their use in cancer have already been identified. Like other targeted therapies, synthetic lethal therapy exerts selective pressure leading to the emergence of resistant sub-clones, so PARP inhibitor resistance is a common phenomenon and new therapeutic strategies must be developed and applied to address resistance (Lord & Ashworth, 2017). Potential concerns for emerging synthetic lethal therapies include the penetrance of the targeted interaction (Ashworth & Lord, 2018; Huang et al., 2020; O'Neil et al., 2017), and there is ongoing debate in the field about the benefits and drawbacks of targeting "private", or context-specific, versus "robust", or broadly penetrant, synthetic lethal

interactions. The prevalence or zygosity of the targeted vulnerability is also a concern: for instance, biallelic inactivation of *BRCA1* or *BRCA2* leads to a much stronger effect of PARP inhibition relative to monoallelic inactivation (Setton et al., 2021). A final major concern is the magnitude of the synthetic lethal effect: cells with *BRCA1* or *BRCA2* loss-of-function mutations are up to 100 times more sensitive to PARP inhibitors relative to wildtype cells (Ashworth & Lord, 2018). Future studies will have to take the lessons learned from the development of PARP inhibitors, as well as knowledge gained from model organisms, to inform target selection and drug development.

1.1.2 *Lessons learned from yeast*

Studies in the genetically tractable, unicellular model organism yeast (*S. cerevisiae*) have been key in characterizing the prevalence and functional importance of synthetic lethal interactions in eukaryotic cells. Synthetic lethality was first described by researchers studying *Drosophila* (Boone et al., 2007; Phillips, 1998). But since technologies for multi-gene loss-of-function screens in yeast have preceded those in multicellular eukaryotic organisms and human cells, yeast geneticists could begin mapping digenic interactions for functionally important genes much earlier (Tong et al., 2004). Yeast geneticists have subsequently mapped digenic interactions at genome scale (Costanzo et al., 2010, 2016) and across multiple environmental conditions (Costanzo et al., 2021) as well as mapping trigenic interactions (Kuzmin et al., 2018). These studies have enabled yeast researchers to identify broad classes of genes that are likely to be synthetic lethal in both yeast and human cells (Deshpande et al., 2013).

In particular, genetic interaction studies in yeast have highlighted the increased prevalence of synthetic lethal interactions among paralogs (Dean et al., 2008; DeLuna et al., 2008; Diss et al., 2017; Guan et al., 2007; Harrison et al., 2007; Kuzmin et al., 2020). These studies have revealed that up to 25% of duplicated genes are synthetic lethal in yeast (Dean et al., 2008; DeLuna et al.,

2008) and that the rate of synthetic lethality varies depending on the type of duplication event that gave rise to the paralogs (Guan et al., 2007; Kuzmin et al., 2020). On the other hand, random gene pairs showed a < 0.5% synthetic lethal hit rate (Segrè et al., 2005; Tong et al., 2004), and even functionally-related singleton (non-duplicated) genes showed a < 3% synthetic lethal hit rate (Dean et al., 2008). Yeast studies have also revealed that many genetic interactions are dependent on genetic background or environmental context (Costanzo et al., 2021; Harrison et al., 2007).

Researchers studying genetic interactions in yeast and other model organisms have also developed statistical models to quantify the severity of synthetic lethal (negative or synergistic) and buffering (positive or antagonistic) genetic interactions (Boone et al., 2007; Collins et al., 2006; DeLuna et al., 2008; van Leeuwen et al., 2017). Many of these models draw on the relatively inclusive definition of "epistasis", or genetic interaction, developed by the statistical geneticist and eugenicist R. A. Fisher in 1919 (Boone et al., 2007; Felsenstein, 1965; Fisher, 1919). In Fisher's definition, epistasis occurs if the fitness consequence of mutating two genes simultaneously differs from the (multiplicative) product of the fitness effects of mutating the two genes individually. In other words, a genetic interaction occurs if:

$$(\text{Gene A mutant}) \times (\text{Gene B mutant}) \neq (\text{Gene A \& B double mutant})$$

Note that for $\log_2$-transformed values (which are often used in human gene fitness screens), this equation would instead be:

$$\log_2(\text{Gene A mutant}) + \log_2(\text{Gene B mutant}) \neq \log_2(\text{Gene A \& Gene B double mutant})$$

The work of geneticists studying yeast, *Drosophila*, and other model organisms has thus been fundamental to our understanding of how to map and quantify genetic interactions, including synthetic lethality.

1.2 CHALLENGES IN SYNTHETIC LETHAL DRUG TARGET DISCOVERY

One major barrier to the identification of synthetic lethal interactions between human genes is simply the number of gene pairs to be tested. Given that the human genome contains approximately 20,000 genes, there are 200 million possible pairwise gene knockout combinations to test. Additional complexity arises when moving beyond gene knockouts to look for synthetic lethal targets in the context of a single cancer driver mutation or of the genetic background associated with a given tumor or cancer subtype. Additionally, since yeast research has highlighted the fact that many genetic interactions are dependent on environmental context, it is important to consider whether the *in vitro* systems used for genetic screens accurately mimic the *in vivo* tumor context.

Another challenge to identifying synthetic lethal interactions is phenotypic robustness due to duplicated genes and multi-genic phenotypes. The human genome exhibits a high degree of redundancy as a result of diploidy, gene duplication, and the functional overlap of signaling pathways (Dandage & Landry, 2019; Dean et al., 2008; Harrison et al., 2007; Lavi, 2015; Ohno, 2013). Given the additional complexity inherent to multicellular organisms (relative to unicellular *S. cerevisiae*) and the extensive degree of duplication in the human genome (Dennis & Eichler, 2016; Lan & Pritchard, 2016; Singh et al., 2012), it is likely that experimental evaluation of human genetic interactions will reveal highly complex patterns of genetic interaction.

The complexity of human genetic interactions due to genetic and functional redundancy is likely to be further exaggerated in highly-mutated cancer genomes. Indeed, functional redundancy among human paralogs in cancer has been indirectly observed via analysis of data from single-gene CRISPR screens in cancer cell lines. These studies showed that inactivating individual paralogs tends to have a smaller effect on cell fitness than inactivating singleton gene knockouts

(Dandage & Landry, 2019; De Kegel & Ryan, 2019; Wang et al., 2015). However, there are a number of drawbacks to this approach which will be discussed in the following section.

To directly identify synthetic lethal cancer targets we must therefore use technologies that are highly scalable. For instance, rather than doing whole-genome single-gene knockout screens in all cancer genetic backgrounds it may be more efficient to first execute dual-gene knockout screens in a subset of cancer cell lines to identify likely synthetic lethal targets. This shorter list of likely synthetic lethal targets can then be studied in more detail in specific cancer-relevant genetic contexts and environmental conditions.

A final challenge in the search for human genetic interactions that can be leveraged for cancer therapy is the consideration of higher-order interactions. Though the classical definition of synthetic lethality is defined as an interaction between two genes, yeast researchers found that trigenic interactions occur ~100 times more frequently than digenic interactions, though their fitness effects tend to be weaker (Kuzmin et al., 2018). Thus, higher-order interactions should be taken into consideration when developing and applying technologies to search for cancer targets in human cells.

1.3 APPLYING GENOME EDITING TOOLS FOR SYNTHETIC LETHAL TARGET DISCOVERY IN CANCER

CRISPR is the current state-of-the-art tool for identifying genes required for human cell growth. Though RNAi was used in many early screens for human gene essentiality, the results of these screens were not reproducible and scientists concluded that RNAi simply had too many off-target effects to be applied effectively for essentiality screening at the genome scale (O'Neil et al., 2017; Smith et al., 2017), particularly in the context of multi-targeting screens to identify synthetic

lethal interactions. Since CRISPR approaches are more sensitive and specific than RNAi approaches, they have become the method of choice for high-throughput identification of essential genes (Doench et al., 2016; Hart et al., 2017; Kleinstiver et al., 2016; Morgens et al., 2017). Essential genes identified via CRISPR screening approaches represent potential targets for cancer treatment.

1.3.1 *Cas9-mediated CRISPR knockout*

The Cas9-mediated CRISPR knockout system is currently the most widely-used method for identifying human essential genes. As illustrated in **Figure 1.2**, this approach uses a guide RNA (gRNA), which is a ~20 nucleotide single stranded RNA that is complementary to a genomic target site of interest, to direct the Cas9 endonuclease to target sites near a protospacer adjacent motif (PAM). Cas9 generates a double strand break (DSB) at the target site of interest, triggering the cell's DNA repair pathway to initiate non-homologous end joining, a messy repair process that frequently generates small insertions and deletions (indels) at the target site of interest. Indels that occur in the DNA coding sequence frequently generate frameshift mutations which lead to premature stop codons, thereby disrupting protein function.

Since CRISPR genome editing was first discovered in *Streptococcus pyogenes* (Gasiunas et al., 2012; Jinek et al., 2012), initial methods used this version of the Cas9 enzyme, referred to as SpCas9. The SpCas9 enzyme has since been engineered to more efficiently target and cut DNA with fewer off-target effects (Chen et al., 2017; Kleinstiver et al., 2016; Slaymaker et al., 2016). The initial SpCas9 variant recognized a 3'-NGG PAM sequence, but the enzyme has since been engineered to recognize additional PAM sequences, expanding the range of sites that can be targeted via CRISPR-Cas9 approaches (Hu et al., 2018; Kleinstiver et al., 2015; Walton et al., 2020).

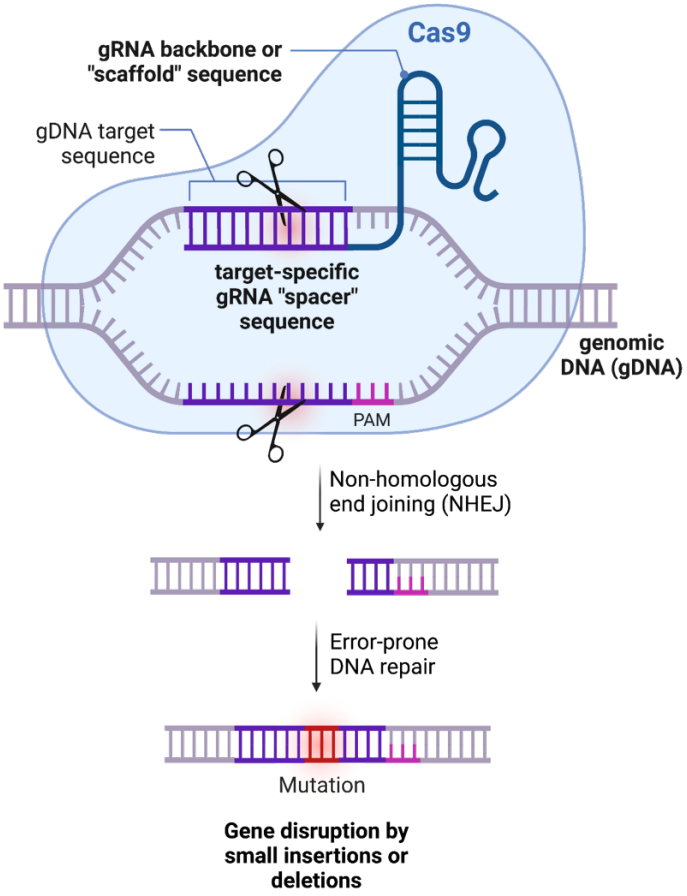


Figure 1.2: Overview of the Cas9-mediated CRISPR knockout method.

Adapted from "2020 Nobel Prize in Chemistry: A Tool for Genome Editing (CRISPR-Cas9)", by BioRender.com (2023). Retrieved from https://app.biorender.com/biorender-templates.

The rules for designing SpCas9 gRNAs with high on-target binding efficiency and low off-target binding have also been elucidated (Doench, 2018; Hanna & Doench, 2020). Researchers have designed pooled libraries of effective gRNAs to target each gene in the human genome (Doench et al., 2016; Gonçalves et al., 2021; Hart et al., 2017; Mair et al., 2019; Morgens et al., 2017; Sanjana et al., 2014; Sanson et al., 2018). These genome-wide gRNA libraries are easily accessible (https://www.addgene.org/crispr/libraries/), and there are many web-based tools for gRNA design (https://www.addgene.org/crispr/reference/#grna). These resources enable the

application of Cas9-mediated CRISPR knockout screening to look for synthetic lethal targets across cancer subtypes, tumor genetic backgrounds, and environmental conditions.

1.3.1.1 Single-gene CRISPR screens for synthetic lethal target identification

An additional approach for identifying synthetic lethal targets is the use of single-gene CRISPR-KO screens across multiple cancer genetic backgrounds. In cases where one gene has been inactivated, such as through a loss-of-function mutation or copy number alteration resulting in gene deletion, single-gene CRISPR approaches can be used to identify synthetic lethal interactions with the mutant gene. Publicly-available loss-of-function screening datasets include the Broad Institute's Cancer Dependency Map (Meyers et al., 2017; Pacini et al., 2021; Tsherniak et al., 2017), which contains results from whole-genome single-gene CRISPR-KO screens in hundreds of cancer cell lines. The Cancer Cell Line Encyclopedia has profiled the same cancer cell lines using omics approaches such as whole exome sequencing, RNA sequencing, and copy number alteration profiling (Barretina et al., 2012; Ghandi et al., 2019). The Wellcome Trust Sanger Institute has also generated similar datasets via loss-of-function screening and omics profiling in cancer cell lines (Behan et al., 2019; Tate et al., 2019; van der Meer et al., 2019). Together, these datasets can be leveraged for synthetic lethal target identification.

On the scale of individual cancer driver mutations, single-gene CRISPR screens have been used to identify potential synthetic lethal targets. Some examples include the gene family *ME1/ME2/ME3* in pancreatic ductal adenocarcinoma (Dey et al., 2017), *VPS4A/VPS4B* in colorectal cancer (Behan et al., 2019; Neggers et al., 2020; Szymańska et al., 2020), and multiple synthetic lethal targets in Ras-mutant acute myeloid leukemia (Wang et al., 2017). However, this type of approach is inefficient for genome-scale synthetic lethal target discovery because the

identification of all possible synthetic lethal cancer targets would require executing whole-genome single-gene CRISPR screens in every possible cancer subtype and genetic background.

Multiple research groups have therefore leveraged these large-scale cancer dependency datasets to computationally predict synthetic lethal interactions in cancer cells (Dandage & Landry, 2019; Lord et al., 2020). These approaches revealed a low synthetic lethal hit rate between unrelated genes, with the most comprehensive analysis finding a synthetic lethal hit rate of < 0.2% (Lord et al., 2020). To address the issue of a low synthetic lethal hit rate, one group turned to predicting synthetic lethal interactions among duplicated genes, which resulted in a higher synthetic lethal hit rate (De Kegel et al., 2021; De Kegel & Ryan, 2019). However, these computational approaches still have other drawbacks including a bias towards identifying interactions with commonly-mutated cancer genes such as *TP53* as well as reliance on computational predictions of variant impact and/or sequencing-based approaches for identifying copy number alterations, which may be relatively inaccurate.

Moreover, as mentioned previously (**Section 1.2**) the human genome exhibits a high degree of redundancy (Dandage & Landry, 2019; Dean et al., 2008; Harrison et al., 2007; Lavi, 2015; Ohno, 2013). This redundancy thus biases the outcome of all single-gene screens (Dede, Kim, et al., 2020; Harrison et al., 2007), preventing drug target identification. One of the first single-gene CRISPR-Cas9 screening papers noted that duplicated genes were underrepresented among hits: "we … observed that genes with paralogs are indeed less likely to be essential, which is consistent with the idea that paralogs can provide functional redundancy at the cellular level (Fig. 2J)" (Wang et al., 2015). Thus, multi-targeting methods to directly identify synthetic lethal cancer targets have been developed.

1.3.1.2 Synthetic lethal target discovery via multi-targeting screens

Multi-targeting, Cas9-mediated CRISPR knockout methods enable the direct experimental identification of synthetic lethal interactions in human cancer cells. One of the first papers describing a method for whole-genome single-gene CRISPR-Cas9 screening in human cells incorporated multi-gRNA arrays to perform combinatorial genome editing (Cong et al., 2013). Given the high on-target specificity of optimized gRNAs, many research groups recognized the possibility of using CRISPR approaches to target and inactivate two or more genes simultaneously. The first wave of these Cas9-mediated multi-targeting CRISPR approaches focused mainly on targeting genes known to be cancer drivers, tumor suppressors, and/or cancer drug targets (Han et al., 2017; Shen et al., 2017; Wong et al., 2016). Though these approaches effectively induced dual-gene knockouts, they had low hit rates overall: methods tested in a single cell line yielded a < 0.3% synthetic lethal hit rate (Han et al., 2017; Wong et al., 2016) while another method showed a < 3% hit rate across three cell lines (Shen et al., 2017). Moreover, many of the interactions uncovered were previously known and given that these approaches employed multiple SpCas9-type gRNAs in one vector, researchers noted relatively high rates of recombination between gRNAs due the presence of two identical scaffold sequences and/or two identical promoters.

To address the gRNA recombination issue, one group developed a multi-targeting CRISPR method that used both SpCas9 and *S. aureus* Cas9 to achieve more specific dual knockouts and applied it to map interactions among gene pairs that were known or computationally predicted to be synthetic lethal (Najm et al., 2018). Another group developed a dual-targeting approach using both SpCas9 and Cas12a to perform dual-gene knockouts of known cancer genes as well as duplicated genes (Gonatopoulos-Pournatzis et al., 2020). However, these methods require the expression of multiple Cas enzymes per cell and are therefore unwieldy relative to systems that use only SpCas9.

An optimized SpCas9-only dual-targeting CRISPR method mainly focused on technology development and failed to uncover many new synthetic lethal interactions between human genes (Sanson et al., 2018). To increase the synthetic lethal hit rate, other groups, including ours, developed and applied dual-targeting Cas9-mediated screening approaches that target only paralogs, resulting in an increased synthetic lethal hit rate of 12-15% (Ito et al., 2021; Parrish et al., 2021; Thompson et al., 2021). These methods incorporate the use of different promoters to reduce recombination between dual gRNAs.

The major benefit of using Cas9-only approaches for synthetic lethal target identification is that SpCas9-based gRNA design rules have been extensively studied (DeWeirdt et al., 2022; Doench, 2018; Doench et al., 2016; Hanna & Doench, 2020; Sanson et al., 2018) and tool-builders or users can thus choose validated gRNAs from existing genome-wide libraries for effective dual-gene inactivation. Additionally, as a field we have a better understanding of the effectiveness of Cas9-based gene inactivation relative to epigenetic silencing of genes or less well-studied Cas enzymes.

The Cas9-mediated CRISPR knockout screening approach has drawbacks including off-target toxicity induced by DSBs and the generation of in-frame indels after CRISPR-mediated DSBs trigger non-homologous end joining. Both these phenomena can mask the true phenotypic effects of a gene knockout (Doench, 2018; Horlbeck et al., 2018). When screening for essential genes, off-target toxicity can be a concern since it may mimic the effects of an essential gene knockout.

The major drawback to using Cas9-mediated CRISPR knockout for multi-targeting screens is the cloning required when generating new CRISPR libraries. There are multiple reasons for this: first, the two SpCas9 gRNAs must be driven off of independent promoters and commonly-used

promoters like human U6 are 200-300 base pairs in length. This makes it expensive to synthesize the entire construct (gRNA1-promoter-gRNA2) needed for a one-step cloning procedure. Additionally, the use of two SpCas9-type gRNA scaffolds may interfere with DNA synthesis methods. Thus, Cas9-only multi-targeting CRISPR methods require a two-step, large-scale cloning procedure to enable the incorporation of both gRNAs and both promoters into the plasmid vector. The use of Cas ortholog Cas12a (**Section 1.3.3**) eliminates the need for this type of laborious cloning procedure.

1.3.2 *CRISPR interference*

1.3.2.1 Background on CRISPRi loss-of-function screens

The CRISPR method has also been adapted for other purposes. One application is the development of CRISPR interference (CRISPRi), in which a catalytically dead SpCas9 enzyme is fused to a transcriptional repressor such as KRAB. A major benefit of CRISPRi relative to CRISPR-KO approaches for identifying essential gene pairs is that CRISPRi does not induce DSBs which can cause off-target toxicity (Horlbeck et al., 2018; Wang et al., 2015). Additionally, since CRISPR-KO can induce in-frame indels, there may be cells present in the polyclonal population that still express the target genes, thereby masking the phenotypic effect of successful knockouts (Doench, 2018; Horlbeck et al., 2018). For this reason, CRISPRi may be preferred for studying essential genes.

However, CRISPRi approaches also have quite a few drawbacks. The biggest limitation of this approach is the variability of transcriptional repression induced. Preexisting chromatin states or epigenetic modifications in the cell can prevent CRISPRi from efficiently silencing gene expression. Additionally, CRISPRi silencing can sometimes extend to genes located near the target in higher-order (or 3D) genomic space (Larson et al., 2013). Moreover, gRNAs for CRISPRi must

be designed in a narrow range of DNA around the transcription start site of a gene (Doench, 2018), so users have fewer options when designing multi-targeting gRNA libraries. Finally, our understanding of how to design optimal gRNAs for CRISPRi is limited relative to the rules for CRISPR-KO gRNA design (Doench, 2018; Larson et al., 2013), which have been studied much more extensively.

1.3.2.2 Multi-targeting applications

One group developed a dual-targeting loss-of-function CRISPRi screening method for genetic interaction screening among known essential gene pairs. However, this approach resulted in a synthetic lethal hit rate of less than 3% (Horlbeck et al., 2018). This is likely because the gene pairs were mostly unrelated as opposed to related paralogs. Additionally, because the authors used an all-by-all approach for dual inactivation of their target gene pairs, they were only able to target 472 total genes despite screening over 200,000 gene pairs.

1.3.3 *Cas12a-mediated multi-gene inactivation*

Since the discovery of the CRISPR-Cas9 system, other groups have identified Cas endonucleases from other bacterial species and engineered these enzymes for new applications. One example, Cas12a, is of particular interest for synthetic lethal target discovery due to its ability to cleave arrays of multiple gRNAs in addition to generating DSBs in target DNA. Cas12a, formerly known as Cpf1, was discovered in *Acidaminococcus* (Zetsche et al., 2015) and has since been engineered for improved genome editing (Kleinstiver et al., 2019) and applied for multi-targeting screens to generate efficient dual and triple knockouts (DeWeirdt et al., 2020). Other groups have applied Cas12a-based approaches to identify cancer-relevant paralog synthetic lethal interactions (Anvar et al., 2023; Dede, McLaughlin, et al., 2020).

The major benefit of using Cas12a versus Cas9 for dual-targeting screens is eliminating the need for multi-step cloning procedures (DeWeirdt et al., 2020), though a large-scale, one-step cloning procedure is still required. For pairwise genetic interaction screening, the benefits of a one- versus a two-step cloning procedure are negligible. However, the real promise of Cas12a lies in its ability to target more than two genes at once, enabling higher-order genetic interaction mapping in human cells.

However, Cas12a-mediated CRISPR knockout screening has limitations as well. Since Cas12a is less well-studied than Cas9, the rules for gRNA design are not yet well understood. Cas12a is thus less useful for synthetic lethal screening applications due to lower on-target knockout efficiency (Zetsche et al., 2017) and lack of validated genome-wide gRNA libraries. Additionally, since Cas9-mediated CRISPR-KO is currently the most widely-used genetic screening approach, this limits the utility of Cas12a approaches for most users, who may not have Cas12a reagents on hand.

1.4 WHERE DO WE GO FROM HERE?

The discovery of new synthetic lethal cancer targets is an important step forward in developing cancer therapies that are better at killing cancer cells while improving patient quality-of-life. Current therapeutic approaches such as chemotherapy and radiation are not tumor-specific and thus can also harm healthy cells. Chemotherapy in particular can cause severe toxicity for cancer patients, since there is a very narrow therapeutic index at which drugs designed to kill all fast-growing cells are more toxic to tumor cells than normal cells. Radiation therapy is a localized approach and thus is less effective in treating metastatic cancer or tumors affecting internal organs than systemic, drug-based therapies.

Targeted approaches like the use of oncogene-targeted drugs or synthetic lethal therapies leverage our improved knowledge of cancer genomics to help slow tumor growth or even cure cancer. Targeted therapies have been successfully applied to treat many types of cancer and have extended the lifespans of non-small cell lung cancer patients while reducing off-target toxicity. Synthetic lethal approaches, meanwhile, are even more tumor-specific and PARP inhibitors have shown clinical success in treating breast, ovarian, and prostate tumors with biallelic inactivation of *BRCA1* or *BRCA2*.

Thanks to advances in genome editing technologies for large-scale, multi-targeting loss-of-function screens in human cells, we can now begin to identify additional synthetic lethal targets so the benefits of this therapeutic approach can be extended to more cancer patients. It makes sense to start applying dual-targeting screens in highly-mutated cancer types, which are most likely to contain inactivating mutations in targetable synthetic lethal genes. Examples of highly-mutated cancer subtypes that have been successfully treated via targeted and/or synthetic lethal therapies include lung adenocarcinoma (Yoda et al., 2019; Yoshida et al., 2020) and cervical cancer (Chalmers et al., 2017; Lord & Ashworth, 2017).

Yet given the frequency of acquired resistance to targeted therapies in lung cancer and synthetic lethal therapy in cervical cancer, additional low-toxicity therapeutic options are urgently needed to extend patient lifespans (Ashworth & Lord, 2018; Dagogo-Jack & Shaw, 2018; Yoda et al., 2019). Moreover, patients with targetable mutations remain in the minority for most cancer types, so new targetable mutations must also be uncovered. Synthetic lethal therapies provide an attractive option for greatly increasing the range of potential targetable mutations while limiting the toxicity that often occurs with chemotherapy and oncogene-targeted therapy.

The challenge of identifying new synthetic lethal cancer targets has motivated my research throughout my PhD. In Chapter 2, I share results from a method I developed to execute dual-targeting, Cas9-mediated CRISPR loss-of-function screens in lung and cervical cancer cells. Chapter 3 describes the software package I developed to enable other researchers to apply my genetic interaction mapping method to identify synthetic lethal targets in any cancer type. In Chapter 4, I share my work identifying gene families that can be targeted by existing cancer drugs to prioritize synthetic lethal targets for follow-up study. Altogether, my thesis work has uncovered hundreds of new synthetic lethal cancer targets that can be further tested and translated to the clinic as effective, low-toxicity cancer treatments.

Chapter 2. DISCOVERY OF SYNTHETIC LETHAL AND TUMOR SUPPRESSIVE PARALOG PAIRS IN THE HUMAN GENOME

A version of this chapter has been published:

Parrish, P. C. R., Thomas, J. D., Gabel, A. M., Kamlapurkar, S., Bradley, R. K., & Berger, A. H. (2021). Discovery of synthetic lethal and tumor suppressor paralog pairs in the human genome. *Cell Reports*, *36*(9). https://doi.org/10.1016/j.celrep.2021.109597

I led this work and my contributions to this paper included designing and conducting experiments, analyzing the data, and writing the manuscript.

CRISPR screens have accelerated the discovery of important cancer vulnerabilities. However, single gene knockout phenotypes can be masked by redundancy among related genes. Paralogs constitute two-thirds of the human protein-coding genome, so existing methods are likely inadequate for assaying a large portion of gene function. Here we develop paired guide RNAs for Paralog gENetic interaction mapping (pgPEN), a pooled CRISPR-Cas9 single and double knockout approach targeting over 2,000 human paralogs. We apply pgPEN to two cell types and discover that 12% of human paralogs exhibit synthetic lethality in at least one context. We recover known synthetic lethal paralogs *MEK1/MEK2*, important drug targets *CDK4/CDK6*, and other synthetic lethal pairs including *CCNL1/CCNL2*. Additionally, we identify ten tumor suppressor paralog pairs whose compound loss promotes cell proliferation. These findings nominate drug targets and suggest that paralog genetic interactions could shape the landscape of positive and negative selection in cancer.

2.1 INTRODUCTION

CRISPR-Cas9 technology has revolutionized functional genomics by enabling high-fidelity, genome-scale, multiplexed loss-of-function screens in human cells. Due to high specificity and ease of application, genome-wide CRISPR screens are increasingly used to identify cancer drug targets and determine mechanisms of drug resistance (Bartha et al., 2018; Blomen et al., 2015; Hart et al., 2015; Tsherniak et al., 2017; Wang et al., 2015, 2017). However, single-gene knockout studies have a major blind spot: they are unable to assay the function of paralogs — ancestrally duplicated genes that frequently retain at least partially overlapping functions. The human genome exhibits a high degree of redundancy as a result of diploidy, gene duplication, and functional overlap of metabolic and signaling pathways (Dean et al., 2008; Harrison et al., 2007; Lavi, 2015; Ohno, 2013). Remarkably, paralogs constitute two-thirds of the human genome, making this blind spot the rule, not an exception, and paralogous genes are less likely to be essential for cell growth than non-paralogous ("singleton") genes in CRISPR knockout screens (Wang et al., 2015). This paralog blind spot therefore obscures our understanding of normal human genome function and impedes the identification of new cancer drug targets.

Genetic interactions between paralogs have been extensively characterized in yeast, revealing fundamental insights about the differences between whole-genome and small-scale duplicates, functional groups that are enriched for interacting paralogs, and paralog mRNA expression patterns (Dean et al., 2008; Diss et al., 2017; Guan et al., 2007; Harrison et al., 2007). Essential paralogs that compensate for one another's function exhibit "synthetic lethality," a genetic interaction in which elimination of the entire family is deleterious but individual loss is tolerated. Yeast geneticists have defined quantitative measures of genetic interactions, which can capture both positive (buffering) and negative (synthetic lethal) interactions (Collins et al., 2006).

While paralog genetic interactions are still poorly characterized in mammalian cells, the extensive degree of duplication in the human genome is similar to that seen in yeast (Dennis & Eichler, 2016; Lan & Pritchard, 2016; Singh et al., 2012), so experimental evaluation of human cells is likely to also reveal complex genetic interactions.

Querying the genetic interaction space of the human genome has been limited by current technology; to survey even every possible pairwise interaction, let alone higher order interactions, would involve ~200 million unique genetic perturbations. Moreover, the landscape of genetic interaction among randomly selected genes is exceedingly sparse; existing studies of much smaller sets of gene pairs in human cells identified genetic interactions in less than 0.1% of unrelated gene pairs (Han et al., 2017). To proactively identify these rare but functionally important interactions, research should therefore focus on high-value sets of genes likely to be enriched for functional interactions, such as paralogs.

Interestingly, the same duplication that makes paralogs difficult to study provides a tactical advantage for cancer therapy: the highly rearranged genomes typical of cancer often harbor paralog deletions and inactivating mutations. Cancer-associated loss-of-function of one paralog can confer a dependency on the continued activity of a duplicated pair (De Kegel & Ryan, 2019; Lord et al., 2020; Viswanathan et al., 2018) and this phenomenon has been used to identify synthetic lethal relationships of paralogs such as *MAGOH/MAGOHB* (Viswanathan et al., 2018), *ARID1A/ARID1B* (Helming et al., 2014), and *SMARCA2/SMARCA4* (Hoffman et al., 2014). If the remaining actively expressed paralog could be targeted in tumors with loss of its pair, then tumor cells may show a selective therapeutic window compared to the surrounding normal cells with expression of both paralog members. A successful example of a therapy based on a synthetic lethal interaction is the enhanced sensitivity to PARP inhibitors in *BRCA1*- and *BRCA2*-mutant tumors

(Bryant et al., 2005; Farmer et al., 2005; Lord & Ashworth, 2017). We hypothesized that paralogs could provide a rich source of genetic interactions and that direct experimental identification of synthetic lethal paralogs could therefore enable future drug discovery efforts.

Recently, several groups have developed innovative methods for assessing human genetic interactions at scale (Boettcher et al., 2018; Dede, McLaughlin, et al., 2020; DeWeirdt et al., 2020; Gier et al., 2020; Gonatopoulos-Pournatzis et al., 2020; Han et al., 2017; Horlbeck et al., 2018; Najm et al., 2018; Shen et al., 2017). Consistently, while the overall rate of genetic interaction among gene pairs is low, many of the interactions identified were in paralogous genes. To comprehensively identify genetic interactions between human paralogs, we here report our direct experimental evaluation of genetic interactions among 1,030 paralog pairs (2,060 genes) in two human cell contexts. Our analysis revealed not only an extraordinarily high rate of paralog synthetic lethality, but also identified positive interactions that nominate ten paralog pairs as tumor suppressor gene families.

2.2 RESULTS

2.2.1 *A paralog blind spot limits discovery of essential genes and cancer dependencies*

The human genome is highly duplicated, with paralogous genes constituting over two thirds of protein coding genes (**Figure 2.1A**). Like other groups (Dandage & Landry, 2019; Dede, McLaughlin, et al., 2020; Wang et al., 2015), we noticed that paralogous genes are less likely to be essential for cell growth than non-paralogous "singleton" genes in single-gene CRISPR knockout screening data ($p < 2.20e{-}16$ by one-tailed Kolmogorov-Smirnov [K-S] test, **Figure 2.1B**) (Vichas et al., 2020). Given the utility of targeting cancer-essential genes for therapy, we reasoned that this paralog blind spot may prevent detection of important druggable cancer dependencies.

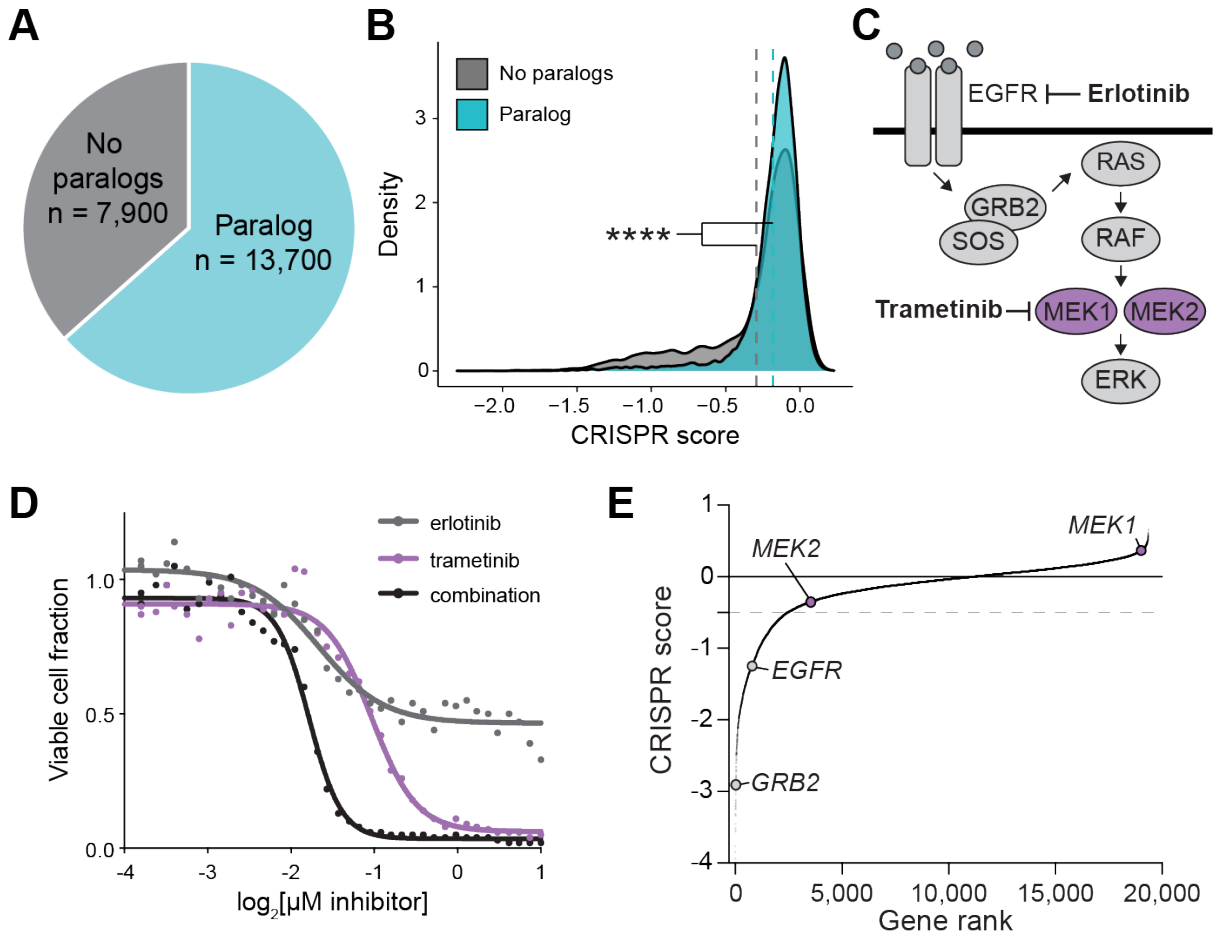


Figure 2.1: Paralog dependencies are missed in single-gene CRISPR knockout screens.

A) Pie chart of human genes classified based on whether they are part of a paralog gene family with 10-99% amino acid sequence identity.

B) Density plot of CRISPR scores for a single gene CRISPR knockout screen in PC9 lung adenocarcinoma cells. Data is from (Vichas et al., 2020). Dashed lines indicate the mean CRISPR score of genes in each group. $P < 2.20e{-}16$ by one-tailed K-S test.

C) Schematic of the EGFR/Ras/MAPK signaling pathway.

D) Dose response curve of PC9-Cas9-EGFRT790M/L858R lung adenocarcinoma cells treated with erlotinib, trametinib, or a 1:1 combination of both drugs. The fraction of viable cells was determined by CellTiterGlo luminescence after 96 hours of treatment. Data was re-analyzed from a larger drug screen from (A. H. Berger et al., 2016).

E) Rank plot of CRISPR scores from an erlotinib sensitization screen in PC9-Cas9-EGFRT790M/L858R cells (Vichas et al., 2020). MEK1 (alias MAP2K1) or MEK2 (alias MAP2K2) single gene knockout does not result in significantly decreased cell growth. Gray dashed line indicates the threshold for negative selection (–0.5).

To determine whether known therapeutic vulnerabilities are missed in CRISPR knockout screens, we compared our previous drug sensitivity profiling of PC9-EGFR$^{L858R/T790M}$ cells (A. H. Berger et al., 2016) to recent genetic vulnerabilities identified in the same system (Vichas et al., 2020). These cells exhibit resistance to the EGFR tyrosine kinase inhibitor, erlotinib, which can be reversed by treatment with trametinib, a kinase inhibitor of MEK1 and MEK2 — protein kinases encoded by the paralogous genes *MEK1* and *MEK2* (also known as *MAP2K1/MAP2K2*), which are part of the Ras/MAPK pathway (**Figure 2.1C-D**). The Ras/MAPK pathway is frequently activated in lung cancer by mutation of upstream receptor tyrosine kinases such as EGFR, or activation of KRAS or mutation of *MEK1* itself (Arcila et al., 2015; Sanchez-Vega et al., 2018; TCGA, 2014). We noted in single-gene CRISPR knockout data in the same cellular context that while *EGFR* and other Ras pathway members such as *GRB2* were essential as expected, neither *MEK1* nor *MEK2* was essential when knocked out individually (**Figure 2.1E**). We reasoned that paralog redundancy might underlie the apparent disconnect between the small molecule and genetic assays. We therefore sought to develop a multiplexed CRISPR approach to directly probe paralog compensation on a genome scale, enabling the discovery of many more paralogous drug targets that may be missed in current CRISPR-based target discovery efforts.

2.2.2 *The pgPEN library enables single and double knockout of 1,030 human paralog families*

To identify synthetic lethal paralogs that could serve as potential lung cancer drug targets, we focused on duplicated genes — paralog families of only two genes. We identified paralog families from Ensembl (Vilella et al., 2009) and then selected families in which a maximum of two genes shared 50-99% amino acid identity (**Figure 2.7A**). Next, we designed a paired guide RNA (pgRNA) CRISPR library to knock out each paralog alone or in combination with its

respective pair. Using single guide RNA (sgRNA) sequences from the Brunello CRISPR library (Doench et al., 2016), we designed 16 four-by-four pairwise double knockout (DKO) pgRNAs for each paralog pair. In addition, we designed single knockout (single KO) pgRNAs containing one targeting sgRNA paired with a non-targeting control sgRNA having no match to the human reference genome. This was done for both paralogs to generate a total of 16 single KO pgRNAs. 500 double non-targeting pgRNAs were included as a control. This "paired guide RNAs for Paralog gENetic interaction mapping (pgPEN)" library (**Table S1**) was synthesized and cloned at 1000-fold coverage using previously-developed methods (Gasperini et al., 2017; Thomas et al., 2020). Next-generation sequencing confirmed that >99.99% of pgRNAs were present in the cloned plasmid pool. The final pgPEN library consists of 33,170 pgRNAs targeting 1,030 paralog pairs (2,060 genes) in single knockout and double knockout combinations. Over half of the paralogs in the pgPEN library are unique to this study while the remainder were also assayed in recent genetic interaction maps (Dede, McLaughlin, et al., 2020; Gonatopoulos-Pournatzis et al., 2020; Thompson et al., 2021) (**Figure 2.7B**). 554 of the gene products of pgPEN-targeted genes are considered "druggable" by recent criteria (Finan et al., 2017) (**Figure 2.7C**).

To map genetic interactions between paralogs, we applied the pgPEN library to PC9 lung adenocarcinoma cells previously engineered to constitutively express Cas9 (Thomas et al., 2020; Vichas et al., 2020) using standard pooled CRISPR screening methodology in triplicate (**Figure 2.2A**). pgRNAs that were positively or negatively selected were identified by Illumina sequencing of pgRNA abundance after ~12 population doublings *in vitro* compared to the starting abundance in the plasmid pool (**Figure 2.2A**). The sequencing strategy used for the pgPEN method is outlined in **Figure 2.7D** and pgRNA library coverage and gRNA pairing statistics are included in **Table S2**. Only properly paired pgRNAs were included in our CRISPR screen analysis. Plasmid pgRNA

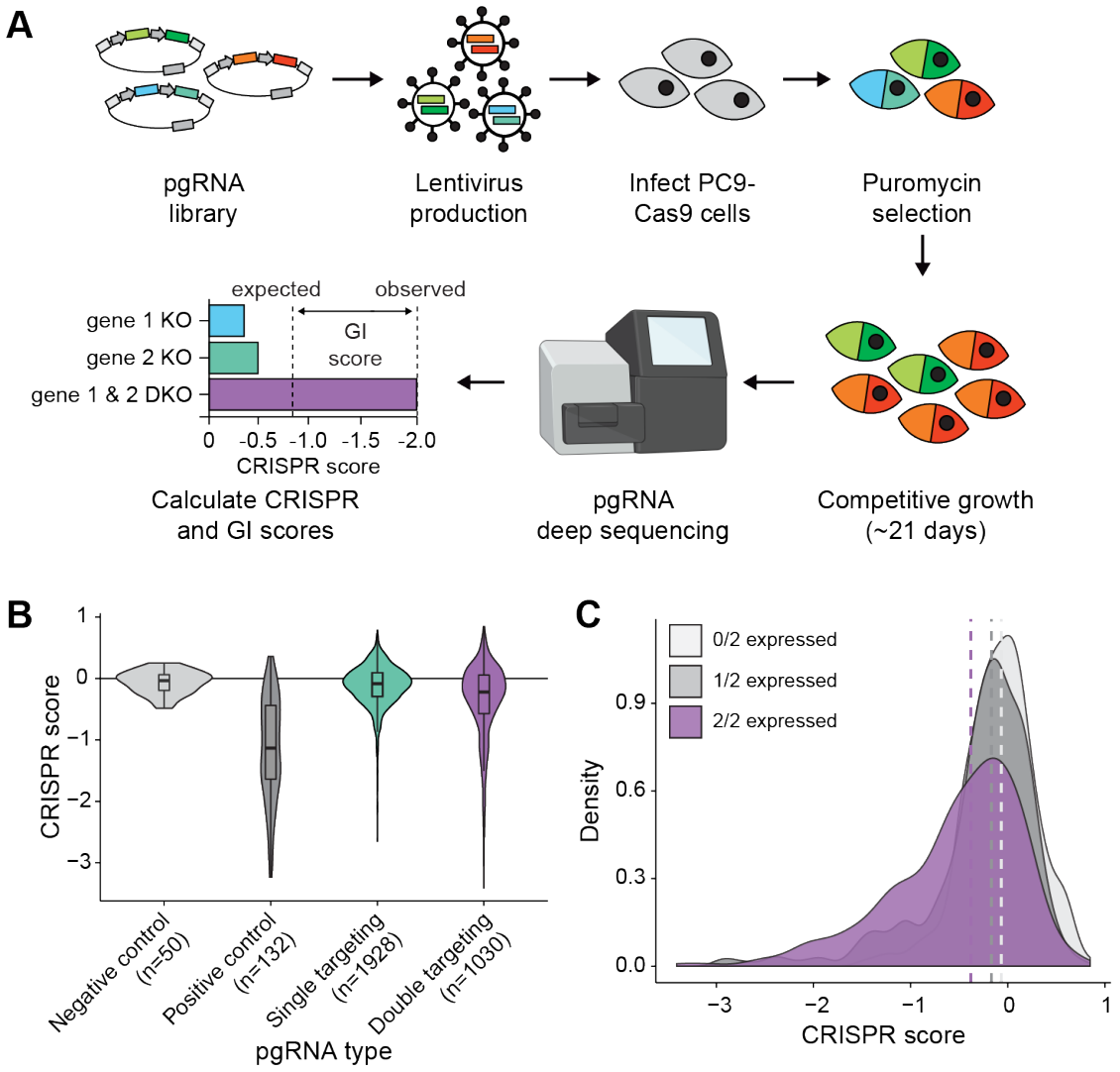


Figure 2.2: The pgPEN CRISPR library enables genetic interaction mapping of 1,030 human paralog pairs.

A) Schematic of pgPEN screening approach for paralog genetic interaction mapping.

B) Violin plots of target-level CRISPR scores for negative control (double non-targeting control), positive control (single KO pgRNAs targeting known essential genes), all other single KO pgRNAs, and DKO pgRNAs in the PC9 screen. The double targeting pgRNA group had significantly lower CRISPR scores than the single targeting pgRNA group ($p < 2.20e{-}16$ by one-tailed K-S test).

C) Density plot of target-level CRISPR scores for DKO pgRNAs grouped by whether zero, one, or both targeted genes are expressed (TPM > 2) in PC9 cells. Dashed lines indicate the median CRISPR score for each group. pgRNAs targeting expressed genes had significantly lower CRISPR scores than those targeting two unexpressed genes for both the 2/2 genes expressed ($p < 2.20e{-}16$ by one-tailed K-S test) and 1/2 genes expressed ($p = 4.03e{-}03$ by one-tailed K-S test) groups.

abundance was highly correlated with early time point samples taken immediately following lentiviral transduction and puromycin selection (mean Pearson's r = 0.93; **Figure 2.8A**). End point samples exhibited expected changes in pgRNA abundance (**Figure 2.8B**) that were highly correlated across replicates (mean Pearson's r = 0.82; **Figure 2.8C**). Single KO pgRNAs targeting pan-essential genes (Meyers et al., 2017) showed the expected dropout in late time point samples (**Figure 2.2B**). These data indicate that the screen was performed without significant bottlenecking and that the pgRNAs performed as expected for known essential genes. Similar to previously established CRISPR screen analysis methods (Meyers et al., 2017), we generated normalized CRISPR scores by scaling pgRNA $\log_2$(fold change) values such that the median CRISPR score of double non-targeting constructs was zero and the median CRISPR score of pan-essential single KO constructs was –1.

CRISPR-Cas9 gene knockout involves the generation of double strand breaks that can themselves inhibit cell proliferation rate (Aguirre et al., 2016). One concern in targeting multiple loci with Cas9 is that the increased generation of double strand breaks could, independent of any specific gene effect, result in enhanced negative selection of DKO compared to single KO pgRNAs. To control for this possibility, we further normalized data such that the median CRISPR score of all single KO pgRNAs targeting non-expressed genes would be zero and the median CRISPR score of all DKO pgRNAs targeting two non-expressed genes would be zero (Methods, **Figure 2.8D-E,** and **Figure 2.2C**). After this normalization, DKO constructs still had significantly lower CRISPR scores than single KO constructs ($p < 2.20e{-16}$ by one-tailed K-S test), indicative of possible genetic interactions in the DKO group. As expected, constructs targeting expressed genes had significantly lower CRISPR scores than those targeting unexpressed genes for both single-targeting ($p = 1.96e{-10}$ by one-tailed K-S test; **Figure 2.8E**) and double-targeting ($p <$

2.20e–16 by one-tailed K-S test; **Figure 2.2C**) constructs. After normalization, only a minimal effect of paralog copy number (**Figure 2.8F**) on CRISPR score was observed (**Figure 2.8G-H**). Scaled CRISPR scores for pgRNAs in the PC9 screen can be found in **Table S3**.

2.2.3 *Direct identification of paralog genetic interactions in human lung cancer cells*

Using the PC9 CRISPR scores, we calculated genetic interaction (GI) scores for each paralog pair under a multiplicative model of genetic interaction following recently developed methods for human GI mapping (DeWeirdt et al., 2020; Han et al., 2017) (**Methods**). Comparison of the expected and observed CRISPR scores for each paralog pair enabled identification of interacting paralogs (**Figure 2.3A-B**) and calculation of GI scores for each paralog pair (**Figure 2.3C-D and Table S4**). This approach identified 87 synthetic lethal and 68 buffering genetic interactions among the 1,030 paralog pairs. Synthetic lethal interactions (GI < –0.5 and FDR < 0.1) included top pairs *CCNL1/CCNL2*, *CDK4/CDK6*, *GSK3A/GSK3B*, *G3BP1/G3BP2*, *CNOT7/CNOT8*, and *OXSR1/STK39* (**Figure 2.3B-D**). Interestingly, *CCNL1/CCNL2* code for cyclins L1 and L2, which activate the paralogous proteins CDK11A/CDK11B. Active CDK11 is involved in regulating pre-mRNA splicing and may also play a role in cell cycle regulation (Loyer & Trembley, 2020; Loyer et al., 2008). We found *CDK11A/CDK11B* were synthetic lethal in PC9 cells (**Figure 2.3B**), and recent work has shown that CDK11A/CDK11B are targeted by the small molecule inhibitor OTS964 (A. Lin et al., 2019). *OXSR1* and *STK39* encode evolutionarily conserved kinases involved in the oxidative stress response, and *STK39* has a possible role in promoting apoptosis (Balatoni et al., 2009; Gagnon & Delpire, 2012). We also found a significant synthetic lethal interaction for *MEK1/MEK2* (**Figure 2.3B-D**), confirming that the discrepancy between genetic and drug data in Figure 1 was indeed due to paralog redundancy. Known synthetic lethal

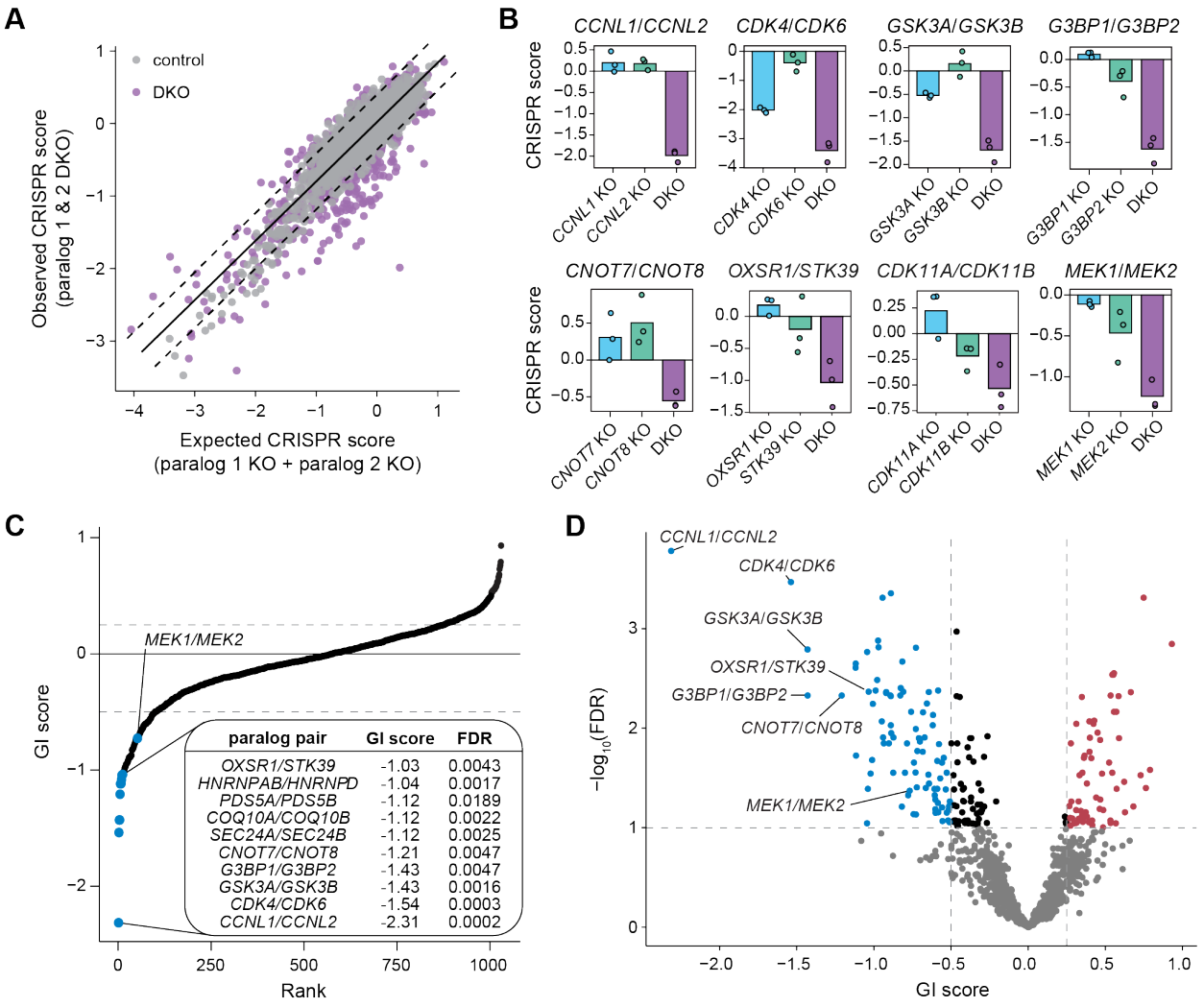


Figure 2.3: pgPEN uncovers synthetic lethal and buffering interactions.

A) Scatter plot of target-level observed versus expected CRISPR scores in the PC9 screen. The solid line is the linear regression line for the negative control (single KO) pgRNAs, while dashed lines indicate ± 2 residuals.

B) CRISPR scores for representative synthetic lethal paralog pairs. Data shown is the mean CS for each single KO or DKO target across three biological replicates with replicate data shown in overlaid points.

C) Rank plot of target-level genetic interaction scores in PC9 cells. Table insert, top paralogs based on GI score.

D) Volcano plot of target-level genetic interaction scores in PC9 cells. FDR indicates the multiple hypothesis-adjusted p values from a two-tailed t test (Methods). Blue, synthetic lethal paralog genetic interactions with GI < –0.5 and FDR < 0.1; red, buffering paralog genetic interactions with GI > 0.25 and FDR < 0.1.

paralogs such as *ARID1A/ARID1B* (Helming et al., 2014) and *MAPK1/MAPK3* (Dede, McLaughlin, et al., 2020; DeWeirdt et al., 2020) were also identified (**Table S4**).

To experimentally validate these findings, we developed a competitive fitness assay in red (mCherry) and green (GFP) labeled PC9-Cas9 cells (**Figure 2.9A-D**). In designing the validation experiment for top synthetic lethal paralog pairs, we used safe-targeting gRNAs (Morgens et al., 2017) in place of non-targeting gRNAs to account for the growth effects observed by generating one versus two double-strand breaks. We transduced PC9-Cas9-GFP-NLS cells with a double safe-targeting pgRNA, while PC9-Cas9-mCherry-NLS cells were transduced with paralog-targeting pgRNAs designed to knock out the expression of each paralog individually or both paralogs together (**Table S5**). Double safe-targeting cells were pooled with paralog-targeting cells at a 1:1 ratio of GFP:mCherry cells.

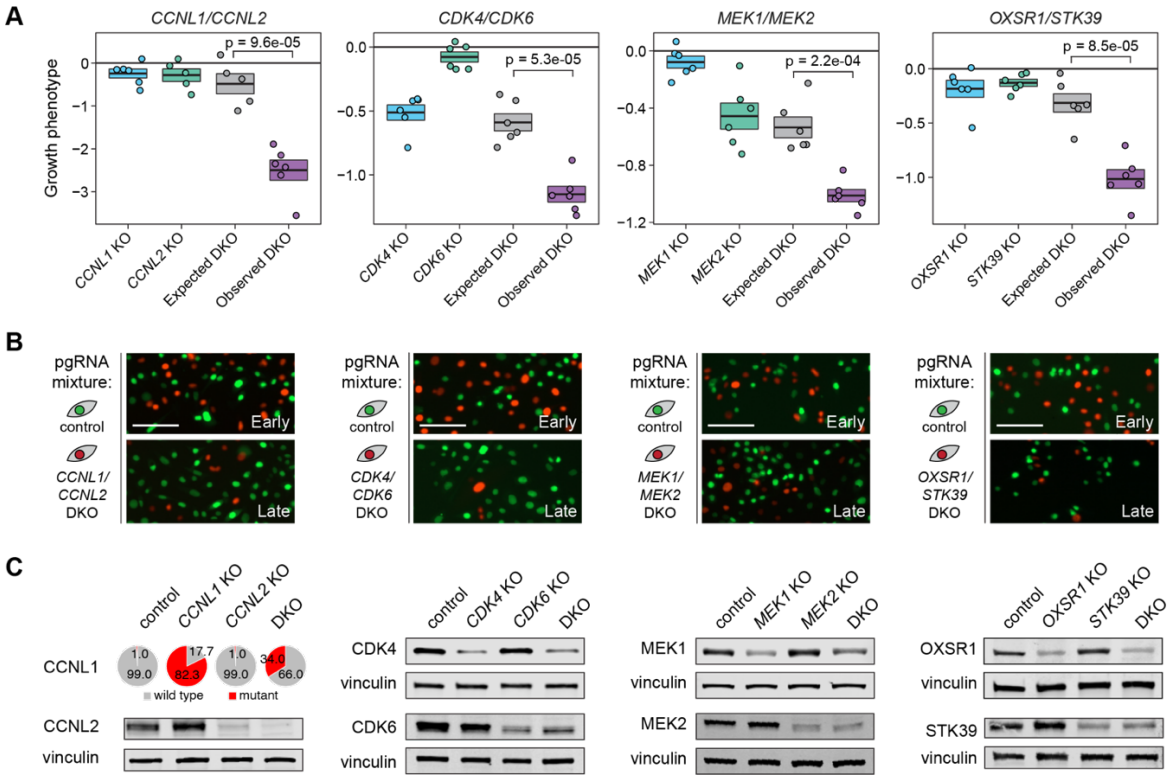


Figure 2.4: CRISPR validation experiments confirm top PC9 synthetic lethal interactions.

A) Box plots of growth phenotypes for PC9-Cas9-mCherry cells expressing the indicated pgRNA compared to PC9-Cas9-GFP cells expressing a double-safe-targeting control pgRNA. Boxes indicate mean ± SEM of six biological replicates, which are shown as overlaid points. Growth phenotype is defined as the $\log_2$ ratio of mCherry:GFP cell counts at the late time point compared to the Day 1 mCherry:GFP cell counts. Expected DKO phenotypes are the sum of single KO growth phenotypes. The expected and observed DKO phenotypes were compared using a one-tailed t test. Data shown are for the time point with the most extreme difference between expected and observed DKO growth phenotypes, termed the late time point: *CCNL1/CCNL2* (Day 12), *CDK4/CDK6* (Day 7), *MEK1/MEK2* (Day 11), and *OXSR1/STK39* (Day 10). Full time course data is shown in **Figure 2.9**.

B) Fluorescence microscopy images of competitive fitness assays on early (Day 1) and late time points as indicated above for panel A. Scale bar, 100 μM.

C) Western blot validation of single KO and DKO pgRNA-induced gene inactivation. For CCNL1, pie charts of percent mutant alleles based on next-generation sequencing are shown due to lack of a suitable CCNL1 antibody for western blotting. Additional genomic DNA-level validation data are presented in **Figure 2.10**.

Using this approach, we determined the effects of targeting four top synthetic lethal paralog pairs from the PC9 screen: *CCNL1/CCNL2*, *CDK4/CDK6*, *MEK1/MEK2*, and *OXSR1/STK39*. The results of these competitive fitness assays mirrored the gene knockout effects observed in the pooled screen format (**Figure 2.4A-B**). For *CCNL1/CCNL2* and *OXSR1/STK39*, individual gene knockouts showed little effect on cell growth, whereas combined knockout of both paralogs resulted in severe growth effects in both the screen (**Figure 2.3B**) and the competitive fitness assay (**Figure 2.4A**). The CRISPR screen data indicated that *CDK4* and *MEK2* single knockouts were essential on their own with the double knockout causing further negative growth effects for each pair (**Figure 2.3B**), and these effects were also observed in the bichromatic competitive fitness assays for these pairs (**Figure 2.4A**). We validated the synthetic lethality of these four pairs by confirming that the observed DKO growth phenotype was significantly less than an expected DKO growth phenotype, which was calculated based on the sum of the two single KO growth effects ($p = 9.56e{-}05$ for *CCNL1/CCNL2*, $p = 5.29e{-}05$ for *CDK4/CDK6*, $p = 2.20e{-}04$ for *MEK1/MEK2*,

p = 8.49e–05 for *OXSR1/STK39*, all by one-tailed t test). Line graphs of growth phenotypes across the entire competitive fitness assay are shown in **Figure 2.9E-H**. We experimentally validated two additional pairs with non-significant genetic interactions, *MAGOH/MAGOHB* and *PSMB5/PSMB8*, confirming a slightly negative but non-significant interaction for *MAGOH/MAGOHB* and that the *PSMB5/PSMB8* DKO growth phenotype was not significantly less than the expected, summed DKO growth phenotype (**Figure 2.9I-L**). Overall, there was good concordance between the genetic interactions observed in the pgPEN PC9 CRISPR screen and in the validation competitive growth assays (**Figure 2.9M**).

We also confirmed the successful generation of genomic DNA CRISPR edits at sgRNA target sites and the loss of target protein expression for the paralog pairs shown in **Figure 2.4A-B**. Confirmation of genomic DNA edits was done by next-generation sequencing for *CCNL1/CCNL2*, as we were unable to identify a suitable antibody for CCNL1 (**Figure 2.4C** and **Figure 2.10A-D**) and by Sanger sequencing for *CDK4/CDK6*, *MEK1/MEK2*, and *OXSR1/STK39* (**Figure 2.10E-I**). Western blots for CCNL2, CDK4/CDK6, MEK1/MEK2, and OXSR1/STK39 showed expected patterns of protein loss based on the pgRNAs expressed in each cell line (**Figure 2.4C**). Taken together, these data suggest that pgPEN is an effective strategy for uncovering synthetic lethal paralog interactions.

2.2.4 *A second pgPEN screen identifies shared versus cell line-specific paralog synthetic lethal interactions*

Next, we applied the pgPEN approach to a different tissue context, HeLa cervical carcinoma cells, using similar methodology with the exception of using a doxycycline-inducible Cas9 system (Cao et al., 2016). Quality control analyses of HeLa screening data again indicated successful generation of expected gene knockout phenotypes (**Figure 2.11A-J** and **Table S6**).

Calculation of GI scores identified 70 significant synthetic lethal interactions and 44 significant buffering interactions (**Figure 2.5A-B**, **Figure 2.11K**, and **Table S4**). Many of the top synthetic lethal pairs were shared between HeLa and PC9 cells, including *CCNL1/CCNL2*, *GSK3A/GSK3B*, and *MEK1/MEK2* (FDR < 0.1 in both cell lines, **Figure 2.5C-D**). Other paralog families were synthetic lethal only in one of the cell lines (FDR < 0.1 in PC9 or HeLa only). These included *SOS1/SOS2*, which were highly essential and synthetic lethal only in HeLa cells (**Figure 2.5C-D**) and *CDK4/CDK6*, which were only required in PC9 cells (**Figure 2.5C-D**). In total, 122 paralog pairs were identified as synthetic lethal in at least one context. Surprisingly, we noted that cell line-specific synthetic lethal interactions were often not explained by expression differences (**Figure 2.5D**), demonstrating that paralog dependencies, like other cancer dependencies, are modified by cellular context or other biological factors besides gene expression.

Some synthetic lethal paralog pairs, including *SEC24A/SEC24B*, *COQ10A/COQ10B*, *CNOT7/CNOT8*, *TIA/TIAL*, and *VPS4/VPS4B* have been highlighted in previous studies (Dede, McLaughlin, et al., 2020; Gonatopoulos-Pournatzis et al., 2020; Lord et al., 2020; Neggers et al., 2020; Szymańska et al., 2020). However, to our knowledge many of the synthetic lethal paralogs identified in the pgPEN screens were not previously known to be functionally redundant in human cells. These include *CCNL1/CCNL2* and *OXSR1/STK39* along with eukaryotic translation initiation factors *EIF1/EIF1B*, DNA and RNA helicase and cGAS/STING pathway members *G3BP1/G3BP2*, hexosamine biosynthesis pathway members *GFPT1/GFPT2*, and *PDS5A/PDS5B*, which regulate sister chromatid cohesion during mitosis. Individual members of many of these synthetic lethal paralog families have been previously implicated in cancer; for instance, high *GFPT2* expression has been linked to tumor metabolic reprogramming in lung adenocarcinoma

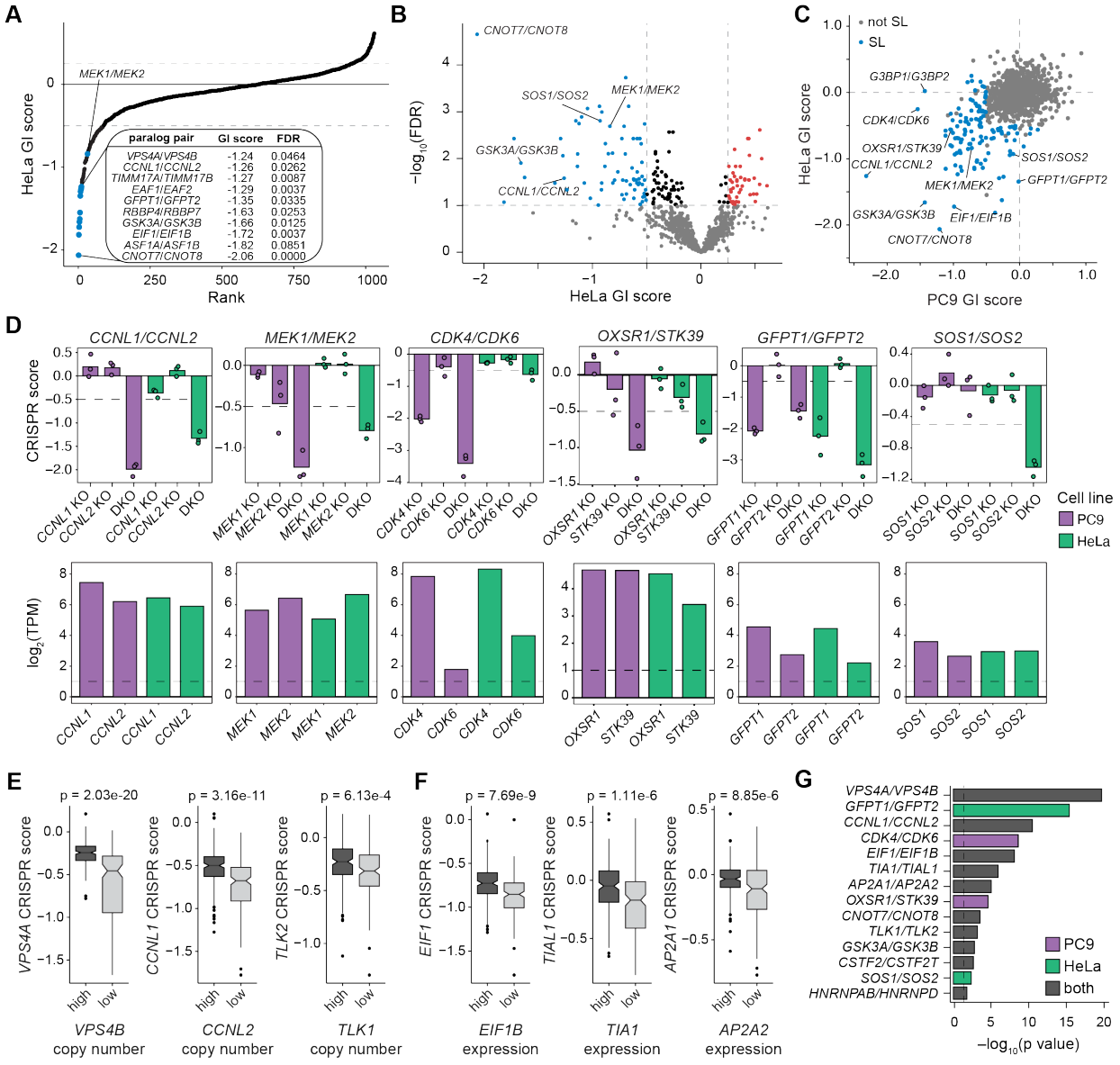


Figure 2.5: Identification of cell line-specific and shared synthetic lethal paralog pairs.

A) Rank plot of target-level genetic interaction scores in HeLa cells. Table insert, top paralogs based on GI score.

B) Volcano plot of target-level genetic interaction scores in HeLa cells. FDR indicates the multiple hypothesis-adjusted p values from a two-tailed t test (Methods). Blue, synthetic lethal paralog genetic interactions with GI < –0.5 and FDR < 0.1; red, buffering paralog genetic interactions with GI > 0.25 and FDR < 0.1.

C) Scatter plot of target-level genetic interaction scores for paralog pairs in PC9 versus HeLa cells. Blue, synthetic lethal paralog pairs with GI < –0.5 and FDR < 0.1 in either PC9 or HeLa cells; gray, all paralog pairs with GI ≥ –0.5 or FDR ≥ 0.1.

D) CRISPR scores for representative synthetic lethal paralog pairs identified in the PC9 and HeLa cell screens. Top row: Data shown is the mean CS for each single KO or

DKO target across three biological replicates with replicate data shown in overlaid points. Shared synthetic lethal paralogs (e.g. *CCNL1/CCNL2* and *MEK1/MEK2*) have FDR < 0.1 in both cell lines, PC9-specific paralogs (e.g. *CDK4/CDK6* and *OXSR1/STK39*) have FDR < 0.1 in PC9 only, and HeLa-specific paralogs (e.g. *GFTP1/GFPT2* and *SOS1/SOS2*) have FDR < 0.1 in HeLa only. Dashed lines indicate CS < -0.5. Bottom row: Paralog gene expression in PC9 and HeLa cells from RNA-seq analysis. Dashed lines indicate $\log_2$TPM = 1, the threshold for gene expression.

E) Box plots comparing the effect of CRISPR-mediated knockout of the indicated gene in DepMap cell lines with high (top quartile) compared to low (bottom quartile) copy number of its paralogous gene. For box plots, the middle line, hinges, notches, and whiskers indicate the median, 25^{th}/75^{th} percentiles, 95% confidence interval and data points within 1.5X the interquartile range (IQR) from the hinge. P values were computed using a two-tailed Wilcoxon Rank-Sum test. CRISPR score and copy number data was obtained from DepMap.

F) As in (F), but for gene expression.

G) Bar plot indicating the p values (computed using a two-tailed Wilcoxon Rank-Sum test) obtained by comparing the effect of a single paralog knockout to the copy number (as in panel (E) or gene expression (as in panel F) of its pair across human cancer cell lines profiled by DepMap. Bar color indicates whether each pair was synthetic lethal in PC9 only, HeLa only, or both cell lines in the pgPEN screens. Dashed line indicates p = 0.05.

(Zhang et al., 2018) and *PDS5B* is a negative regulator of cell proliferation and has been highlighted as a possible tumor suppressor gene in prostate cancer (Maffini et al., 2008).

As a complementary approach to validate our PC9 and HeLa screens, we used single gene knockout data to determine the essentiality of one paralog in the context of low expression or spontaneous copy number loss of its pair in hundreds of cancer cell lines profiled by DepMap (Tsherniak et al., 2017). For this analysis, we grouped cell lines according to whether "gene 1" of a paralog pair was highly (top quartile) or lowly (bottom quartile) expressed and calculated the median CRISPR score for "gene 2" in these groups (**Table S7**). This strategy was also used to determine the influence of gene copy number, and the inverse analysis (i.e., "gene 1" dependency in the context of low "gene 2" expression/copy number) was also performed. We found that for *VPS4A*/*VPS4B*, *CCNL1*/*CCNL2*, and *TLK1*/*TLK2*, reduced copy number of one family member

was significantly associated with greater dependency of the paralogous genes (**Figure 2.5E**). We observed similar dependencies when considering gene expression for *EIF1*/*EIF1B*, *TIA1*/*TIAL1*, *AP2A1*/*AP2A2* (**Figure 2.5F**). All of the top 10 shared synthetic lethal paralogs identified in both PC9 and HeLa screens displayed evidence of genetic interactions using these approaches (**Figure 2.5G** and **Table S7**).

2.2.5 *Identification of tumor suppressor paralog pairs*

In addition to synthetic lethal interactions, pgPEN screens can identify positive genetic interactions. We noticed that these positive interactions include both buffering interactions, where loss of one paralog prevents the deleterious phenotype of loss of the other — we identified 108 of such interactions in at least one cell context — as well as cases where the combined loss of both genes synergistically promotes cell growth. The latter are likely to be paralog families with tumor suppressive functions that require complete loss of the family to reveal the cellular phenotype. To identify these tumor suppressive paralogs, we restricted our analysis to significant buffering interactions (GI > 0.25, FDR < 0.1) between expressed paralogs in which the double knockout was positively enriched in each CRISPR screen (CS > 0.25). Under these relatively stringent criteria, four tumor suppressor interactions were identified in PC9, and six in HeLa cells (**Figure 2.6A**). None of the ten interactions were shared across cell lines, potentially reflecting the differing biology of HeLa and PC9 cells and the difficulty in achieving positive selection in basal culture conditions of rapidly proliferating cancer cell lines.

Tumor suppressor pairs identified in PC9 cells include *RAB27A/RAB27B*, encoding Rab-family GTPases involved in vesicle trafficking (Z. Li et al., 2018), and the BTB/POZ-domain genes *BTBD10/KCTD20* (**Figure 2.6B**). In HeLa cells, one of the top pairs identified was *CDKN2A/CDK2NB* (**Figure 2.6C**), frequently deleted tumor suppressors that encode the CDK4/6

inhibitors INK4A/ARF and INK4B (Kim & Sharpless, 2006). Another top tumor suppressor paralog pair in HeLa cells was *FBXO25/FBXO32* (**Figure 2.6C**), which encode SCF-type E3 ligase proteins. Although little is known about the function of these proteins and their substrates, the genetic interaction between *FBXO25* and *FBXO32* suggests that these two proteins may share similar functions or substrates. *FBXO25* and *FBXO32* have been individually proposed to have tumor suppressive function in previous studies (Xue et al., 2012; Zhou et al., 2017).

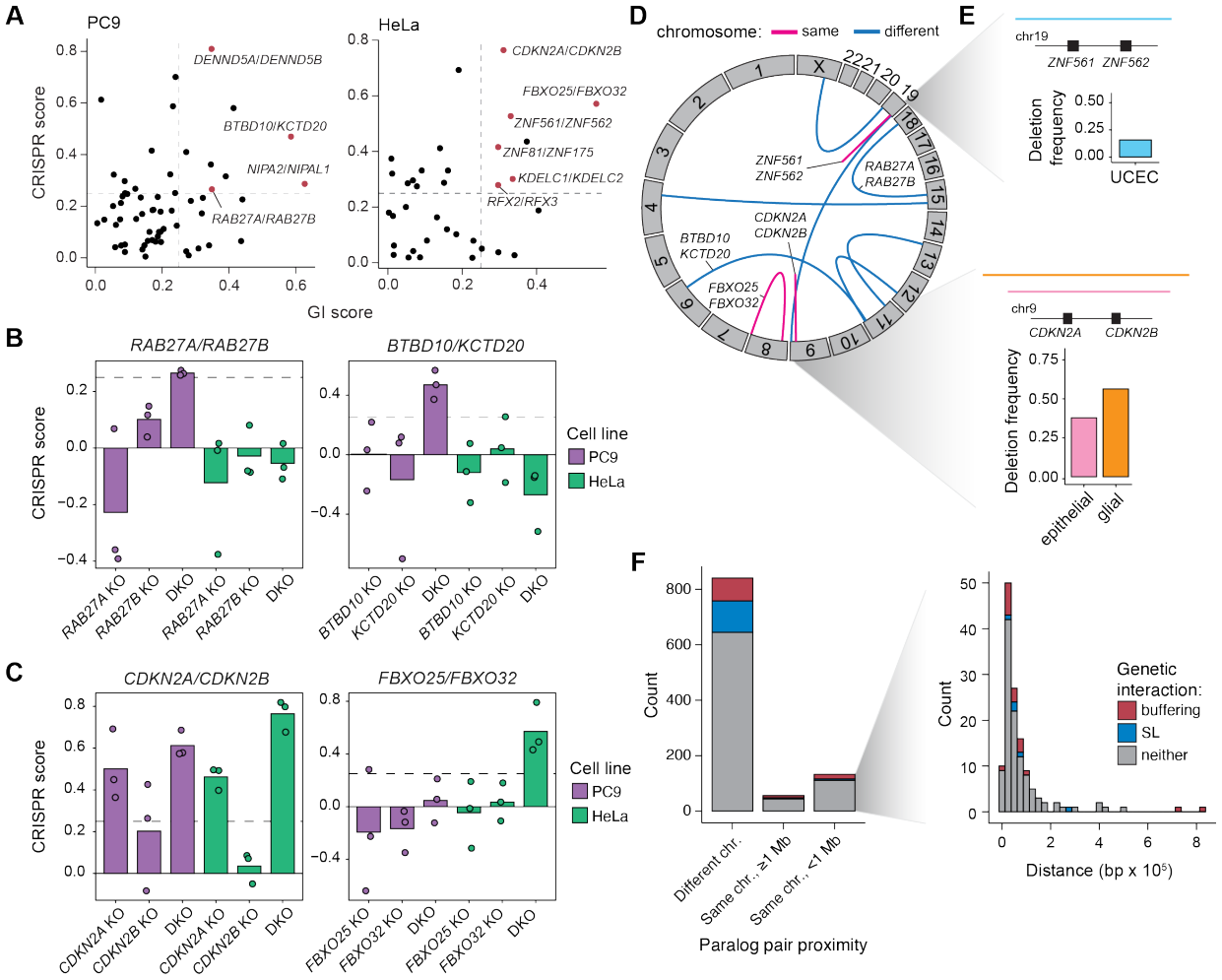


Figure 2.6: Paralog buffering interactions include tumor suppressor paralogs.

A) Identification of tumor suppressive paralog interactions (GI score > 0.25; FDR < 0.1; CRISPR score > 0.25).

B) CRISPR scores of PC9-specific tumor suppressive paralog pairs. Data shown is the mean of three biological replicates with replicate data shown in overlaid points. Dashed lines indicate CRISPR score = 0.25.

C) CRISPR scores of HeLa-specific tumor suppressive paralog pairs. Data shown as in (B).

D) Circos plot showing the genomic locations of tumor suppressive paralog pairs. Blue arcs indicate paralog pairs located on different chromosomes, while pink arcs represent paralog pairs located on the same chromosome.

E) Top, diagram of a recurrent deletion seen in uterine corpus endometrial carcinoma (UCEC) data from TCGA that spans the genomic locus containing *ZNF561* and *ZNF562*, and a bar plot indicating the deletion frequency. Bottom, diagram of recurrent deletions in epithelial and glial cancers that span the genomic locus containing *CDKN2A* and *CDKN2B*, and a bar plot showing the deletion frequency in each cancer subtype.

F) Genomic distance between paralogs for the 1,030 paralogs pairs included in the pgPEN library in three proximity categories: on different chromosomes, on the same chromosome but ≥1 megabase apart, and on the same chromosome within 1 megabase. Inset: Histogram of paralog distance for pairs that are within 1 megabase of one another.

While this direct identification of tumor suppressor paralog pairs has merit for understanding basic genome function, spontaneous loss of two unlinked genes in cancer should be rare, and therefore it is unlikely that double knockout of paralogs contributes to tumorigenesis for most paralog families. Interestingly however, we noted that two tumor suppressor pairs contained genes co-located in the same chromosomal locus (**Figure 2.6D**). In addition to *CDKN2A/CDKN2B*, whose combined loss is well known to promote tumorigenesis, *ZNF561* and *ZNF562* are also co-located and reside on chromosome 19p13.2, a frequently deleted region in uterine corpus endometrial cancer (A. C. Berger et al., 2018; Cherniack et al., 2017) (**Figure 2.6E**). Beyond the tumor suppressor paralogs, 13% of all the paralog pairs in the pgPEN library are located within 1 MB of each other in the human genome (**Figure 2.6F**), which raises the possibility that the cell fitness consequences of double knockout of human paralogs could contribute to the selective forces that drive aneuploidy patterns in human cancer (Ben-David & Amon, 2020; Taylor et al., 2018).

2.3 DISCUSSION

This work provides, to our knowledge, the largest direct experimental assessment of paralog genetic interactions in the human genome to date. The pgPEN library we developed uses two Cas9-type sgRNAs driven from independent promoters to enable knockout of two paralogs simultaneously and targets 2,060 duplicate human paralogs. Complementing three other recent studies of human paralog genetic interactions (Dede, McLaughlin, et al., 2020; Gonatopoulos-Pournatzis et al., 2020; Thompson et al., 2021), our library adds over 1,000 unique paralogs and brings the total set of human paralogs assayed to date to just over 3,900. Both Thompson et al. and the present study use Cas9-type CRISPR systems, whereas the other two studies include Cas12a-derived enzymes. Cas12a systems have the benefit of using an array of sgRNAs on a single transcript that is processed by Cas12a, enabling programmable delivery of multiple sgRNAs to the same cell (DeWeirdt et al., 2020). Continued application of Cas12a for CRISPR screening will enable the experimental identification of higher order combinatorial genetic interactions in human cells. However, for pairwise interactions of paralogs, the pgPEN library may provide an ease of application to investigators with Cas9-expressing cell systems already developed.

Remarkably, our pgPEN screens revealed that 12% of duplicate paralogs exhibit synthetic lethality, demonstrating that paralogs are a rich source of genetic interactions. A recent meta-analysis of multiplexed paralog CRISPR screens revealed that this hit rate was consistent with work published by other groups, and pairwise comparisons of the four studies showed that many shared paralogs (i.e., those that were screened by both groups) were consistently classified as synthetic lethal (De Kegel et al., 2021). These findings strengthen our conclusions, underscoring the importance of simultaneously targeting redundant genes and demonstrating that a large fraction of cancer dependencies are missed by current single-gene knockout approaches. Like others (De

Kegel & Ryan, 2019; Viswanathan et al., 2018), we propose that synthetic lethal interactions among paralogs could be harnessed for cancer therapy, since the aneuploid genomes typical of cancer cells commonly harbor deletions and inactivating mutations in one or more paralogs. Targeting lineage-specific essential paralogs or paralog families with partial loss in cancer could provide an orthogonal approach for cancer therapy to be applied in combination with existing therapies to provide durable cancer control and improved patient outcomes. In addition, even paralogs that are not lost in cancer may represent tractable cancer targets; the same homology and redundancy that complicates genetic identification of paralogs as cancer dependencies could enable simultaneous targeting of each protein with ease. This strategy is exemplified by the current use of small molecules targeting several of the top synthetic lethal paralogs we identified, such as *CDK4/CDK6* and *GSK3A/GSK3B*. Indeed, a recent study demonstrated in cell lines and retrospective clinical analyses that tumors with low CDK6 levels rely on CDK4 expression and show increased sensitivity to CDK4/6 inhibitors (Wu et al., 2021).

Last, we provide a systematic identification of tumor suppressive paralog pairs. We identified ten paralog pairs whose combined loss significantly promotes cancer cell line growth. Although combined loss of some of these pairs is likely to be rare, 2 of 10 we identified are located in the same chromosomal locus. Many of these loci are frequently deleted in cancer. These data therefore shed light on the basis for the positive selection of these genome deletions and suggest that combined paralog loss may shape the landscape of positive and negative selection in human cancer.

2.4 MATERIALS AND METHODS

2.4.1 *Experimental Model Details*

PC9 cells were originally derived from a metastatic lung adenocarcinoma from a 45 year old male patient. PC9-Cas9 cells were previously generated (Thomas et al., 2020) and cultured in RPMI-1640 (Gibco) supplemented with 10% Fetal Bovine Serum (FBS, Sigma). PC9-Cas9-GFP-NLS and PC9-Cas9-mCherry-NLS cells were generated by transducing PC9-Cas9 cells with lentivirus containing GFP-NLS or mCherry-NLS-encoding vectors (parental backbone was a gift from Dider Trono, Addgene #12252). mCherry or GFP positive cells were selected using flow cytometry. HeLa cells were originally derived from a cervical carcinoma from a 31 year old female patient. HeLa/iCas9 cells were previously generated (Cao et al., 2016) and cultured in Dulbecco's Modified Eagle's Medium (DMEM, Genesee Scientific) supplemented with 10% FBS. The HEK293T cell line was originally derived from kidney tissue from a female fetus. HEK293T cells were obtained from ATCC (CRL-3216) and cultured in DMEM supplemented with 10% FBS. All cells were maintained at 37ºC in 5% CO_2 and confirmed mycoplasma-free.

2.4.2 *Method Details*

2.4.2.1 Human paralog analysis and selection

For analysis of human paralog versus singleton essentiality (**Figure 2.1A-B**), a list of human protein-coding genes was obtained from Ensembl (Vilella et al., 2009). Mitochondrial genes and splice variants were removed from the analysis. The remaining genes were divided into two groups: (1) paralogous genes with >10% amino acid sequence identity and (2) singleton genes. For the pgPEN library, the list of human paralogs was further filtered to include only those with >50% reciprocal amino acid sequence identity with only one other gene. Genes encoding components of olfactory signaling and T cell receptors were also excluded. As shown in **Figure**

2.7A, a total of 2,060 paralogous genes (1,030 pairs) were included in the pgPEN library. Note that paralogs *MEK1/MEK2* may also be referred to as *MAP2K1/MAP2K2*.

2.4.2.2 PC9 single-gene CRISPR screen

PC9-Cas9 and PC9-Cas9-EGFR$^{T790M/L858R}$ single-gene CRISPR knockout screen data was re-analyzed from previously published data (Vichas et al., 2020). The relative essentiality of singletons versus paralogs in the PC9-Cas9 CRISPR knockout screen using the Brunello library (Doench et al., 2016) was assessed via a two-tailed Kolmogorov-Smirnov test (**Figure 2.1B**).

2.4.2.3 PC9 drug sensitivity profiling

PC9-Cas9 and PC9-Cas9-EGFR$^{T790M/L858R}$ erlotinib/trametinib drug sensitivity data was re-analyzed from previously published data (A. H. Berger et al., 2016). For combination dosing, erlotinib and trametinib were delivered to cells in a 1:1 molar ratio.

2.4.2.4 pgPEN library design and cloning

The pgPEN library was designed using sgRNA sequences selected from the Brunello library (Doench et al., 2016). sgRNAs containing BsmBI restriction target sequences and U6 termination signals were excluded from the library. Given that previous data demonstrated no position effects using the pgRNA approach (Gasperini et al., 2017), the sgRNA targeting a given gene was located at the same site in every pgRNA.

pgRNA oligonucleotides were synthesized by Twist Biosciences and cloned per published protocols (Gasperini et al., 2017; Thomas et al., 2020). Briefly, the pgRNA oligonucleotides were amplified (primers RKB1169 and RKB1170, **Table S5**) using NEBNext High Fidelity 2X Ready Mix (New England Biolabs) and purified via a 1.8X Ampure XP SPRI bead (Beckman Coulter) clean-up. Amplified oligonucleotides were then cloned into BsmBI (FastDigest Esp3I, Thermo Fisher Scientific)-digested lentiGuide-Puro (Addgene #52963) (Sanjana et al., 2014) plasmid

backbone via the NEBuilder HiFi (New England Biolabs) assembly system. Cloned plasmids were purified using a 0.8X Ampure bead clean-up and transformed into Endura ElectroCompetent *E. coli* cells (Lucigen) via electroporation to generate the pLGP-2xSpacer vector. The pLGP-2xSpacer vector was isolated using the NucleoBond Xtra Maxiprep kit (Macherey-Nagel) and linearized by BsmBI digest. A GBlock (synthesized by Integrated DNA Technologies) containing a second gRNA backbone and H1 promoter sequence was digested with BsmBI, purified via a 1.8X Ampure bead clean-up, and ligated into the pLGP-2xSpacer backbone using NEB Quick Ligase (New England Biolabs). The reaction product was purified using an 0.8X Ampure bead cleanup and transformed into Endura Electrocompetent *E. coli* via electroporation to propagate the final pLGP-pgRNA vectors. The pLGP-pgRNA plasmids were again isolated using the NucleoBond Xtra Maxiprep kit, and the cloned library was amplified and sequenced as described below to confirm high coverage. At each cloning step, individual *E. coli* colonies were sequence verified via colony PCR and Sanger sequencing with primer RKB1148 (**Table S5**). Over 1000X coverage of each pgRNA was maintained throughout plasmid library cloning, amplification, and sequencing; coverage depth was selected based on our previous screen experience as well as published recommendations (Doench, 2018; Joung et al., 2017).

2.4.2.5 Lentivirus production and titration

With our cloned library, we produced lentivirus via a large-format transfection in HEK293T cells using a protocol adapted from Joung *et al*. (Joung et al., 2017). Briefly, we used TransIT-LT1 (Mirus Bio) as a transfection reagent, with packaging plasmid psPAX2 (Addgene #12260) (Stewart et al., 2003) and envelope plasmid pCMV-VSV-G (Addgene #8454; plasmid was a gift from Didier Trono) and Opti-MEM (Thermo Fisher Scientific). Plasmids were added at a 4:2:1 ratio of transfer to packaging to envelope plasmid. At 18 hours post-transfection, media

was changed to high-serum DMEM (30% FBS). Lentivirus was harvested 48 hours post-transfection. Over 500X coverage of each pgRNA was maintained throughout; coverage depth was selected based on our previous screen experience as well as published recommendations (Doench, 2018; Joung et al., 2017).

2.4.2.6 pgPEN CRISPR screens

PC9-Cas9 and HeLa/iCas9 cells were transduced with the pgPEN library at low multiplicity of infection (~0.3) to ensure the integration of a single pgRNA construct into >95% of transduced cells (Doench, 2018). Transduced cells were then selected using puromycin (1.0 μg/mL, Sigma) for 48-72 hours until all uninfected control cells were dead. For the PC9-Cas9 screen, cells were split into three biological replicates after infection but before puromycin selection, and genomic DNA (gDNA) was harvested from each replicate after puromycin selection for an early time point sample. For the HeLa/iCas9 screen, cells were kept in the pooled format until puromycin selection was complete, resulting in a single early time point sample. HeLa/iCas9 cells were then induced using doxycycline (1.0 μg/mL, Sigma) and split into three biological replicates. For both screens, cells were then passaged for approximately 12 population doublings while maintaining over 500X coverage of each pgRNA at every step. An endpoint gDNA sample was harvested from each biological replicate and stored at –80ºC. Genomic DNA was extracted using the QIAamp DNA Blood Maxi Kit (Qiagen).

2.4.2.7 pgPEN library preparation and sequencing

Plasmid and gDNA samples were amplified and sequenced at >500X coverage per pgRNA according to our previously established methods (Thomas et al., 2020). All primer sequences used for library preparation are included in **Table S5**. First, 2.5 μg of gDNA was used as input for each reaction, with a total of 48 reactions (120 μg total input gDNA) to ensure >500X coverage per

sample. Input DNA was amplified using NEBNext High Fidelity 2X Ready Mix with primers RKB2713/RKB2714 followed by 1.8X Ampure bead clean-up. Second, the amplicon from PCR #1 was used as input for PCR #2, with 10 ng input DNA in one reaction per sample. The input DNA was amplified using primers RKB2715/RKB2716 followed by 1X Ampure bead clean-up. Third, 10 ng of the amplicon from PCR #2 was used as input for PCR #3 and was amplified using a common forward primer (RKB2717) and a sample-specific barcoded reverse primer (see **Table S5**) to allow for multiplexed sequencing. Product from PCR #3 was purified using a 1X Ampure bead clean-up, quantified by a Qubit assay (Thermo Fisher Scientific), and pooled at equimolar amounts prior to Illumina sequencing. The custom sequencing strategy used for pgPEN is outlined in **Figure 2.7D**.

2.4.2.8 pgRNA cloning for validation

Validation pgRNA oligonucleotides consisted of two sgRNA sequences separated by the H1 promoter and were synthesized by Genewiz (Brooks Life Sciences). All pgRNA sequences used for validation experiments are available in **Table S5**. Each validation pgRNA was cloned into a BsmBI-digested LentiGuide-Puro backbone using a one-step Gibson reaction using the NEBuilder HiFi (New England Biolabs) DNA assembly system, as described above for pgPEN library cloning. Cloned plasmids were transformed into One Shot Stbl3 Chemically Competent *E. coli* (Invitrogen). Individual colonies were sequence verified via colony Sanger sequencing with primer RKB1148 (**Table S5**), then the final pgRNA vector was isolated using the Plasmid Plus Midi Kit (Qiagen).

2.4.2.9 Competitive fitness assay

For the bichromatic competitive fitness assay, PC9-Cas9-GFP-NLS cells were transduced with a control pgRNA and PC9-Cas9-mCherry-NLS cells with either a paralog single KO pgRNA

or a paralog DKO pgRNA. After 48-72 hours of selection with puromycin (1 µg/mL), cells were pooled at an equal ratio and seeded in tissue culture-treated plates (Corning). The day cells were pooled was termed Day 0. For the *MAGOH/MAGOHB* paralog pair, non-targeting control (NTC) gRNAs were used as controls and each competition (double NTC vs. *MAGOH* single KO, double NTC vs. *MAGOHB* single KO, double NTC vs. *MAGOH/MAGOHB* DKO) was carried out in triplicate. For all other competitive fitness assays (*CCNL1/CCNL2*, *CDK4/CDK6*, *MEK1/MEK2*, *OXSR1/STK39*, and *PSMB5/PSMB8*), safe-targeting gRNAs that target intergenic regions (Morgens et al., 2017) were used as controls to account for the different number of double-strand breaks generated by single KO versus DKO pgRNAs and each competition was carried out in six biological replicates. After pooling (Day 0), cells were imaged 24 hours later (Day 1) using a Cytation 5 imager (BioTek Instruments). Raw counts of mCherry- and GFP-expressing cells were computed using Gen5 v3.02 software (BioTek Instruments) to determine the initial paralog-targeting and safe-targeting pgRNA abundance. Cells were imaged and mCherry and GFP cell counts were taken every 1-3 days.

2.4.2.10 Genomic DNA sequencing for validation

Genomic DNA was extracted from validation cell lines using the DNeasy Blood & Tissue Kit (Qiagen). On-target editing efficiencies for each gene target were determined via PCR and Sanger or next-generation sequencing. sgRNA target regions were amplified using NEBNExt High-Fidelity 2X PCR Master Mix (New England Biolabs) and custom primers designed for each target; primer sequences are available in **Table S5**. PCR products were then purified via Ampure bead clean-up and submitted for sequencing by Genewiz. For Sanger sequencing, results were analyzed using the online tool Tracking of Indels by Deconvolution (TIDE, http://shinyapps.datacurators.nl/tide/) which uses Sanger traces to approximate CRISPR editing

efficiencies (Brinkman et al., 2014). Next-generation sequencing results were mapped, aligned and CRISPR indels were quantified using the CRISPREsso v2 pipeline (Clement et al., 2019).

2.4.2.11 Cell lysis and western blotting for validation

Whole-cell extracts for immunoblotting were prepared by incubating cells on a rocker at 4ºC in RTK lysis buffer [20 mM Tris (pH 8.0), 2 mM EDTA (pH 8), 137 mM NaCl, 1% IGEPAL CA-630, 10%Glycerol] plus Pierce protease and phosphatase inhibitors (Thermo Scientific) for 20 minutes. Following centrifugation (>15,000 x g for 20 minutes at 4ºC), protein lysates were quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). Lysates were separated by SDS–PAGE and transferred to PVDF membranes using the Trans-blot Turbo Transfer System (BioRad). Membranes were blocked in Intercept Blocking Buffer (LiCOR) with 0.1% Tween 20 Solution (BioRad) for 1 hour at room temperature followed by overnight incubation at 4°C with primary antibodies diluted in blocking buffer. IRDye (LiCOR) secondary antibodies were used for detection and were imaged on Odyssey CLx Imaging system (LiCOR). Loading control and experimental protein were probed on the same membrane in all cases. For clarity, loading control is cropped and shown below experimental condition in all panels regardless of the relative molecular weights of the two proteins.

Primary antibodies used for Western blotting: CCNL2 (Novus Biologicals #NB100-87009, 1:2000), MEK1 (Cell Signaling Technology #2352, 1:1000), MEK2 (Cell Signaling Technology #9147, 1:1000), OXSR1 (alias OSR1, Cell Signaling Technology #3729, 1:1000), STK39 (alias SPAK, Cell Signaling Technology #2281, 1:500), CDK4 (Cell Signaling Technology #12790, 1:1000), CDK6 (Cell Signaling Technology #13331, 1:1000), vinculin (Sigma #V9264, 1:10,000).

2.4.3 *Quantification and Statistical Analysis*

Unless otherwise noted, results were analyzed for statistical significance with Rv3.6.3 in an Rstudio v1.2.5 environment. Statistical details of all experiments can be found below or in the corresponding figure legends.

2.4.3.1 pgPEN CRISPR screen sequencing analysis

Sequencing, image analysis, and base calling for pgPEN screens were performed on the Illumina HiSeq 2500 with RTA 1.18.66.3 software. FASTQ files were generated using Illumina's bcl2fastq v2.20 conversion software. Reads were trimmed using FASTX-Toolkit v0.014, and samples were demultiplexed using idemp (https://github.com/yhwu/idemp). Sequencing reads for each pgRNA were mapped separately to the pgPEN library annotation using Bowtie v1.2.2 (Langmead et al., 2009). Aligned SAM files were converted to BAM format and sorted using SAMtools v1.9 (H. Li et al., 2009). pgRNA counts were obtained using a custom R script (https://doi.org/10.5281/zenodo.5081113) with R v3.6.2 and R packages Rsamtools v1.34.1 (accessed via Bioconductor v1.3.0, (Huber et al., 2015)) and Tidyverse v1.2.1 (Wickham et al., 2019). Based on the reference set, correctly-paired pgRNAs were retained while incorrectly-paired gRNAs were discarded. pgRNAs with <2 reads per million (RPM) in the plasmid pool or with a read count of zero at any time point were also removed. The $\log_2$-scaled fold change (LFC) of each pgRNA was then computed using the MAGeCK v0.5.9.2 (W. Li et al., 2014) test command to compare initial abundance in the plasmid pool to abundance at early and late time points.

LFC values were scaled so that the median of negative control (double non-targeting) pgRNAs was set to zero, while the median of positive control (single-targeting pgRNAs targeting Project Achilles pan-essential genes (Meyers et al., 2017)) pgRNAs was set to –1 (**Figure 2.2B** and **Figure 2.11D**). We also used RNA-seq data from each cell line (Thomas et al., 2020) to control

for growth defects caused by the double-strand break generation and repair process. To do this, we adjusted pgRNA LFCs so that the median LFC of single- and double-targeting pgRNAs targeting unexpressed genes (TPM < 2) was set to zero (**Figure 2.2C**, **Figure 2.8D-E**, and **Figure 2.11E-G**). Finally, we analyzed copy number effects using data from DepMap, accessed via Bioconductor v1.3.0 (Huber et al., 2015) package depmap v1.0.0. We grouped pgRNAs by the combined copy number of targeted genes for each construct and analyzed the CRISPR scores of each copy number group (**Figure 2.8F-H** and **Figure 2.11H-J**). Given that the copy number of the vast majority of paralogs included in our library was close to 2, we did not adjust for copy number effects. The scaled and normalized LFC for each pgRNA was termed a CRISPR score (CS). Target-level CRISPR scores were calculated by taking the mean across pgRNAs with the same single KO or DKO paralog target. Final CRISPR scores were computed by taking the mean across the three biological replicates for each screen.

2.4.3.2 Genetic interaction score calculations

To compute a genetic interaction (GI) score for each paralog pair, we combined two previously published methods for genetic interaction mapping in human cells (DeWeirdt et al., 2020; Han et al., 2017). We first calculated an expected and observed CS for each pgRNA. For DKO pgRNAs (pgRNA-Paralog1_Paralog2), we calculated the expected CS by first taking the mean CRISPR scores of each single KO pgRNA with the same targeting sgRNA sequence paired with a non-targeting control (NTC) sgRNA sequence (i.e., mean(pgRNA-Paralog1_NTC1, pgRNA-Paralog1_NTC2) and mean(pgRNA-NTC1_Paralog2, pgRNA-NTC2_Paralog2)). We summed these two single KO mean CS values to calculate an expected CS for each paralog pair, and compared this expected CS to the observed DKO CS (pgRNA-Paralog1_Paralog2). To establish a distribution of non-interacting GI scores, we used single KO pgRNAs as a negative

control. We calculated an expected CS for single KO pgRNAs by computing the sum of (1) the CS for the other single KO pgRNA containing the same targeting sgRNA sequence paired with a different NTC sgRNA sequence (pgRNA-Paralog1_NTC2) and (2) the mean CS of double NTC pgRNAs (pgRNA-NTC1_NTC2) containing the same NTC sgRNA sequence (i.e., mean(pgRNA-NTC1_NTC2, pgRNA-NTC1_NTC3)). This single KO expected CS was then compared to the observed single KO CS (pgRNA-Paralog1_NTC1 or pgRNA-NTC1_Paralog2). Target-level single KO and DKO expected and observed CRISPR scores were calculated by taking the mean across pgRNAs.

We then obtained the distribution of CRISPR scores for control (single KO) pgRNAs by calculating the linear regression of control expected versus observed CS values (**Figure 2.3A** for PC9 and **Figure 2.11K** for HeLa). GI scores were determined by calculating the residual of each observed CS value for each paralog pair from the control regression line. Statistical significance of DKO GI scores was determined using a t test compared to the distribution of control (single KO) GI scores. A Benjamini-Hochberg false discovery rate (FDR) correction (Benjamini & Hochberg, 1995) was then applied, and FDR < 0.1 was considered significant.

Synthetic lethal paralogs were defined as those with a GI score < –0.5 and FDR < 0.1, while buffering paralogs were defined as those with GI score > 0.25 and FDR < 0.1. Tumor suppressor paralogs were defined as buffering paralogs with an additional filter for DKO CS > 0.25 in either PC9 or HeLa cells. Cancer deletion data for paralog tumor suppressor analysis shown in **Figure 2.6E** were obtained from The Cancer Genome Atlas Copy Number Portal (Beroukhim et al., 2010).

2.4.3.3 Competitive fitness assay analysis

For each competition, the ratio of mCherry (targeting) to GFP (control) cells was computed for each sample replicate at each time point. The $\log_2$ fold enrichment of pgRNAs relative to Day 1 was calculated by dividing each subsequent day's mCherry:GFP ratio by the Day 1 ratio. The expected DKO growth effect under a null model of no interaction was calculated by arbitrarily pairing single KO target (i.e., Gene1_Safe and Gene2_Safe) replicates to calculate the sum of single KO growth effects. For each paralog-targeting pgRNA and for the previously calculated expected (single KO sum) growth phenotypes, the mean and standard error of the mean (SEM) of across replicates were calculated at each time point. Failed replicates were excluded from the analysis. The expected growth effect was compared to the observed DKO growth effects using a one-tailed t test (**Figure 2.4A** and **Figure 2.9E-H, K-L**).

2.4.3.4 DepMap validation analysis

For each paralog pair, we determined the effect of CRISPR-mediated knockout of one gene in cell lines with high (top quartile) compared to low (bottom quartile) expression or copy number of its paralogous gene (**Figure 2.5E-G**). CRISPR score, expression, and copy number data was obtained from DepMap via the Bionconductor package depmap v1.0.0. We compared the effect of paralog 1 knockout in paralog 2 low vs. high cell lines using a two-tailed Wilcoxon Rank-Sum test.

2.5 ACKNOWLEDGMENTS

We thank Dr. Athea Vichas (Fred Hutchinson Cancer Research Center) for advice on CRISPR screens in PC9 cells and Madeleine Duran (University of Washington) for assistance with improving the efficiency and reproducibility of custom scripts.

2.5.1 *Funding*

This work was funded with support from the Lung Cancer Research Foundation. P.C.R.P. was supported by NSF DGE-1762114 and NIH T32-HG000035. J.D.T. is a Washington Research Foundation Postdoctoral Fellow. A.H.B. was supported in part by the NIH/NCI (R00 CA197762 and R37 CA252050); the Devereaux Outstanding Investigator Award from the Prevent Cancer Foundation; the Stephen H. Petersdorf Lung Cancer Research Award; Seattle Translational Tumor Research program; and the Innovators Network Endowed Chair. R.K.B. was supported in part by the NIH/NIDDK (R01 DK103854); NIH/NHLBI (R01 HL128239 and R01 HL151651); NIH/NCI (R01 CA251138); Edward P. Evans Foundation; Blood Cancer Discoveries Grant program through the Leukemia & Lymphoma Society, Mark Foundation for Cancer Research, and Paul G. Allen Frontiers Group (8023-20); Dept. of Defense Breast Cancer Research Program (W81XWH-20-1-0596); and the McIlwain Family Endowed Chair in Data Science. R.K.B. is a Scholar of The Leukemia & Lymphoma Society (1344-18). Sequencing was performed by the Fred Hutch Genomics Shared Resource (supported by NIH/NCI Cancer Center Support Grant P30 CA015704). Computational studies were supported in part by FHCRC's Scientific Computing Infrastructure (ORIP S10 OD028685).

2.5.2 *Author Contributions*

A.H.B conceived of the project. A.H.B. and R.K.B. supervised the project. P.C.R.P, J.D.T., and A.H.B. designed the study. P.C.R.P., J.D.T., S.K., and A.G. conducted experiments. P.C.R.P, J.D.T., and A.G. analyzed the data. P.C.R.P., J.D.T., and A.H.B. wrote the manuscript, with input from all authors.

2.5.3 *Competing Interests*

The authors declare no competing interests.

2.5.4 *Data and Materials Availability*

Data and materials used for the pgPEN study are outlined in **Table 2.1**, the Key Resources Table.

Table 2.1: Key Resources Table.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-CCNL2 (Cyclin L2)	Novus Biologicals	Cat #NB100-87009; RRID:AB_1201144
Rabbit polyclonal anti-OXSR1 (OSR1)	Cell Signaling Technology	Cat #3729; RRID:AB_2157610
Rabbit polyclonal anti-STK39 (SPAK)	Cell Signaling Technology	Cat #2281; RRID:AB_2196951
Mouse monoclonal anti-MEK1	Cell Signaling Technology	Cat #2352; RRID:AB_10693788
Rabbit monoclonal anti-MEK2	Cell Signaling Technology	Cat #9147; RRID:AB_2140641
Rabbit monoclonal anti-CDK4	Cell Signaling Technology	Cat #12790; RRID:AB_2631166
Rabbit monoclonal anti-CDK6	Cell Signaling Technology	Cat #13331; RRID:AB_2721897
Mouse monoclonal anti-vinculin	Sigma	Cat #V9264; RRID:AB_10603627
Deposited data		
Raw and analyzed CRISPR screen data	This paper	GEO: GSE178179
DepMap	Tsherniak et al., 2017	10.1016/j.cell.2017.06.010
TCGA Copy Number Portal	Beroukhim et al., 2010	10.1016/j.cell.2017.06.010
PC9 and HeLa RNAseq data	Thomas et al., 2020	GEO: GSE120703

EnsemblCompara GeneTrees	Vilella et al., 2009	10.1101/gr.073585.107
Experimental models: Cell lines		
Human: PC9-Cas9 cells	Thomas et al., 2020	https://doi.org/10.1038/s41588-019-0555-z
Human: iCas9/HeLa cells	Thomas et al., 2020	https://doi.org/10.1038/s41588-019-0555-z
Human: PC9-Cas9-mCherry-NLS cells	This paper	N/A
Human: PC9-Cas9-GFP-NLS cells	This paper	N/A
Human: HEK293T cells	ATCC	CRL-3216
Oligonucleotides		
All oligos used, see Table S5	This paper	N/A
Recombinant DNA		
pgPEN plasmid library	This paper	Addgene #171172
pLentiGuide-Puro	Sanjana et al., 2014	Addgene #52963
pRRLSIN.cPPT.PGK-GFP.WPRE	N/A (unpublished)	Addgene #12252
psPAX2	Stewart et al. 2003	Addgene #12260
pCMV-VSV-G	N/A (unpublished)	Addgene #8454
Software and algorithms		
pgRNA_sequencing_analysis	This paper; Zenodo	https://doi.org/10.5281/zenodo.5081113
MAGeCK v0.5.9.2	Li et al., 2014	https://sourceforge.net/projects/mageck/
Bowtie v1.2.2	Langmead et al., 2009	https://sourceforge.net/projects/bowtie-bio/; RRID:SCR_005476
SAMtools v1.9	Li et al., 2009	https://www.htslib.org; RRID:SCR_002105
Tidyverse v1.3.0	Wickham et al., 2019	https://CRAN.R-project.org/package=tidyverse; RRID:SCR_019186
Bioconductor v3.1.0	Huber et al., 2015	https://www.bioconductor.org/install; RRID:SCR_006442

Gen5 v3.02	BioTek Instruments	https://www.biotek.com/products/software-robotics-software/gen5-microplate-reader-and-imager-software/; RRID:SCR_017317
TIDE	Brinkman et al., 2014	http://shinyapps.datacurators.nl/tide/
CRISPResso	Clement et al., 2019	https://github.com/pinellolab/CRISPResso2

2.5.4.1 Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Alice Berger (ahberger@fredhutch.org).

2.5.4.2 Materials Availability

The pgPEN CRISPR plasmid library has been deposited to Addgene as the Human Paralog Knockout Library (pgPEN), #171172. All other plasmids and cell lines generated for this study will be shared by the lead contact upon request.

2.5.4.3 Data and Code Availability

RNA-seq data and raw and processed CRISPR screen (GEO GSE120703) and sequencing data for the PC9-Cas9 and HeLa/iCas9 pgPEN CRISPR screens (GEO GSE178179) has been deposited to GEO. All GEO-deposited data is listed in the key resources table (**Table 2.1**) and is publicly available as of the date of publication. This paper also analyzes existing, publicly available datasets. The accession numbers for these datasets are listed in the key resources table. All other data reported in this paper will be shared by the lead contact upon request.

Custom pgRNA counting code can be downloaded from Zenodo: https://doi.org/10.5281/zenodo.5081113.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

2.6 SUPPLEMENTAL DATA

2.6.1 *Supplemental Tables*

All Supplemental Tables referenced in this chapter can be found at the following link: https://doi.org/10.1016/j.celrep.2021.109597.

2.6.2 *Supplemental Figures*

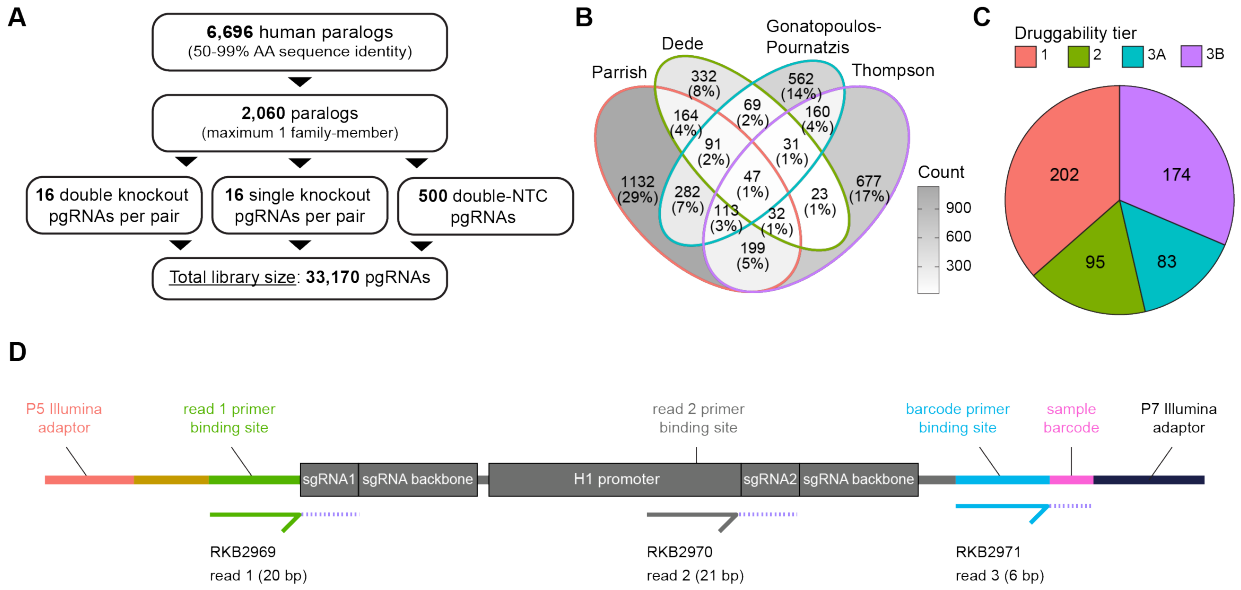


Figure 2.7: pgPEN CRISPR library information and sequencing strategy.

A) Schematic of filtering strategy used to select paralogous genes for inclusion in the pgPEN library. AA, amino acid.

B) Venn diagram of paralogs included in this study, Dede et al. (2020), Gonatopoulos-Pournatzis et al. (2020), and Thompson et al. (2021).

C) Pie chart depicting the number of genes considered "druggable" based on classification from Finan et al. (2017). Tier 1 genes encode proteins targetable by approved and clinical trial-phase drugs; Tier 2 genes encode proteins targetable by known small molecules or have high sequence identity to known drug targets; Tiers 3A and 3B include secreted and extracellular proteins, as well as those with lower sequence identity to known drug targets and other members of known druggable families not included in higher tiers.

D) Schematic of pgPEN sequencing strategy. Custom sequencing primers and read lengths to sequence sgRNA1 (standard Illumina read 1), sgRNA2 (standard Illumina i7/index 1 read), and the sample barcode (standard Illumina i5/index 2 read) are indicated.

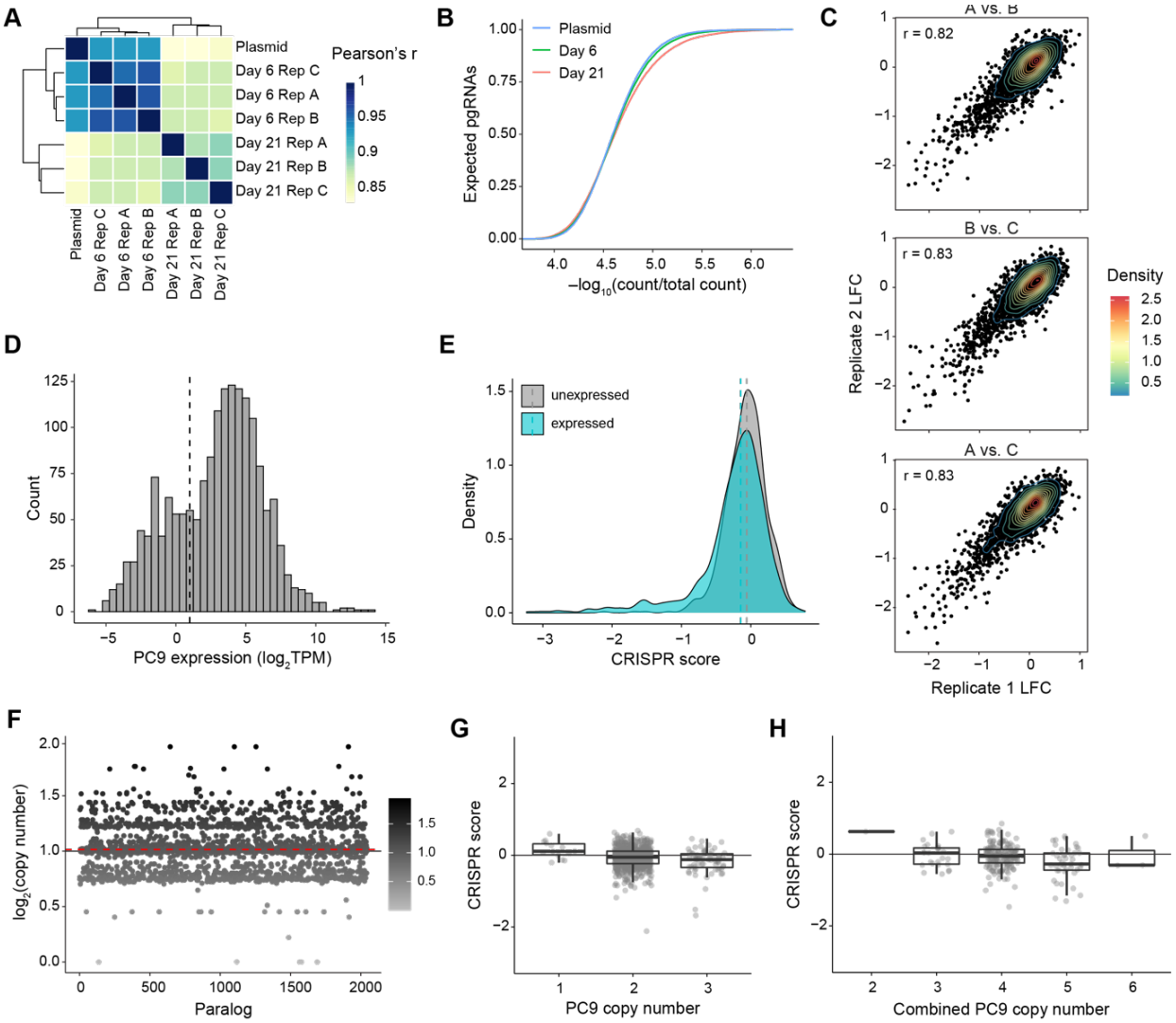


Figure 2.8: PC9 paralog screen quality control, gene expression, and copy number information.

A) Heat map illustrating Pearson correlations between sequencing reads in counts per million (CPM) supporting each pgRNA for all samples from the PC9 screen. Dendrogram, unsupervised clustering of CPM by the complete-linkage method.

B) Cumulative distribution of pgRNA counts at each time point (mean across three biological replicates) in the PC9 screen. Each point represents the proportion of total pgRNAs with the corresponding count in a given sample.

C) Scatter plot of pgRNA target $\log_2$(fold change) (LFC) values for the late time point compared to plasmid for the indicated replicate comparisons in the PC9 screen. Contour lines indicate the density of points in 2D space.

D) Histogram of gene expression in PC9 cells for all paralogs in the pgPEN library. Dashed line indicates $\log_2$(TPM) = 1, the cutoff used for expression. TPM, transcripts per million.

E) Density histogram of CRISPR scores for single KO pgRNA targets, grouped by target gene expression in PC9 cells. Dashed lines indicate the median CRISPR score for each group. pgRNAs targeting expressed genes had significantly lower CRISPR scores than those targeting unexpressed genes ($p = 1.96e{-10}$ by one-tailed Kolmogorov-Smirnov test).

F) Scatter plot of PC9 $\log_2$(copy number) for paralogs in the pgPEN library. Red dashed line indicates the mean PC9 copy number of pgPEN library paralogs. Copy number data was obtained from DepMap.

G) Box and overlaid dot plot of CRISPR scores for single KO paralog target genes, grouped by PC9 copy number. Box plots indicate median ± interquartile range (IQR) for each group.

H) Box and overlaid dot plot of CRISPR scores for double KO paralog target gene pairs, grouped by the combined PC9 copy number of the two genes. Box plots indicate median ± IQR for each group.

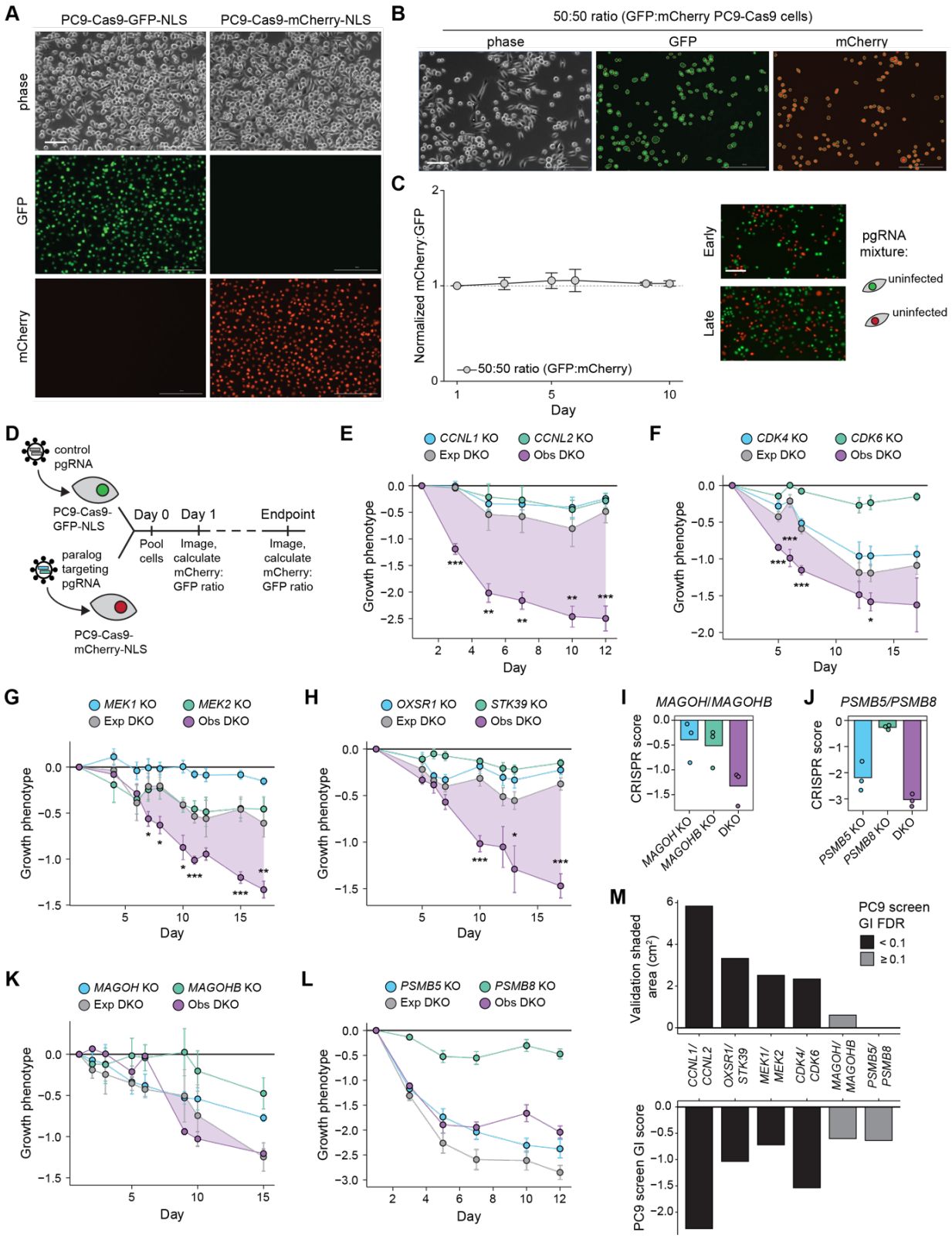


Figure 2.9: Validation of PC9 screen data using competitive fitness assays.

A) Representative images of PC9-Cas9-GFP-NLS and PC9-Cas9-mCherry-NLS cells. Scale bar, 200 μM.

B) Representative images of detecting GFP and mCherry positive cells within a 50:50 mixture. Scale bar, 200 μM.

C) Line graph of PC9-Cas9-GFP-NLS and PC9-Cas9-mCherry-NLS cell relative viability across a 10-day time course. Data shown is the mean ± SEM of three biological replicates. Right, representative images of mixture at early (Day 1) and late (Day 10) time points are shown. Scale bar, 100 μM.

D) Schematic of the competitive fitness assay.

E) Competitive fitness assay time course for CCNL1/CCNL2 showing line graph of growth phenotypes for PC9-Cas9-mCherry-NLS cells expressing a pgRNA to generate the indicated gene KO compared to PC9-Cas9-GFP-NLS cells expressing a double-safe-targeting control pgRNA. Growth phenotype is defined as the log2 ratio of mCherry:GFP cell counts at the late time point compared to the day 1 mCherry:GFP cell count ratio. Data shown is the mean ± SEM of six biological replicates. Single KO pgRNAs contained one gene-targeting sgRNA and one safe-targeting sgRNA. Expected DKO growth phenotypes (labeled "Exp DKO") were calculated as the sum of single KO growth phenotypes, with purple shaded regions indicating the excess negative growth effect seen in the observed DKO (labeled "Obs DKO") compared to the null model of no interaction. One-tailed t test p values for expected vs. observed DKO growth phenotypes are indicated as follows: * = $p < 0.05$, ** = $p < 0.01$, and *** = $p < 0.001$.

F) As (E), but for CDK4/CDK6.

G) As (E), but for MEK1/MEK2.

H) As (E), but for OXSR1/STK39.

I) CRISPR scores for the MAGOH/MAGOHB paralog pair in the PC9 screen. Data shown is the mean CRISPR score for each single KO or DKO target across three biological replicates with replicate data shown in overlaid points.

J) As in (I), but for PSMB5/PSMB8.

K) Line graph of MAGOH/MAGOHB competitive fitness assay time course, performed as in (E) except that the data shown is the mean ± SEM of three biological replicates and non-targeting sgRNAs were used instead of safe-targeting sgRNAs. Statistical analysis and significance symbols are as described in (E).

L) As in (E), but for PSMB5/PSMB8.

M) Bar plots comparing the GIs observed in the validation competitive growth assays to those seen in the PC9 CRISPR screen. Top panel shows the area of the purple shaded region from panels E-H and K-L, and bottom panel shows the GI score calculated for each paralog pair in the PC9 pgPEN screen. Bars are colored to indicate whether the PC9 screen GI FDR was considered statistically significant at a cutoff of FDR < 0.1.

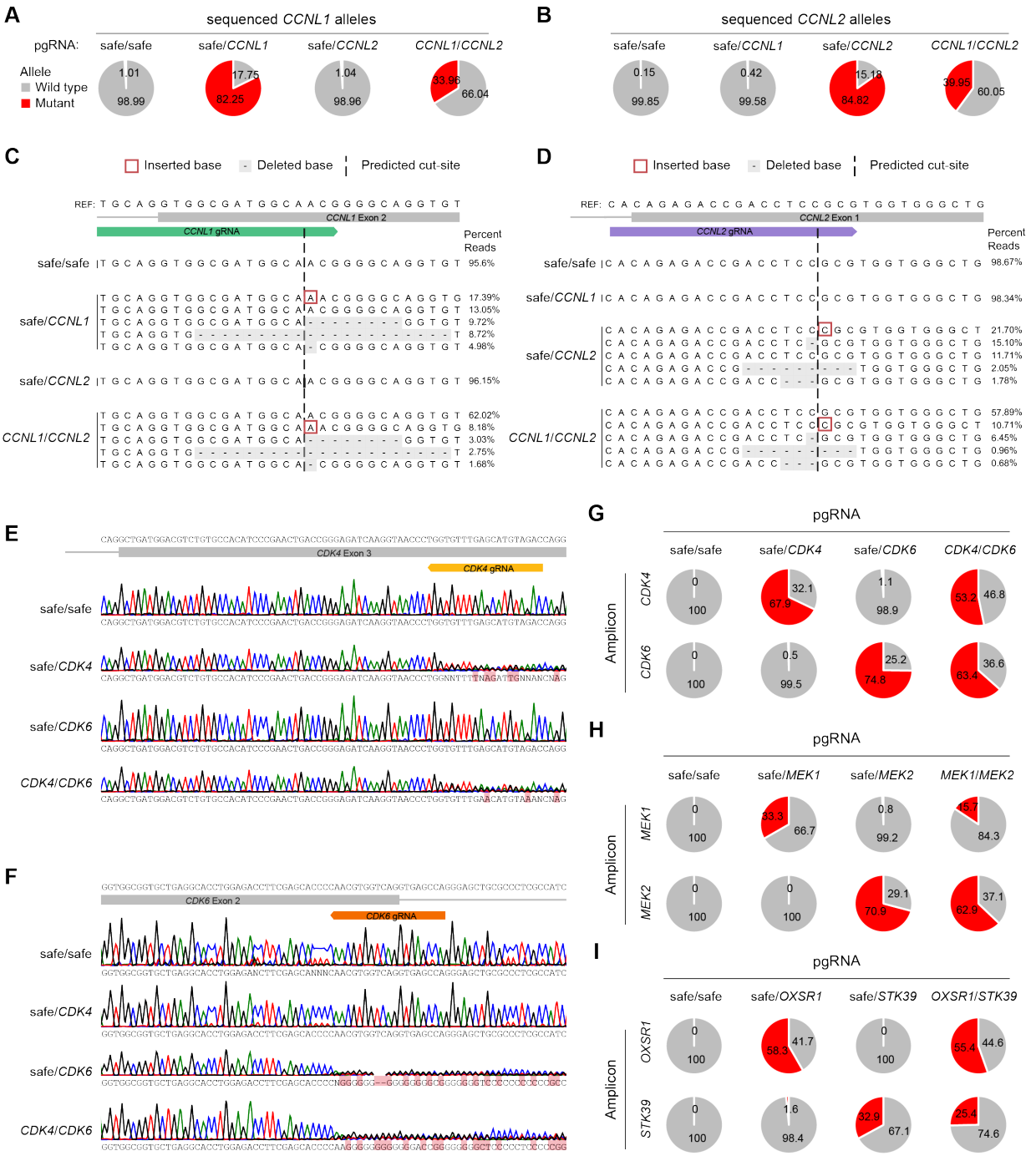


Figure 2.10: Genomic DNA sequencing data for PC9 screen validation paralog pairs.

A) Pie charts summarizing next-generation sequencing of *CCNL1* alleles from PC9 cells expressing Cas9 and the indicated pgRNAs. Percent wild-type (gray) and mutant (red) alleles are indicated.

B) As in (A), but for *CCNL2*.

C) Sequences of the five most frequent alleles from (A). The location of the *CCNL1*-targeting sgRNA is also shown.

D) Sequences of the five most frequent alleles from (B). The location of the *CCNL2*-targeting sgRNA is also shown.

E) Representative Sanger sequencing of *CDK4* from PC9 cells expressing Cas9 and the indicated pgRNAs. The location of the *CDK4*-targeting sgRNA is also shown.

F) As in (E), but for *CDK6*.

G) Pie charts summarizing Sanger sequencing of *CDK4* and *CDK6* alleles from PC9 cells expressing Cas9 and the indicated pgRNAs. Percent wild-type (gray) and mutant (red) alleles are indicated.

H) As in (G), but for *MEK1* and *MEK2*.

I) As in (G), but for *OXSR1* and *STK39*.

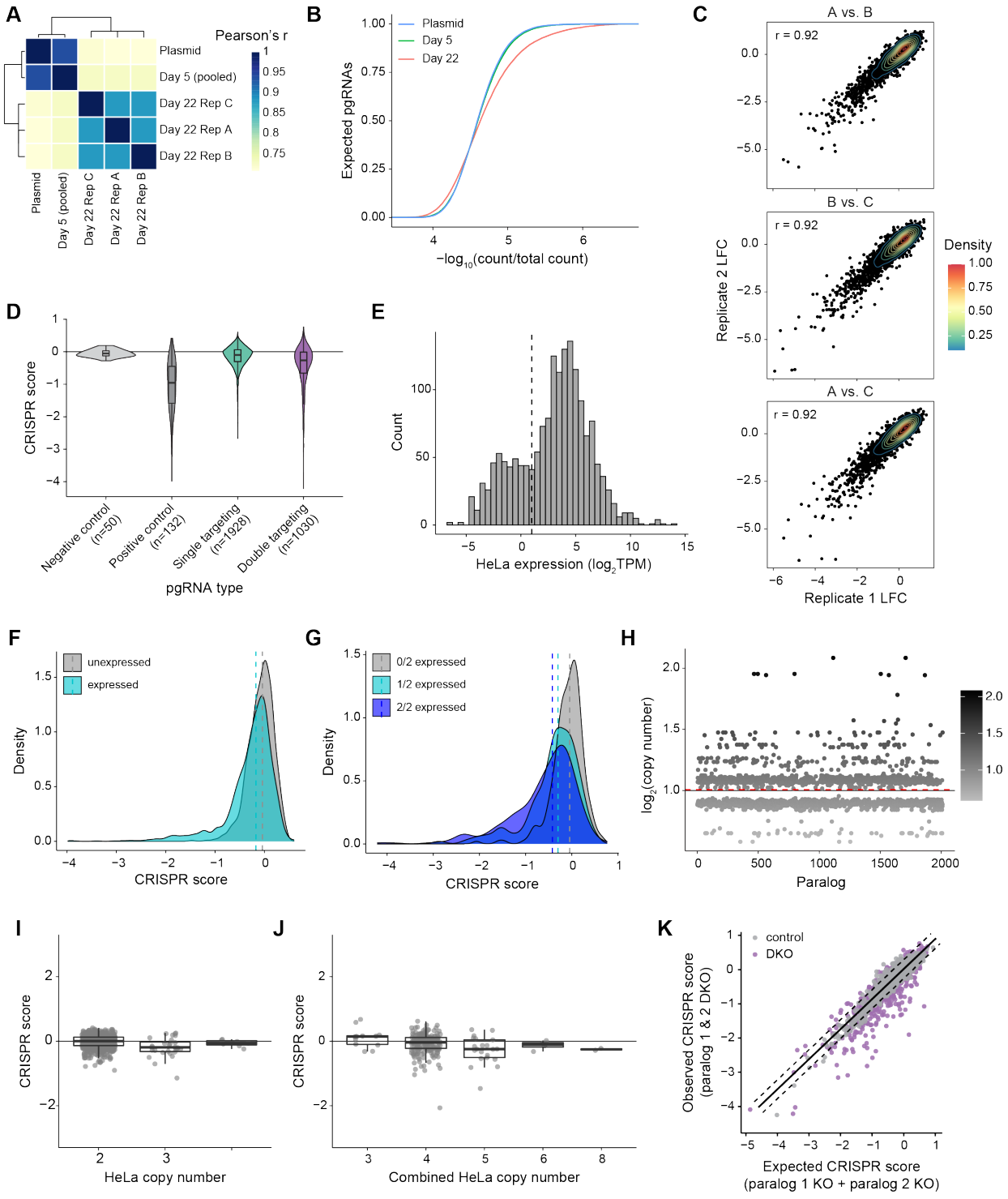


Figure 2.11: HeLa paralog screen quality control, gene expression, copy number, and genetic interaction score information.

A) Heat map illustrating Pearson correlations between sequencing reads in counts per million (CPM) supporting each pgRNA for all samples from the HeLa screen. Dendrogram, unsupervised clustering of CPM by the complete-linkage method.

B) Cumulative distribution of pgRNA counts at each time point (mean across three biological replicates) in the HeLa screen. Each point represents the proportion of total pgRNAs with the corresponding count in a given sample.

C) Scatter plot of pgRNA target LFC values for the late time point compared to plasmid for the indicated replicate comparisons in the PC9 screen. Contour lines indicate the density of points in 2D space.

D) Violin plots of target-level CRISPR scores for negative control (double non-targeting control), positive control (single KO pgRNAs targeting known essential genes), all other single KO pgRNAs, and DKO pgRNAs in the HeLa screen.

E) Histogram of gene expression in HeLa cells for all paralogs in the pgPEN library. Dashed line indicates $\log_2(\text{TPM}) = 1$, the cutoff used for expression.

F) Density histogram of CRISPR scores for single KO pgRNA targets, grouped by target gene expression in HeLa cells. Dashed lines indicate the median CRISPR score for each group.

G) Density plot of target-level CRISPR scores for DKO pgRNAs grouped by whether zero, one, or both target genes are expressed in HeLa cells. Dashed lines indicate the median CRISPR score for each group.

H) Scatter plot of HeLa cell $\log_2$(gene copy number) for paralogs in the pgPEN library. Red dashed line indicates the mean HeLa cell gene copy number of pgPEN library paralogs. Copy number data was obtained from DepMap.

I) Box and overlaid dot plot of CRISPR scores for single KO paralog target genes, grouped by HeLa cell gene copy number. Box plots indicate median ± IQR for each group.

J) Scatter plot of target-level observed versus expected CRISPR scores in the HeLa screen. The solid line is the linear regression line for the negative control (single KO) pgRNAs, while dashed lines indicate ± 2 residuals.

Chapter 3. PGMAP: A PIPELINE TO ENABLE GUIDE RNA READ MAPPING FROM DUAL-TARGETING CRISPR SCREENS

A version of this chapter is being prepared for prepared for submission to a peer-reviewed journal.

A similar version has been posted on arXiv as a preprint:

Parrish, P. C. R., Groso, D. J., Thomas, J. D., Bradley, R. K., and Berger, A. H. (2023). pgMAP: a pipeline to enable guide RNA read mapping from dual-targeting CRISPR screens. In *arXiv [q-bio.GN]*. arXiv. http://arxiv.org/abs/2306.00944

I led this work and my contributions to this paper included building a reproducible and scalable Snakemake pipeline, processing next-generation sequencing data, improving efficiency of code shared by collaborators, writing custom R and Python scripts, and writing the manuscript.

Motivation:

Multiple groups have recently developed dual-targeting CRISPR screen methods for genetic interaction mapping in human cells. But most of these methods use custom sequencing strategies and analysis code, limiting reproducibility and accessibility for other users. We thus need user-friendly software tools to enable reproducible identification of gene pairs required for cell growth.

Results:

We developed pgMAP, an analysis pipeline to map gRNA sequencing reads from dual-targeting CRISPR screens. pgMAP output includes a dual gRNA read counts table and quality control metrics including the proportion of correctly-paired reads and CRISPR library sequencing coverage across all time points and samples.

Availability and Implementation:

pgMAP is implemented using Snakemake and is available open-source under the MIT license at https://github.com/FredHutch/pgMAP_pipeline.

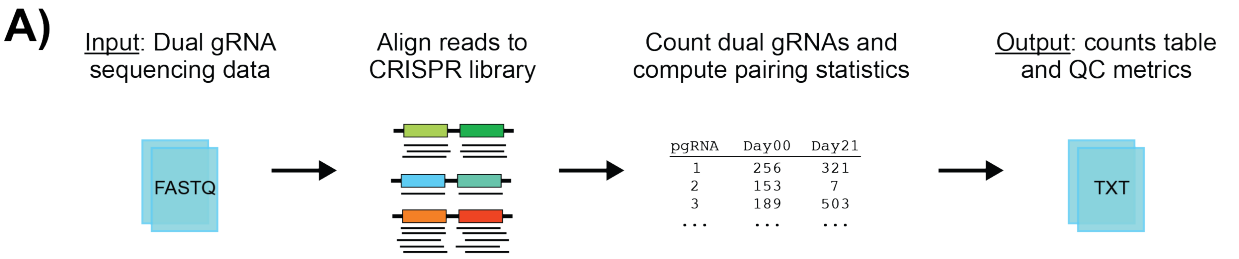
3.1 INTRODUCTION

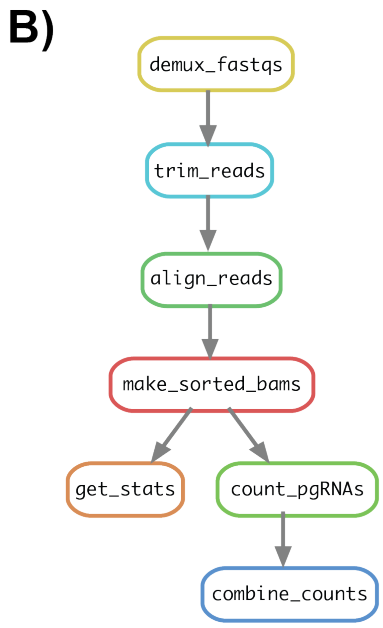
Cancer therapies that target synthetic lethal interactions have the potential to expand treatment options for patients. In recent years, several genome-scale, dual-targeting CRISPR-mediated knockout (KO) screening approaches have been developed to map genetic interactions (GIs) in the human genome (Dede, McLaughlin, et al., 2020; Gonatopoulos-Pournatzis et al., 2020; Ito et al., 2021; Köferle et al., 2022; Parrish et al., 2021; Tang et al., 2022; Thompson et al., 2021). These methods enable functional profiling of duplicated gene families and expand the range of potentially targetable synthetic lethal interactions in cancer (Dandage & Landry, 2021; Ryan et al., 2023). However, computational methods for GI mapping from human CRISPR screen data are poorly established, which may impede interpretation of dual-targeting CRISPR screen data and thus prevent identification of actionable synthetic lethal targets.

Commonly-used tools for analyzing single-targeting CRISPR KO screen data, e.g. MAGeCK (W. Li et al., 2014) and BAGEL (Hart & Moffat, 2016), have limited utility for multi-targeting CRISPR screens. These tools can be used to quantify the fitness effects of each dual gRNA construct once sequencing reads have been mapped to a reference. However, neither MAGeCK nor BAGEL can be used to directly map sequencing reads generated by dual-targeting CRISPR approaches, nor can they map GIs since they were designed to be used only with single-targeting CRISPR approaches.

Some software packages to analyze sequencing data from dual-targeting CRISPR screens have been developed (Ward et al., 2021; Zamanighomi et al., 2019). However, these methods have some drawbacks. One method requires a counts table as input, meaning users must use custom code to map their dual gRNA sequencing reads prior to using the tool (Zamanighomi et al., 2019). Moreover, since both existing tools are implemented as R packages, their dependencies are not

version-controlled and individual scripts must be run one-by-one, which limits scalability, reproducibility, and ease-of-use for users with limited computational experience.





C)

Sample name	# correctly paired reads	% correctly paired reads	Library coverage	Runtime (mins.)
Plasmid	38,471,205	78.84	1,160X	120
Day 5 (pooled)	32,028,290	70.21	965X	85
Day 22, Rep A	29,144,511	72.17	879X	78
Day 22, Rep B	36,628,141	72.28	1,104X	95
Day 22, Rep C	36,448,291	71.56	1,099X	120

Figure 3.1: pgMAP reproducibly and scalably maps sequencing reads from dual gRNA CRISPR screens.

A) Illustrated overview of the pgMAP pipeline.
B) Key Snakemake rules used to align sequencing reads and count dual gRNAs.
C) Sample QC metrics for dual gRNA pairing and library coverage from the HeLa-iCas9 pgPEN CRISPR screen.

Here, we present pgMAP (paired guide RNA Mapper), a Snakemake-based analysis pipeline to map sequencing reads and generate a dual gRNA counts table from dual-targeting CRISPR-mediated KO screening data (**Figure 3.1A**). This read-mapping strategy was originally designed to analyze data from the pgPEN (paired guide RNAs for paralog genetic interaction

mapping) screening method (Parrish et al., 2021). pgMAP represents several improvements to the code base, including the use of the Snakmake workflow manager (Mölder et al., 2021) and Conda environments (*Anaconda Software Distribution*, 2016) to ensure scalable and reproducible analysis. pgMAP also calculates and outputs summary statistics to enable quality control analysis. pgMAP will allow users to apply pgPEN to model systems of interest and enable discovery of genetic interactions across many organisms, genetic backgrounds, tissues, and cancer types.

3.2 Materials and Methods

pgMAP computes dual gRNA counts and quality control (QC) statistics from reads generated by the pgPEN Illumina sequencing strategy in an automated fashion. Reads are trimmed with FastX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/) and demultiplexed using idemp (https://github.com/yhwu/idemp). Next, demultiplexed reads are mapped to the pgPEN library reference with Bowtie v1.2.2 (Langmead et al., 2009), generating files of alignments in Sequence Alignment/MAP (SAM) format. The SAM files are sorted and converted to Binary Alignment/Map (BAM) format using SAMtools (H. Li et al., 2009).

In the last steps of the pipeline, dual gRNAs present in each sample are counted and the percentage of correctly-paired dual gRNAs is computed using `<counter_efficient.R>`, an R script (https://www.R-project.org) that incorporates the libraries Tidyverse v1.2.1 (Wickham et al., 2019) and Rsamtools v1.34.1 via Bioconductor (Huber et al., 2015). The final step of the pipeline runs a Python script `<combine_counts.py>` to generate a file that contains dual gRNA counts for each time point, replicate, and condition in tab-separated values (TSV) format.

3.3 USAGE AND EXAMPLES

To begin, clone the repository from https://github.com/FredHutch/pgMAP_pipeline into an empty folder on a high-performance computing (HPC) cluster. pgMAP is run with Snakemake and requires an up-to-date version of the Mamba package manager (https://github.com/mamba-org/mamba) and an associated Conda environment containing Snakemake and its dependencies. The Conda environment for running pgMAP will be automatically deployed by the Bash script `<run_snakemake.sh>`, and can be found via the following path: `<workflow/envs/snakemake.yaml>`. Additional instructions and links for installation can be found in the file `<README.md>`, located in the base folder or on the homepage of the GitHub website.

The config folder contains sample versions of the configuration files required to run the Snakemake pipeline, which can also be used to analyze the tutorial dataset included in the pgMAP package (tutorial data is located in the folder `<input/tutorial-fastqs/>`). Information on running the tutorial and analyzing sequencing data from pgPEN CRISPR screens can be found in the file `<README.md>`. For more information on the pgPEN sequencing method, see Parrish et al., 2021, Figure S1A.

pgMAP is run on an interactive HPC node via the Bash script `<run_snakemake.sh>`, which activates the Conda environment and executes the pipeline:

```
snakemake --snakefile "workflow/Snakefile" \
  --use-conda \
  --conda-prefix "~/tmp/" \
  --conda-frontend mamba \
  -k -p --reason \
  --jobs 50 --latency-wait 180
```

Upon successful execution of the pgMAP pipeline, results sub-directories containing the output from each Snakemake rule will be generated (see **Figure 3.1B)**. Dual gRNA count tables

will be found in the <results/pgRNA_counts/> directory. Reports containing information about pipeline runtime and graphs of the Snakemake rules that were run will be generated in the <workflow/reports/> directory. Example QC output from the pgPEN HeLa screen is shown in **Figure 3.1C**. This output includes per-sample statistics about dual gRNA pairing rates, runtime, and the fold-coverage of the reference dual gRNA CRISPR library based on the number of correctly-paired reads and the input library size.

Additional example results from the HeLa-iCas9 pgPEN screen are included as supplemental tables. **Table 3.1** shows statistics about the de-multiplexing step, where the FastQ file containing reads from all timepoints and replicates is split up based on sample barcodes. **Table 3.2** shows output from the alignment step, including the number and percentage of reads that successfully aligned to the reference and those that failed to align.

pgMAP can also be executed on an HPC cluster via a job scheduler such as Slurm, where jobs are submitted via the <sbatch> command, or Moab/Torque, where jobs are submitted via the <qsub> command. For this approach, the commands used to execute pgMAP will depend on the workload manager used by the user's HPC cluster. An example script for submitting to Slurm is included in the development branch of the repository as <run_snakemake_cluster.sh>:

```
snakemake --snakefile "workflow/Snakefile" \
  --use-conda \
  --conda-prefix "~/tmp/" \
  --conda-frontend mamba \
  -k -p --reason \
  --jobs 50 --latency-wait 180 --restart-times 3 \
  --cluster-config config/cluster_slurm.yaml \
  --cluster "sbatch -p {cluster.partition} --mem={cluster.mem} -t
   {cluster.time} -c {cluster.ncpus} -n {cluster.ntasks} -o
   {cluster.output}"
```

More information about running Snakemake pipelines via HPC cluster schedulers can be found in the documentation: https://snakemake.readthedocs.io/en/stable/executing/cluster.html. Note that it

may be necessary to set up a configuration profile to save your job submission settings. An example Slurm file can be found at the following path: `<config/cluster_slurm.yaml>`.

3.4 CONCLUSION

We developed pgMAP, an analysis pipeline for aligning and counting Illumina sequencing reads from dual-targeting CRISPR screens to produce a dual gRNA counts table and quality control metrics. The pgMAP pipeline is designed to be implemented by computational scientists as well as bench researchers who are familiar with Linux-based command line interfaces and/or HPC environments and the Python programming language. pgMAP enables researchers with all levels of computational skills to apply dual-targeting CRISPR screening approaches to their own model systems of interest, offering novel insights into human genetic interactions on a genome scale.

3.5 ACKNOWLEDGMENTS

The authors thank April Lo, Mitchell Vollger, and Robin Aguilar for advice on developing and implementing Snakemake pipelines.

3.5.1 *Funding*

This work was funded with support from the Lung Cancer Research Foundation and an American Cancer Society Research Scholar Grant RSG-21-09-01-ET. P.C.R.P. was supported by NSF DGE-1762114. D.J.G was supported in part by the NIH/NCI (R37 CA252050). A.H.B. was supported in part by the NIH/NCI (R00 CA197762 and R37 CA252050); the Devereaux

Outstanding Investigator Award from the Prevent Cancer Foundation; the Stephen H. Petersdorf Lung Cancer Research Award; the Seattle Translational Tumor Research program; and the Innovators Network Endowed Chair. R.K.B. was supported in part by the NIH/NHLBI (R01 HL128239 and R01 HL151651); NIH/NCI (R01 CA251138); Edward P. Evans Foundation; Blood Cancer Discoveries Grant program through the Leukemia & Lymphoma Society, Mark Foundation For Cancer Research, and Paul G. Allen Frontiers Group (8023-20); Department of Defense Breast Cancer Research Program (W81XWH-20-1-0596); and the McIlwain Family Endowed Chair in Data Science. R.K.B. is a Scholar of The Leukemia & Lymphoma Society (1344-18). Sequencing was performed by the Fred Hutch Genomics Shared Resource (supported by NIH/NCI Cancer Center Support Grant P30 CA015704). Computational studies were supported in part by FHCRC's Scientific Computing Infrastructure (ORIP S10 OD028685).

3.5.2 *Data availability*

Raw and processed CRISPR screen sequencing data for the pgPEN CRISPR screens (GEO: GSE178179) is publicly available via GEO. All other data analyzed in this paper is available from the lead contact upon request.

3.5.3 *Conflicts of interest*

The authors declare no competing interests.

3.6 SUPPLEMENTAL DATA

3.6.1 *Supplemental Tables*

Table 3.1: Sample output from pgPEN de-multiplexing step. The number and percentage of Illumina sequencing reads assigned to each barcode is shown.

Sample name	Number of reads	Percent of total reads
Plasmid	66,312,306	23.53%
Day 5 (pooled)	52,384,943	18.59%
Day 22, Rep A	45,532,803	16.16%
Day 22, Rep B	56,415,566	20.02%
Day 22, Rep C	57,436,578	20.38%
Unassigned	3,765,019	1.34%
Total	**281,847,215**	**100%**

Table 3.2: Sample quality control output from pgPEN alignment step.

gRNA number	Sample name	Number of reads processed	Number of reads with ≥1 alignment	Percent of reads with ≥1 alignment	Number of reads failing to align	Percent of reads failing to align	Number of alignments
1	Plasmid	52,384,943	51,166,948	97.67%	1,217,995	2.33%	342,746,179
1	Day 5 (pooled)	45,532,803	44,366,837	97.44%	1,165,966	2.56%	305,262,688
1	Day 22, Rep A	56,415,566	55,120,573	97.70%	1,294,993	2.30%	379,282,821
1	Day 22, Rep B	57,436,578	56,035,123	97.56%	1,401,455	2.44%	386,262,446
1	Day 22, Rep C	66,312,306	63,633,494	95.96%	2,678,812	4.04%	427,572,077
2	Plasmid	10,999,712	10,800,478	98.19%	199,234	1.81%	72,494,492
2	Day 5 (pooled)	52,384,943	46,308,031	88.40%	6,076,912	11.60%	314,079,728
2	Day 22, Rep A	45,532,803	41,118,664	90.31%	4,414,139	9.69%	287,211,825
2	Day 22, Rep B	56,415,566	51,530,773	91.34%	4,884,793	8.66%	360,110,895
2	Day 22, Rep C	57,436,578	51,812,303	90.21%	5,624,275	9.79%	362,482,993

3.6.2 *Genetic interaction mapping pipeline*

The goal of pgMAP was to generate counts tables that can be used to identify essential gene pairs via genetic interaction mapping. With the counts table in hand, users can thus use existing R packages for genetic interaction mapping (Ward et al., 2021; Zamanighomi et al., 2019). However, these tools have some drawbacks: since they are implemented as R packages without clearly-defined environments of software dependencies, the results from these tools may not be reproducible. Additionally, they require users to be familiar with R and Rstudio for statistical computing and to run individual analysis steps in an interactive fashion.

Thus, I also developed a second Snakemake pipeline for GI mapping using the pgPEN approach. This software is still in development and can be accessed via GitHub: https://github.com/FredHutch/GI_mapping. Though pgMAP is currently limited to input data generated via the pgPEN sequencing approach, the GI mapping pipeline is designed to be more broadly applicable: users can input a dual gRNA counts table from any sequencing strategy to identify statistically significant genetic interactions. This pipeline performs quality control analyses, quantifies the growth effects of genetic perturbations and drug treatments, and identifies synthetic lethal interactions. Together, these two software tools will empower scientists of all computational skill levels to uncover targetable synthetic lethal interactions across any cell line, cancer type, or genetic background.

The relevant data are located on the Berger lab fast drive at the following path:

```
center/fh/fast/berger_a/grp/bergerlab_shared/Projects/paralog_pgRNA/pgPEN_lib
   rary/GI_mapping
```

Chapter 4. FINDING DRUGGABLE GENE FAMILIES

In this final part of my dissertation work, I determined if viability of cancer cells with loss of one paralog gene is significantly reduced when treated with a drug targeting one or more other genes in that paralog family. If loss of only one paralog family-member sensitizes cancer cells to a drug that targets the whole paralog family, that drug could still be used to kill cancer cells while avoiding toxicity to healthy cells. I analyzed an existing dataset generated by treating 578 cancer cell lines with over 4,500 drugs. I focused on 367 gene families that were predicted to be synthetic lethal via a recent computational approach.

To identify druggable paralog dependencies, I used a two-tiered approach: first, I applied a discovery-focused linear regression analysis to identify cases where low expression of one paralog was significantly associated with lower viability when cancer cells were treated with an inhibitor targeting that paralog family. This first analysis identified 68 cases where loss of expression of one family-member was significantly associated with increased sensitivity to drugs targeting that gene family (FDR < 0.05). I then used a more stringent outlier-based analysis to confirm my findings for 13 promising paralog pair/drug combinations, which validated 5 of the hits from my first analysis. These results suggest that drugs targeting multiple paralogs in a synthetic lethal family can be used to selectively harm tumor cells while having limited effects on healthy cell viability. Overall, the results from this computational analysis support the use of paralog synthetic lethal therapies in cancer.

4.1 INTRODUCTION

Two potential concerns arose when we considered the feasibility of implementing our paralog synthetic lethal therapy approach in the clinic. First, pgPEN and other CRISPR screening methods rely on genetic inactivation as a proxy for drug activity. However, this does not fully mimic the activity of drugs that are used to treat cancer patients in the clinic. A second potential drawback to targeting synthetic lethal paralogs is that duplicated genes often retain similar sequences, so drugs designed to inactivate one paralog may also affect other members of the same gene family. This means that a synthetic lethal therapy used to treat tumors with loss of one paralog family-member could have off-target effects in healthy cells due nonspecific binding of the paralog-targeting drug to other related genes. I therefore developed a computational approach to ensure that paralog-targeting drugs can be used effectively to kill cancer cells without harming normal cells.

The goal of this part of my thesis work was to determine whether the synthetic lethal paralog families identified via pgPEN or other screening approaches could be effectively targeted using small molecules rather than genetic knockouts. We hypothesized that loss of expression of one paralog would confer sensitivity to inhibitors targeting other family-members due to increased dependency on the remaining genes (**Figure 4.1A**). To validate this hypothesis and to narrow down our list of potential cancer targets, we decided to take a chemical biology-based approach to look for evidence that targeting paralog families with a drug has negative effects on cancer cell viability.

Specifically, we wanted to look at drug sensitivity and gene expression data from cancer cell lines to determine whether we could find cases where loss of a single paralog led to decreased cell viability upon treatment with a drug targeting other members of that paralog family (**Figure 4.1B**). In other words, we asked: if two or more paralogs in a gene family are essential together

(or synthetic lethal) and one of them is lost, does the cell become more dependent on the remaining family-members? The results of this analysis will shed light on whether paralog synthetic lethal targeting via small molecules is a viable approach.

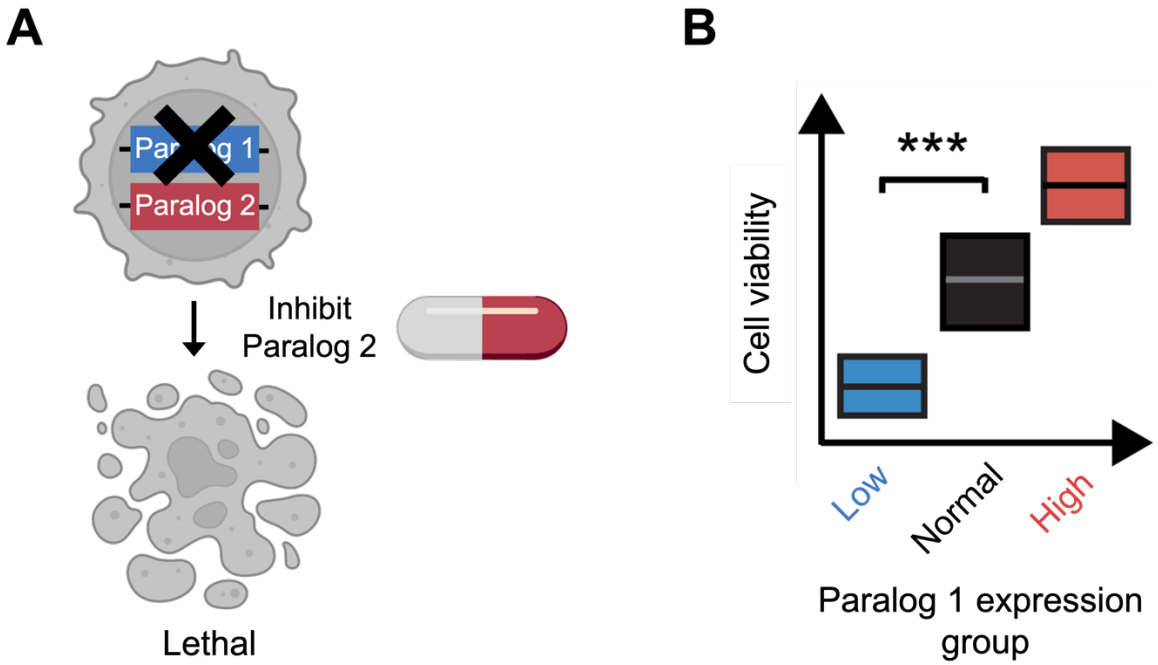


Figure 4.1: Overview of paralog druggability analysis.

A) Illustration of using paralog-targeting drugs to treat cancer cells.

B) Hypothetical results from paralog druggability analysis. Figure created with BioRender.com.

To answer this question, we leveraged existing datasets to find cancer dependencies. An overview of our analysis can be found in **Figure 4.2**. We used the Broad Institute's Cancer Dependency Map (DepMap) project, a publicly-available dataset of CRISPR screens, gene expression, mutation and copy number profiling, and other omics data for over 1,000 cancer cell lines (Corsello et al., 2017, 2020; Ghandi et al., 2019; Pacini et al., 2021; Tsherniak et al., 2017). For this analysis, I used the PRISM (Profiling Relative Inhibition Simultaneously in Mixtures) drug repurposing dataset where 4,518 existing small molecules were screened in 578 cancer cell lines (Corsello et al., 2017, 2020). In a separate study, gene expression levels from the same 578

cell lines were profiled using RNA sequencing (RNAseq) (Tsherniak et al., 2017). I used these data together to address the question: does loss of expression of any individual paralog family-member confer drug sensitivity to paralog inhibitors?

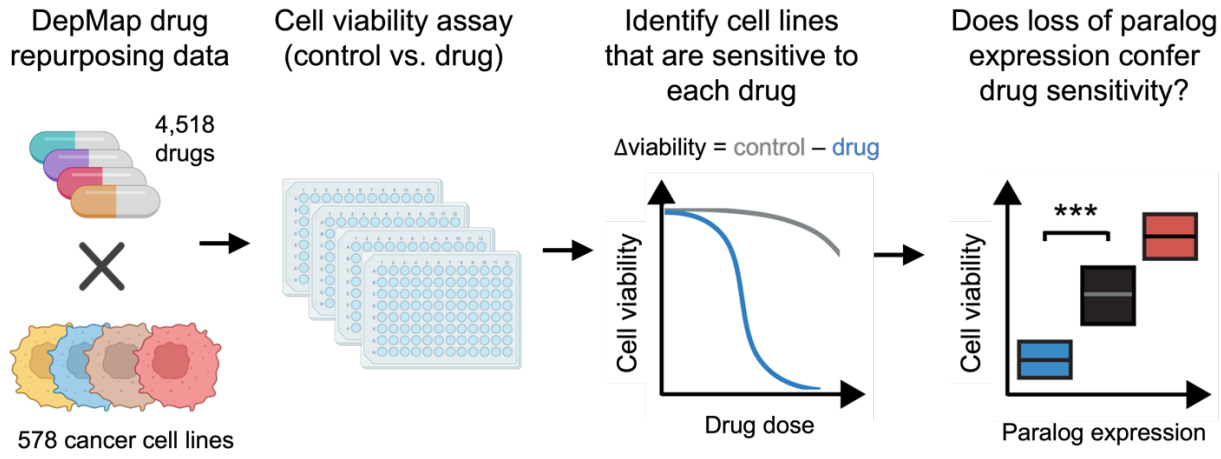


Figure 4.2: Overview of PRISM screening approach and paralog druggability analysis. Figure created with BioRender.com.

4.2 Materials and Methods

4.2.1 *Code and package versions*

All code used in this section can be found in the following GitHub repository: https://github.com/FredHutch/Paralog_druggability. Software and package versions can be found in **Table 4.1**.

Table 4.1: Software and package versions used for paralog druggability analysis.

Name	Version	Link
R	4.1.0	https://www.r-project.org/
Rstudio	1.4.1717	https://posit.co/download/rstudio-desktop/
Tidyverse	1.3.1	https://tidyverse.tidyverse.org/
ggrepel	0.9.1	https://github.com/slowkow/ggrepel
ggforce	0.3.3	https://ggforce.data-imaginist.com/
ggsignif	0.6.3	https://const-ae.github.io/ggsignif/

Rcolorbrewer	1.1-3	https://cran.rstudio.com/package=RColorBrewer
ggpubr	0.4.0	https://rpkgs.datanovia.com/ggpubr/
ggprism	1.0.3	https://csdaw.github.io/ggprism/
scales	1.2.0	https://scales.r-lib.org/
ggtext	0.1.1	https://wilkelab.org/ggtext/
DescTools	0.99.44	https://andrisignorell.github.io/DescTools/
biomaRt	2.48.3	https://www.bioconductor.org/packages/release/bioc/html/biomaRt.html
depmap	1.6.0	https://bioconductor.org/packages/release/data/experiment/html/depmap.html
ExperimentHub	2.0.0	https://www.bioconductor.org/packages/release/bioc/html/ExperimentHub.html

4.2.2 *Linear regression analysis*

For each inhibitor in the PRISM dataset with one or more known gene targets, I compared the gene expression of each target to the viability of DepMap cell lines when treated with the drug-of-interest. The RNAseq gene expression levels for each target gene were measured in transcripts per million (TPM) and $\log_2$-transformed with a pseudocount of 1 as described below:

$$\log_2(\text{TPM} + 1)$$

Cell viability was quantified as the $\log_2$-transformed ratio of cell counts when treated with an inhibitor of interest versus under DMSO control treatment conditions.

I plotted gene expression values versus cell viability across all cancer cell lines in the PRISM dataset and fit a linear model using the `<lm()>` function in R. For each drug-gene comparison, I extracted the linear regression r^2 and p-value and adjusted p-values for multiple testing across all comparisons via the Benjamini-Hochberg method (Benjamini & Hochberg, 1995) using the `<p.adjust()>` R function with the `<method="BH">` argument specified. Additionally, I extracted the slope of the fit line for each linear model to determine whether inhibitor treatment

was associated with increased or decreased cell viability across the range of target gene expression seen in PRISM cell lines.

4.2.3 *Robustly scaled outlier analysis*

I robustly scaled the DepMap gene expression data, quantified as $\log_2$(TPM), using the `<RobScale()>` R function with arguments `<center="TRUE">` and `<scale="TRUE">`. I grouped the cell lines based on the following criteria: the "low" expression group had a scaled $\log_2$(TPM) < -1.5, the "high" expression group had a scaled $\log_2$(TPM) > 1.5, and the "neither" group contained cell lines with all other $\log_2$(TPM) values. In other words, the "neither" group contained cell lines with relatively normal expression of the paralog of interest. The z-scaled $\log_2$(TPM) cutoff of 1.5 represents 1.5 times the interquartile range of gene expression for the paralog of interest across all DepMap cell lines.

To determine which paralog-targeting drugs had a significant effect on the viability of cell lines with low expression of each paralog family-member, I used a Wilcoxon rank-sum test (WRST, also known as a Mann-Whitney U test) to compare the median viability of the low versus normal expression groups. I then adjusted the resulting p-values for multiple testing using the Benjamini-Hochberg method as described previously (**Section 4.2.2**).

4.2.4 *Filtering for druggable, synthetic lethal paralogs*

To improve our statistical power to find significant paralog vulnerabilities, I limited my search space to a set of gene families that were predicted to be synthetic lethal via a recent computational modeling approach (De Kegel et al., 2021). I filtered for paralogs in the top 5th percentile of predicted synthetic lethal gene families or above. I also filtered the PRISM drug sensitivity dataset to extract cell line viability data for drugs targeting one or more members of our

paralog families of interest. I then combined my filtered drug sensitivity dataset with DepMap gene expression and copy number information for each gene of interest (see script <make_drug_dfs.Rmd>). The final RDS-formatted data file is located at the following path:

```
01_Research/01_Projects/01_Paralog_pgRNA/01_Data/Paralog_SL_meta_analysis/02_
   output_files/d.DeKegel_top5_pairs_drug_sens_expr_annot
```

4.3 RESULTS

We took a two-pronged approach to look for paralog dependencies using the DepMap datasets. First, I executed a discovery-based analysis using linear regression modeling. For each gene of interest, I plotted RNAseq gene expression against cell viability across all cancer cell lines in the PRISM dataset. I then fit a linear model to these data to determine whether there was a significant association between decreased gene expression and decreased cell viability upon treatment with drugs targeting the gene-of-interest. I also examined the RNAseq-based gene expression levels of each gene-of-interest to determine whether there was a meaningful difference in expression between cell lines with "low" gene expression versus "normal" gene expression. Past research suggests that a bimodal gene expression pattern can be useful as a biomarker (Ba-Alawi et al., 2022; Huang et al., 2020). The ability to clearly distinguish between "low" and "normal" expression is thus important for downstream clinical applications of our analysis (see **Figure 4.7**).

For paralog pair/drug combinations that emerged as hits from the linear regression analysis, and for which there seemed to be a meaningful difference in expression levels across cancer cell lines, I confirmed the feasibility of targeting these paralogs using an outlier-based analysis. To do this, I robustly scaled gene expression across PRISM cancer cell lines and grouped cell lines by target gene expression using a cutoff of 1.5X the interquartile range (IQR). I then compared the median viability of "low" versus "normal" (non-outlier) expression groups via WRST. For both

the linear regression- and the outlier-based analyses, I used the Benjamini-Hochberg method to adjust the resulting p-values for multiple testing, yielding a false discovery rate (FDR) value. The FDR represents the proportion of hits that are false positives — thus, FDR < 0.05 means that 5% of hits below this cutoff are actually false positives, FDR < 0.1 means that 10% of hits below this cutoff are false positives, and so on.

4.3.1 *Analysis of a known drug target supports the efficacy our approach*

To determine what a "strong" drug dependency would look like, I first carried out the linear regression and outlier analyses using all genes in the PRISM dataset versus inhibitors targeting those genes. I then focused the *EGFR* gene as proof-of-concept. This analysis provides a baseline expectation for both the range of expression of a known targetable cancer gene, and for the viability effects resulting from treating cancer cells with EGFR inhibitors. Since driver mutations in *EGFR* are activating mutations, we hypothesized that cell lines with higher *EGFR* expression would show lower viability when treated with an EGFR inhibitor (**Figure 4.3A**). Indeed, the data supported our hypothesis: cell lines with higher expression of *EGFR* tended to show decreased viability when treated with EGFR inhibitors such as PD-168393 (r^2 = 0.87 and FDR = 3.98e–10, **Figure 4.3B**).

I then applied the outlier-based approach to confirm the findings of our linear regression analysis. This analysis showed that cell lines in the "low" *EGFR* expression outlier group had significantly higher viability when treated with EGFR inhibitor PD-168393 relative to cell lines with "normal" EGFR expression (WRST p-value = 5.6e–06). As expected, cell lines in the "high" *EGFR* expression outlier group showed significantly lower viability when treated with PD-168393 relative to cell lines in the "normal" *EGFR* expression group (WRST p-value = 1.4e–03).

These findings validated our hypothesis that as *EGFR* gene expression increases, the viability of cells treated with EGFR inhibitors decreases. This result also confirmed that our

analysis method was able to identify known cancer dependencies upon treatment with targeted inhibitors. I thus proceeded to look for paralog dependencies using the same PRISM dataset and analysis approach.

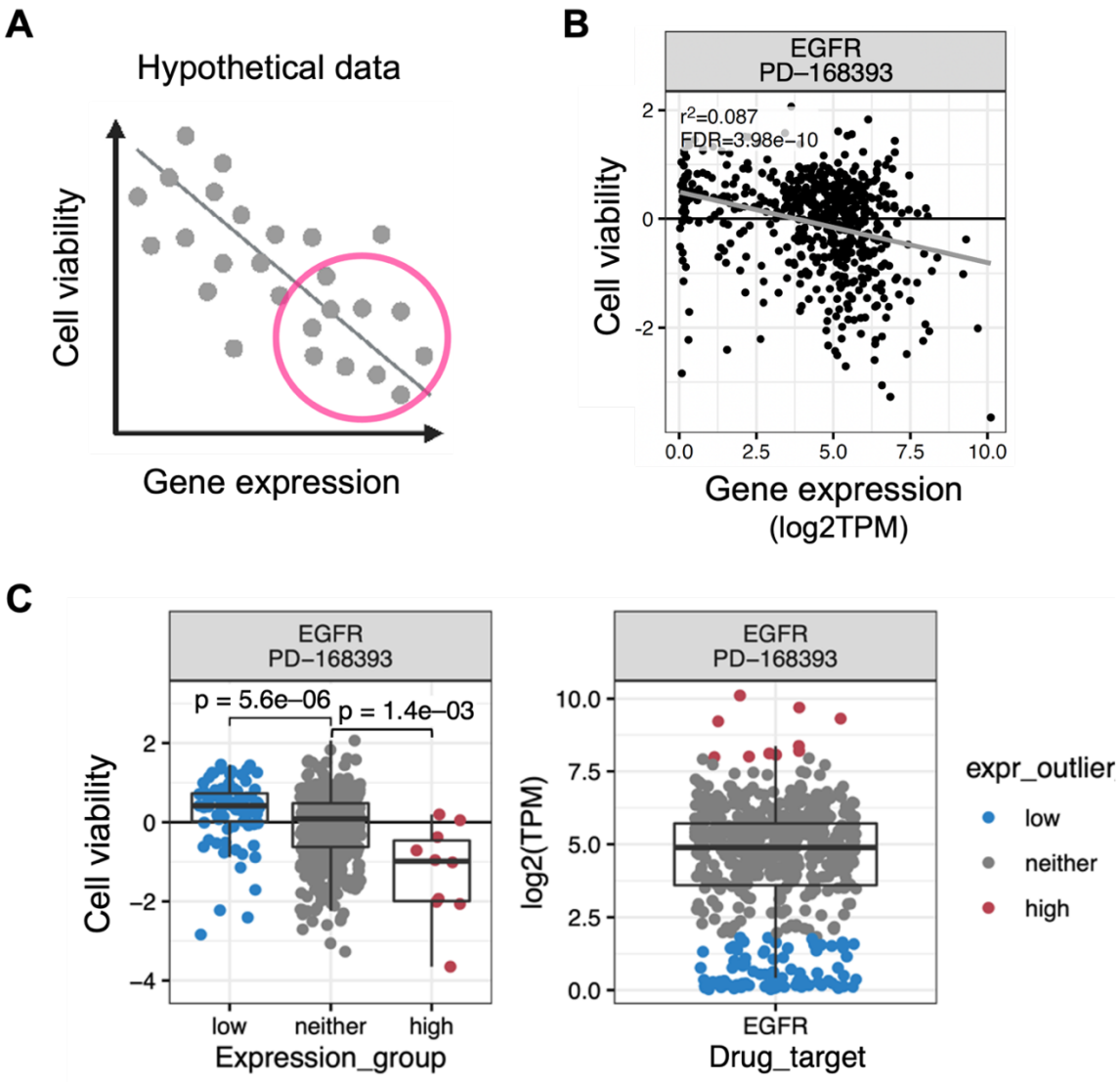


Figure 4.3: Validating PRISM analysis strategy by comparing *EGFR* expression to cell viability after EGFR inhibitor treatment.

A) Illustration of expected dependency results for EGFR. Figure panel created with BioRender.com.

B) Scatterplot and linear regression analysis of *EGFR* expression versus cell viability after treatment with EGFR inhibitor PD-168393. Gene expression is measured as $\log_2$(TPM) and cell is measured as the $\log_2$(fold-change) of cell viability in the EGFRi-treated vs. control (DMSO-treated) groups.

C) Outlier-based analysis results used to validate findings from panel B. Left panel, scatterplot of *EGFR* expression groups versus cell viability. The median cell viability of each gene expression group was compared using a WRST. Right panel, scatterplot of the distribution of *EGFR* gene expression across PRISM cell lines. Cell viability and gene expression were measured as in panel B.

4.3.2 *Discovery-focused analysis of predicted synthetic lethal gene families*

A total of 367 paralog pairs were included in my analysis after filtering for both the top 5% predicted synthetic lethal gene families and for paralogous drug targets included in the PRISM dataset. I applied the linear regression-based analysis approach to look for cases where cell lines with low expression of either gene in the pairs showed increased sensitivity to drugs targeting either or both members of that paralog family. I found 68 significant hits ($r^2 < 0$ and FDR < 0.05, see **Table 4.2**). In 29 cases, the expression of gene A1 affected drug sensitivity and in 39 cases, the expression of gene A2 affected drug sensitivity. Note that "A1" and "A2" are general terms that indicate the (arbitrary) position of each gene in a given paralog pair. Thus, for a gene pair labeled MAP2K1_MAP2K2, A1 refers to *MAP2K1* and A2 refers to *MAP2K2*.

Table 4.2: Summary of hits from paralog linear regression analysis.

Type of dependency	Paralog	Number of hits at FDR < 0.05	Number of hits at FDR < 0.1	Dataframe name
Sensitivity (lm slope > 0)	A1	29	48	`d.DeKegel_top5_pairs_drug_sens_A1_expr_lm_pos_slope`
Sensitivity (lm slope > 0)	A2	39	67	`d.DeKegel_top5_pairs_drug_sens_A2_expr_lm_pos_slope`
Resistance (lm slope < 0)	A1	113	203	`d.DeKegel_top5_pairs_drug_sens_A1_expr_lm_neg_slope`
Resistance (lm slope < 0)	A2	35	61	`d.DeKegel_top5_pairs_drug_sens_A2_expr_lm_neg_slope`

Hits from the linear regression analysis reveal paralog-targeted inhibitors that are associated with decreased viability of cancer cell lines with low expression of a gene family-member. The top hit, GSK2110183 and *AKT1/AKT3*, shows that cell lines with lower expression

of *AKT3* (A2) are significantly more sensitive to the AKT inhibitor GSK2110183 (r^2 = 0.098 and FDR = 4.81e–11, **Figure 4.4A**). The top hit from the analysis of "A1" paralogs demonstrates that cell lines with lower *ATP1A1* (A1) expression are more sensitive to the ATPase inhibitor k-strophanthidin (r^2 = 0.059 and FDR = 3.46e–06, **Figure 4.4B**). Other hits from the linear regression analysis include: three other AKT inhibitors and the *AKT1/AKT2/AKT3* family, bosutinib and *MAP2K1/MAP2K2*, saracatinib and *ABL1/ABL2*, bosutinib and *CDK2/CDK5*, BVT-948 and *PTPN11/PTPN6*, NU6027 and *CDK1/CDK2*, and romidepsin and *HDAC4/HDAC5*. Results for a subset of these top drug/paralog pair combinations are visualized in **Error! Reference source not found.**.

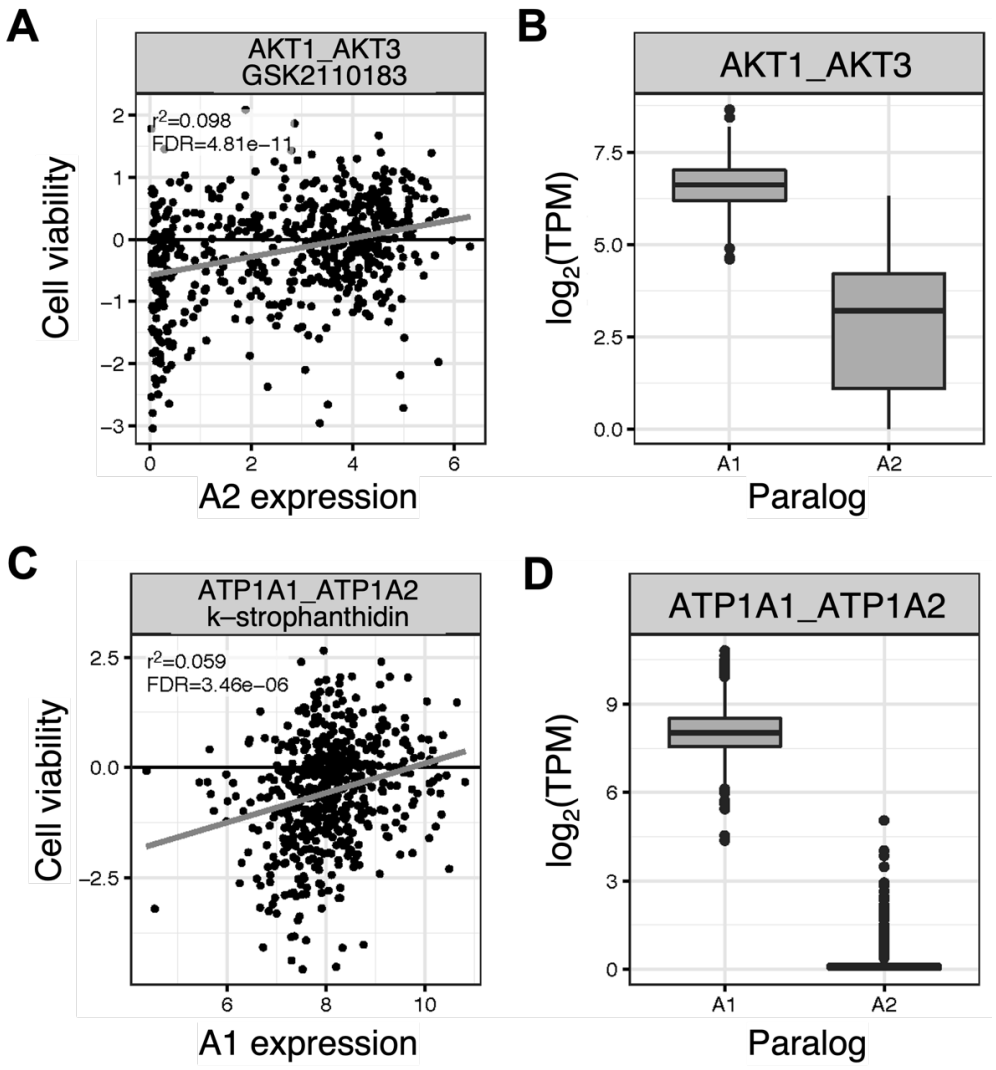


Figure 4.4: Top hits from paralog linear regression analysis.

A) Scatterplot and linear regression analysis of *AKT3* expression versus cell viability after treatment with AKT inhibitor GSK2110183. Gene expression is measured as $\log_2$(TPM) and cell is measured as the $\log_2$(fold-change) of cell viability in the inhibitor-treated vs. control (DMSO-treated) groups.

B) Box-and-whisker plots of gene expression for *AKT1* (referred to as A1) and *AKT3* (referred to as A2) across the PRISM cancer cell lines. Gene expression is measured as $\log_2$(TPM)

C) As in **(A)**, but for *APT1A1* (A1) expression and k-strophanthidin, an ATPase inhibitor.

D) As in **(B)**, but for *ATP1A1* (A1) and *ATP1A2* (A2).

Tab-delimited and R Data Serialization-formatted files containing the full results of this analysis can be found at the following path on the Berger Lab Researcher drive (file names can be found in **Table 4.2**):

```
pparrish/01_Research/01_Projects/01_Paralog_pgRNA/01_Data/Paralog_SL_meta_ana
   lysis/02_output_files/
```

The above folder also contains plots for the top 50 hits from the linear regression analyses:

```
2022-02-02_DeKegel_top5_pairs_drug_sens_A1_expr_lm_pos_slope_fdr_L0.1all.pdf
```

```
2022-02-02_DeKegel_top5_pairs_drug_sens_A2_expr_lm_pos_slope_fdr_L0.1_top50.
   pdf
```

4.3.3 *Outlier analysis confirms efficacy of using AKT inhibitors to target the synthetic lethal AKT gene family*

Though many interesting hits came out of our linear regression analysis, we chose to focus our follow-up efforts on the *AKT1/AKT2/AKT3* gene family. We selected this family first because multiple *AKT* inhibitors showed strong effects on cell viability (**Table 4.3**), suggesting that this was a true effect rather than a false positive. Second, *AKT3* shows both a relatively broad range of expression (0-6 $\log_2$TPM) and a bimodal expression pattern in which a relatively distinct group of cell lines had *AKT3* expression levels of approximately 0 $\log_2$TPM (**Figure 4.4A**). Many other paralogs whose expression levels affect drug sensitivity show a narrower range of expression across cell lines (e.g. the limited IQR seen for *CDK1* in **Figure 4.8D**, Gene A2), or they were expressed constitutively across all cell lines in the PRISM dataset (e.g. the relatively high expression levels seen for *ATP1A3* in **Figure 4.8F**, Gene A1). Given work from other groups suggesting that bimodal gene expression may be a biomarker for drug sensitivity (Ba-Alawi et al., 2022; Huang et al., 2020), we thus chose to focus on the *AKT3*-driven synthetic lethal response to AKT inhibitors.

I applied the outlier-based approach to group cell lines based on their robustly-scaled gene expression values (outliers were considered to be ±1.5X the IQR). This approach validated the effects seen for the *AKT1/AKT2/AKT3* gene family and AKT inhibitors GSK2110183 and AZD5363 (**Figure 4.9**). The top hits from this analysis included multiple cases where low expression of members of the *AKT1/AKT2/AKT3* gene family resulted in increased sensitivity to AKT inhibitors (**Figure 4.8**). The strongest effects were seen for cell lines with low *AKT3* expression when treated with AKT inhibitors GSK2110183 and AZD5363 (**Figure 4.5**)

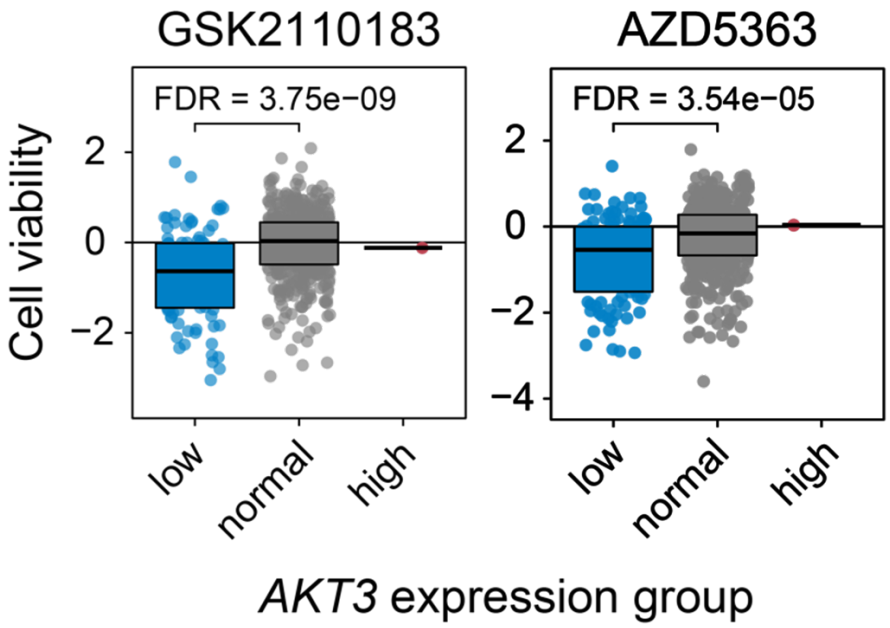


Figure 4.5: Outlier-based analysis supports synthetic lethal targeting of cell lines with *AKT3* loss using AKT inhibitors.

We determined that *AKT3* expression drives this phenotype, because *AKT1* and *AKT2* are constitutively expressed across the DepMap cell lines, with most cell lines showing expression levels of between 5-7 $\log_2$(TPM) for both genes. On the other hand, *AKT3* expression varies more widely across the cell lines, ranging between approximately 0-6 $\log_2$(TPM) (**Figure 4.8A-D**, right-hand panels). Thus *AKT3* expression could serve as a possible biomarker for stratifying patients who might respond to synthetic lethal targeting of the AKT gene family.

To rule out the possibility that this effect could be driven by *AKT3* copy number or mutations in the PI3K/AKT pathway, I analyzed results from copy number profiling and whole-exome sequencing of the DepMap cell lines (Ghandi et al., 2019; Pacini et al., 2021). The copy number analysis confirmed that there was no correlation between low *AKT3* RNA expression and low *AKT3* copy number—indeed, cell lines in the "low" outlier group showed a range of values for $\log_2$(copy number). This suggests that decreased *AKT3* expression levels are not caused by recurrent deletions of *AKT3* across a subset of cancer cell lines (**Figure 4.10**).

Another explanation for our results could be the presence of recurrent mutations in the PI3K/AKT pathway that influence downstream *AKT3* expression. Thus, I analyzed mutation calling results from whole exome sequencing data from the PRISM cancer cell lines. Mutation calling results were grouped into predicted deleterious mutations (**Figure 4.11**) and cancer hotspot mutations seen in The Cancer Genome Atlas (TCGA) data (**Figure 4.12**). Overall, very few of the PRISM cell lines had mutations in *AKT1*, *AKT2*, or *AKT3*. I also did not see a strong association between cell lines with low *AKT3* expression and the presence of mutations in other PI3K/AKT pathway genes. A relatively high number of *AKT3*-low cell lines had TCGA hotspot mutations in *PIK3CA*, and a Fisher's exact test showed a statistically significant association between the presence of TCGA hotspot mutations in *PIK3CA* in PIK3CA and low *AKT3* expression ($p = 3.424e{-}06$). However, two-thirds of the *AKT3*-low cell lines did not have a *PIK3CA* mutation so additional data is needed to fully explain *AKT3* expression loss across cancer cell lines.

A key consideration for translating synthetic lethal therapies to the clinic is whether there is an easily distinguishable biomarker that can be used to stratify patients that are likely to respond to synthetic lethal therapy.

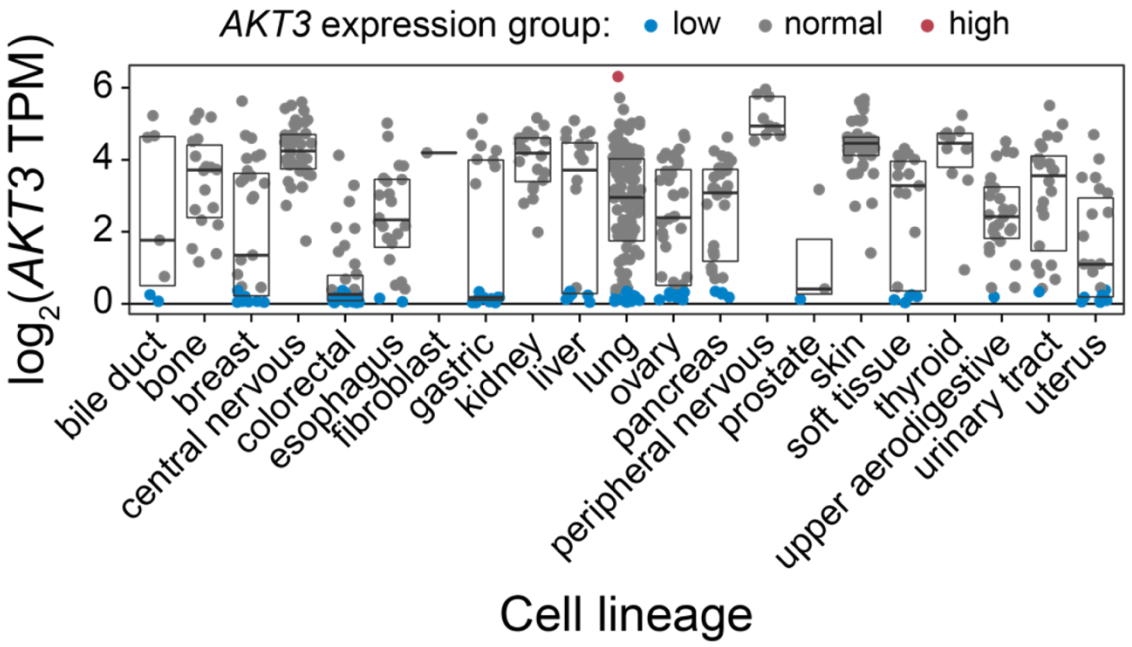


Figure 4.6: *AKT3* shows bimodal expression across cell lineages in the PRISM and DepMap datasets.

I thus analyzed *AKT3* expression patterns across cell lineages in the PRISM dataset and saw that many cell types do show a bimodal expression pattern, with nearly complete *AKT3* expression loss in some cell lines and a relatively large dynamic range of expression. Together, these findings nominate *AKT3* as a target for further study as a potential synthetic lethal cancer target.

4.4 Conclusion

With these data in hand, our major question was: how can we apply our findings about AKT inhibitor sensitivity to cancer therapy? We propose an approach where cancer patients could be stratified based on *AKT3* expression. In patients whose cancer showed *AKT3* expression loss, their tumor cells would be more dependent on the expression of the remaining gene family-members *AKT1* and *AKT2*. Thus, use of an AKT inhibitor in those patients would hopefully lead to tumor regression.

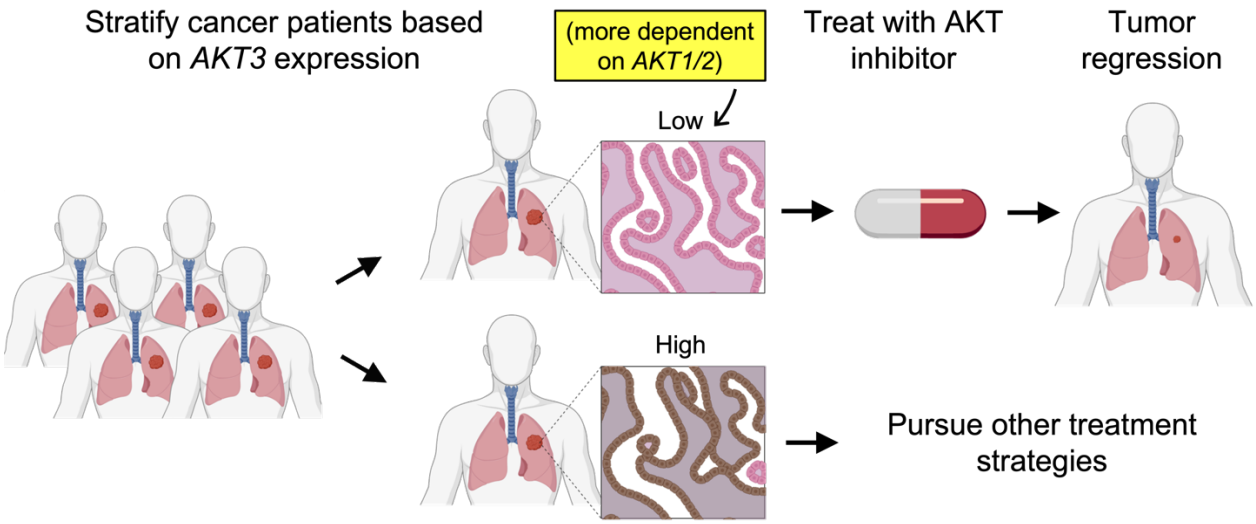


Figure 4.7: Schematic for targeting druggable, synthetic lethal paralogs in cancer. Figure created with BioRender.com.

The *AKT* gene family is known to be important for cancer. The PI3K/AKT pathway is key for survival during cellular stress. Given that tumors have limited access to oxygen and nutrients, cancer cells are often dependent on this pathway and it therefore plays a key role in the development of both primary tumors and acquired resistance to chemotherapy or other cancer drugs. AKT specifically is involved in regulating cell survival and cell cycle progression. There are a number of AKT inhibitors in Phase I and Phase II clinical trials. This means that with further study, AKT inhibitors have the potential for use as effective synthetic lethal cancer drugs in tumors with loss of *AKT3* expression.

4.5 SUPPLEMENTAL DATA

4.5.1 *Code and Results*

All code and results from this analysis are located in the following sub-folder within Phoebe's folder in the Berger Lab Researcher drive:

`01_Research/01_Projects/01_Paralog_pgRNA/01_Data/Paralog_SL_meta_analysis/`

4.5.2 *Supplemental Figures and Tables*

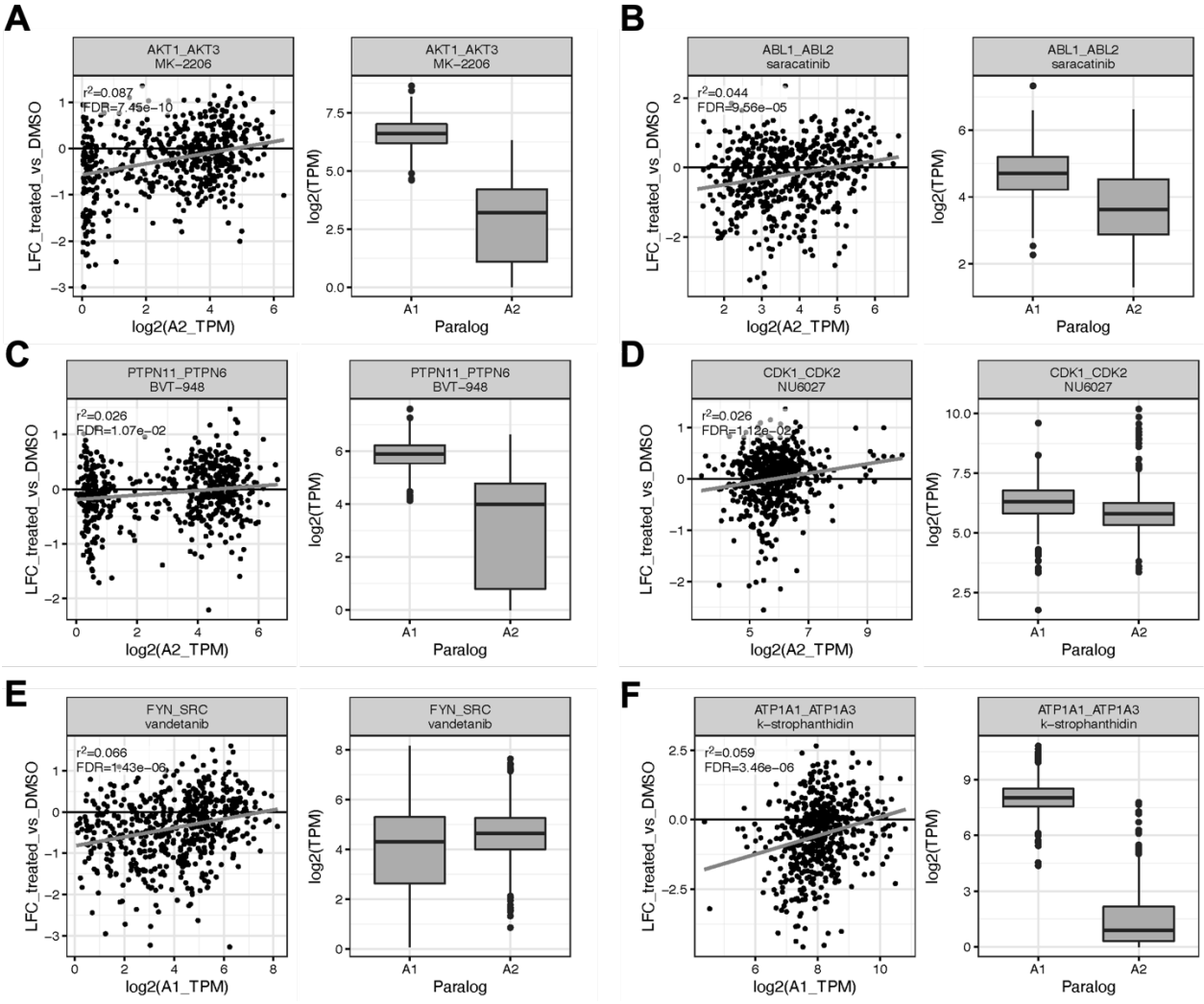


Figure 4.8: Selected hits from linear regression-based paralog drug sensitivity analysis.

A) Left panel, scatterplot of Gene A2 expression versus cell viability in drug treatment for *AKT1/AKT3* and MK-2206. Right panel, Gene A1 and Gene A2 expression across PRISM cell lines for *AKT1/AKT3*. Cell viability is measured as the log2(fold-change) in drug versus control (DMSO) treatment. Gene expression is measured as $\log_2$(TPM).

B) As in **(A)**, but for *ABL1/ABL2* and saracatinib.

C) As in **(A)**, but for *PTPN11/PTPN6* and BVT-948.

D) As in **(A)**, but for *CDK1/CDK2* and NU6027.

E) Left panel, scatterplot of Gene A1 expression versus cell viability in drug treatment for *FYN/SRC* and vandetanib. Right panel, Gene A1 and Gene A2 expression across PRISM cell lines for *FYN/SRC*. Cell viability and gene expression are measured as in **(A)**.

F) As in **(E)**, but for *ATP1A1/ATP1A3* and k-strophanthidin.

Table 4.3: Linear regression summary statistics for AKT inhibitors and the *AKT* gene family.

AKT inhibitor	Paralog pairs	R^2	FDR
GSK2110183	AKT1_AKT3, AKT2_AKT3	0.098	4.81e–11
MK-2206	AKT1_AKT3, AKT2_AKT3	0.087	7.45e–10
GDC-0068	AKT1_AKT3, AKT2_AKT3	0.061	1.27e–06
AZD5363	AKT1_AKT3, AKT2_AKT3	0.047	4.64e–06

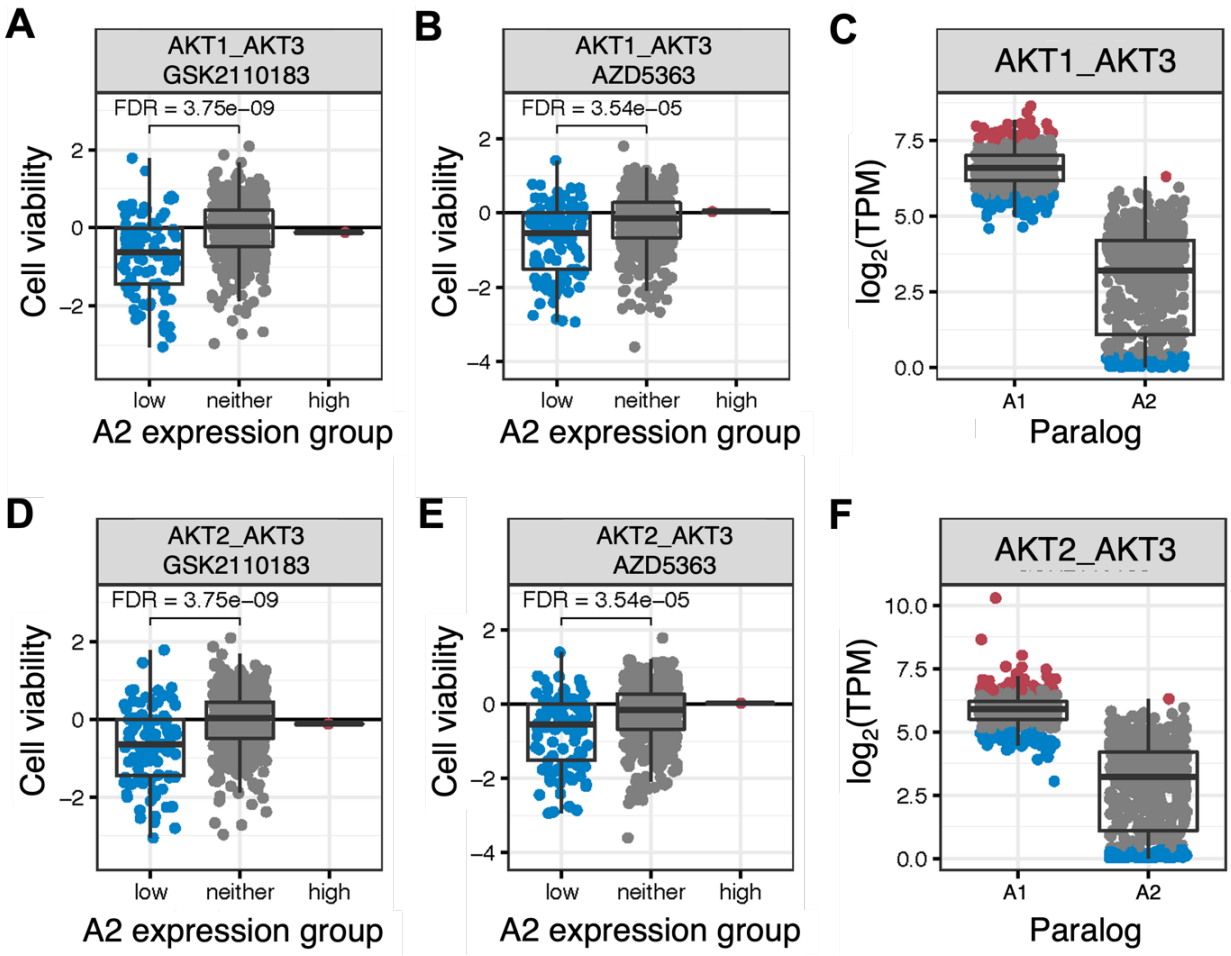


Figure 4.9: Outlier analysis confirms increased effectiveness of AKT inhibitors GSK2110183 and AZD5363 in cell lines with low *AKT3* expression.

A) Grouped scatterplot of *AKT3* (A2) expression versus cell viability in GSK2110183 relative to control (DMSO) treatment for the *AKT1/AKT3* paralog pair. Overlaid box-and-whisker plots indicate group median, IQR, and outliers. WRST with Benjamini-Hochberg FDR correction was used to compare the median of low versus normal groups.

B) As in **(A)**, but for AZD5363.

C) Grouped scatterplot of *AKT1* and *AKT3* expression across PRISM cell lines. Gene expression is measured as $\log_2$(TPM). Colors indicate the expression outlier group of each cell line: blue is low, gray is normal, red is high.

D) As in **(A)**, but for *AKT2/AKT3*.

E) As in **(B)**, but for *AKT2/AKT3*.

F) As in **(C)**, but for *AKT2* and *AKT3*.

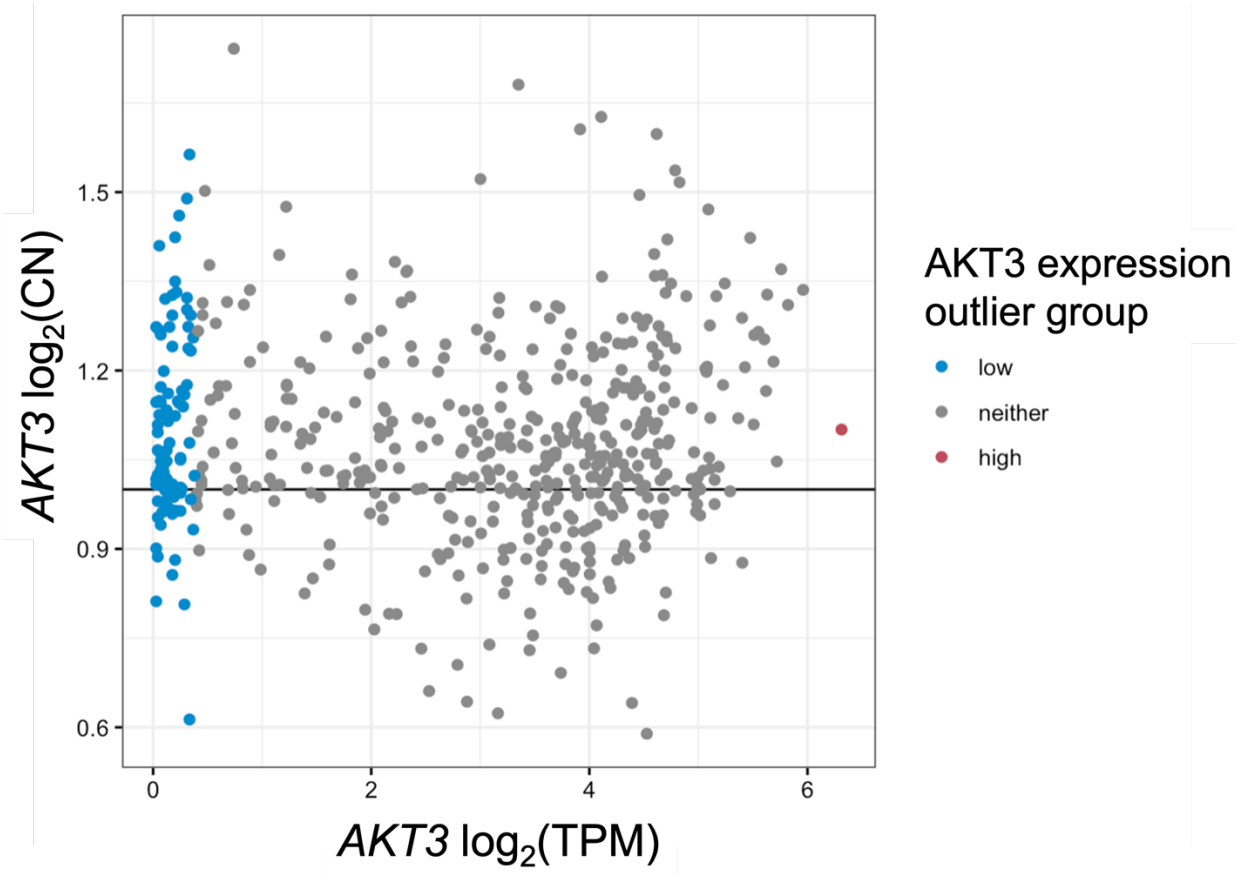


Figure 4.10: Analysis of AKT3 expression versus copy number across DepMap cancer cell lines shows no association between loss of *AKT3* expression and decreased *AKT3* copy number.

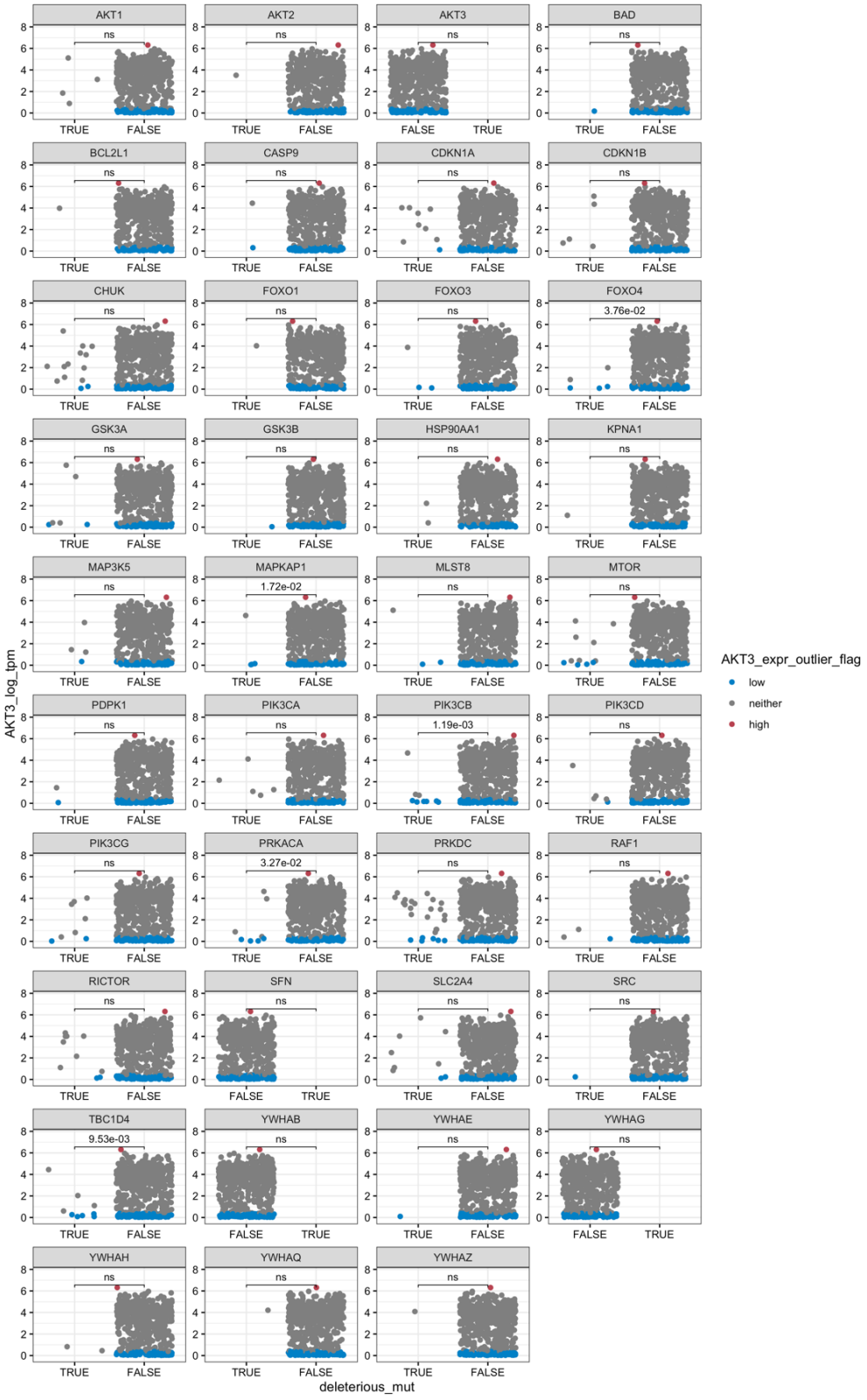


Figure 4.11: The presence of deleterious mutations in PI3K/AKT pathway genes is not strongly associated with *AKT3* gene expression loss (by Fisher's exact test).

Figure 4.12: The presence of TCGA hotspot mutations in PI3K/AKT pathway genes is not strongly associated with *AKT3* gene expression loss (by Fisher's exact test).

Chapter 5. DISCUSSION

5.1 CONCLUSIONS

Altogether, my dissertation research has greatly expanded the set of genes that can be targeted to specifically kill tumor cells without affecting healthy cells. By developing and implementing the pgPEN method, I uncovered over 100 novel, potentially targetable synthetic lethal interactions in human lung and cervical cancer cells. My approach has been cited and adopted by researchers around the world. By comparing my data to other recent paralog synthetic lethal screens, I confirmed that over 75% (n = 96) of pgPEN hits are predicted to be broadly synthetic lethal and nearly 10% (n = 10) of pairs were synthetic lethal in multiple paralog screens. This suggests that many pgPEN hits likely represent penetrant synthetic lethal interactions that could be targeted across tissue types and environmental contexts.

I also applied the skills I gained in my Data Science degree option to facilitate the discovery of synthetic lethal targets for other cancer types. I developed a user-friendly, open-source software package that other researchers can use to map sequencing data from dual-targeting CRISPR screen data. Additionally, I developed a computational analysis approach and applied it to a large-scale data set of drug screening in cancer cell lines. Results from this analysis nominated additional druggable, synthetic lethal paralogs for follow-up study and lend support the use of drugs to target synthetic lethal paralogs in cancer cells.

The pgPEN library was initially the largest human paralog dual-targeting CRISPR library ever developed. Since then, other groups have developed approaches to identify synthetic lethal cancer targets, demonstrating the promise of targeting paralogs for cancer therapy. The synthetic lethal therapeutic targets that I and others have identified can be further tested and ultimately

translated to the clinic, with the goal of improving quality-of-life for cancer patients and ultimately curing a broad spectrum of cancers.

5.2 Broader Implications

Paralog synthetic lethal therapies hold great promise for treating cancer, but we must also consider the broader clinical context when pursuing any new cancer therapeutic approach. The main goal of solving therapeutic challenges in cancer drug discovery is to make better therapies that are more effective at treating disease and less toxic for patients. The identification of paralog synthetic lethal targets is part of a much wider toolkit of cancer therapeutics that are applied in the clinic. Though synthetic lethal paralogs could potentially be targeted as a first-line therapeutic approach, they could also be targeted in combination with chemotherapy, radiation, oncogene-targeted therapy, or immunotherapy to treat primary tumors, metastatic lesions, and tumors that have acquired resistance to other therapies.

I would like to highlight two key considerations regarding the clinical implications of paralog-targeted synthetic lethal therapies. First, context dependency must be considered when pursuing synthetic lethal targets. Hits from *in vitro* human gene essentiality screens can fail to validate *in vivo*, and synthetic lethal interactions may suffer from this to an even larger degree given that they are more likely than single essential genes to be context-dependent. Potential paralog targets should therefore be validated in organoid and animal models. Second, the therapeutic window for targeting duplicated genes is also an important consideration requiring further study and careful experimental analysis.

With these considerations in mind, there are many exciting new technologies for genome editing that can be used to expand the range of targetable alterations in cancer. For instance, though

most of this work focused on the use of CRISPR knockout screens for drug target discovery, many exciting new genome editing approaches are currently being developed to functionally profile mutations such as SNVs and small, in-frame indels. These CRISPR-based approaches include saturation genome editing via homology-directed repair, as well as the use of base editors and prime editors to engineer precise genomic edits in a pooled fashion.

Additionally, though my dissertation work focused mainly on identifying genes that can be targeted by small molecules that are either currently approved or in development for cancer treatment, the definition of "druggability" in cancer is expanding beyond small molecule therapeutics. Chemical biologists have developed exciting new tools such as proteolysis-targeting chimeras (PROTACs) that can be used to degrade target proteins rather than simply inhibiting their function. Moreover, CRISPR gene therapies are in clinical development for *ex vivo* use in treating blood disorders. If these therapies are successful, this could pave the way for the use of CRISPR to treat cancer as well as other diseases.

Lastly, I would like to highlight the fact that the genome editing and genetic interaction mapping approaches I developed have built on decades of work by researchers studying basic science in model organisms. Though scientific research is often divided into "basic" and "clinical" realms, none of the biomedical research we do would be possible without foundational research in the basic sciences. Basic molecular biology research in bacteria led to the development of CRISPR for genome editing, and yeast genetics research laid the groundwork for the development of paralog synthetic lethal therapies. I have been lucky to complete my dissertation research at institutions that value both approaches. I hope researchers of all backgrounds across many scientific disciplines can continue to learn from each other to advance scientific research and improve human health.

BIBLIOGRAPHY

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., Kost-Alimova, M., Gill, S., Xu, H., Ali, L. D., Jiang, G., Pantel, S., Lee, Y., Goodale, A., Cherniack, A. D., … Hahn, W. C. (2016). Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discovery*, *6*(8), 914–929. https://doi.org/10.1158/2159-8290.CD-16-0154

Anaconda Software Distribution (Version 2-2.4.0). (2016). [Anaconda]. https://anaconda.com

Anvar, N. E., Lin, C., Wilson, L. L., Sangree, A. K., Ma, X., Colic, M., Doench, J. G., & Hart, T. (2023). Combined genome-scale fitness and paralog synthetic lethality screens with just 44k clones: the IN4MER CRISPR/Cas12a multiplex knockout platform. In *bioRxiv* (p. 2023.01.03.522655). https://doi.org/10.1101/2023.01.03.522655

Arcila, M. E., Drilon, A., Sylvester, B. E., Lovly, C. M., Borsu, L., Reva, B., Kris, M. G., Solit, D. B., & Ladanyi, M. (2015). MAP2K1 (MEK1) Mutations Define a Distinct Subset of Lung Adenocarcinoma Associated with Smoking. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *21*(8), 1935–1943. https://doi.org/10.1158/1078-0432.CCR-14-2124

Ashworth, A., & Lord, C. J. (2018). Synthetic lethal therapies for cancer: what's next after PARP inhibitors? *Nature Reviews. Clinical Oncology*, *15*(9), 564–576. https://doi.org/10.1038/s41571-018-0055-6

Ba-Alawi, W., Nair, S. K., Li, B., Mammoliti, A., Smirnov, P., Mer, A. S., Penn, L. Z., & Haibe-Kains, B. (2022). Bimodal Gene Expression in Patients with Cancer Provides Interpretable Biomarkers for Drug Sensitivity. *Cancer Research*, *82*(13), 2378–2387. https://doi.org/10.1158/0008-5472.CAN-21-2395

Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M. C., Kim, J., Reardon, B., Ng, P. K.-S., Jeong, K. J., Cao, S., Wang, Z., Gao, J., Gao, Q., Wang, F., Liu, E. M., Mularoni, L., … Ding, L. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*, *173*(2), 371-385.e18. https://doi.org/10.1016/j.cell.2018.02.060

Balatoni, C. E., Dawson, D. W., Suh, J., Sherman, M. H., Sanders, G., Hong, J. S., Frank, M. J., Malone, C. S., Said, J. W., & Teitell, M. A. (2009). Epigenetic silencing of Stk39 in B-cell lymphoma inhibits apoptosis from genotoxic stress. *The American Journal of Pathology*, *175*(4), 1653–1661. https://doi.org/10.2353/ajpath.2009.090091

Baldwin, A., Grueneberg, D. A., Hellner, K., Sawyer, J., Grace, M., Li, W., Harlow, E., & Munger, K. (2010). Kinase requirements in human cells: V. Synthetic lethal interactions between p53 and the protein kinases SGK2 and PAK3. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(28), 12463–12468. https://doi.org/10.1073/pnas.1007462107

Baldwin, A., Li, W., Grace, M., Pearlberg, J., Harlow, E., Münger, K., & Grueneberg, D. A. (2008). Kinase requirements in human cells: II. Genetic interaction screens identify kinase requirements following HPV16 E7 expression in cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(43), 16478–16483. https://doi.org/10.1073/pnas.0806195105

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. V., Sonkin, D., Reddy, A., Liu, M., Murray, L., Berger, M. F., Monahan, J. E., Morais, P., Meltzer, J., Korejwa, A., Jané-Valbuena, J., … Garraway, L. A. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, *483*(7391), 603–607. https://doi.org/10.1038/nature11003

Bartha, I., di Iulio, J., Venter, J. C., & Telenti, A. (2018). Human gene essentiality. *Nature Reviews. Genetics*, *19*(1), 51–62. https://doi.org/10.1038/nrg.2017.75

Baskar, R., Lee, K. A., Yeo, R., & Yeoh, K.-W. (2012). Cancer and radiation therapy: current advances and future directions. *International Journal of Medical Sciences*, *9*(3), 193–199. https://doi.org/10.7150/ijms.3635

Bedard, P. L., Hyman, D. M., Davids, M. S., & Siu, L. L. (2020). Small molecules, big impact: 20 years of targeted therapy in oncology. *The Lancet*, *395*(10229), 1078–1088. https://doi.org/10.1016/S0140-6736(20)30164-1

Behan, F. M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C. M., Migliardi, G., Santos, R., Rao, Y., Sassi, F., Pinnelli, M., Ansari, R., Harper, S., Jackson, D. A., McRae, R., Pooley, R., Wilkinson, P., van der Meer, D., Dow, D., Buser-Doepner, C., … Garnett, M. J. (2019). Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature*, *568*(7753), 511–516. https://doi.org/10.1038/s41586-019-1103-9

Ben-David, U., & Amon, A. (2020). Context is everything: aneuploidy in cancer. *Nature Reviews. Genetics*, *21*(1), 44–62. https://doi.org/10.1038/s41576-019-0171-x

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *57*(1), 289–300. http://www.jstor.org/stable/2346101

Berger, A. C., Korkut, A., Kanchi, R. S., Hegde, A. M., Lenoir, W., Liu, W., Liu, Y., Fan, H., Shen, H., Ravikumar, V., Rao, A., Schultz, A., Li, X., Sumazin, P., Williams, C., Mestdagh, P., Gunaratne, P. H., Yau, C., Bowlby, R., … Akbani, R. (2018). A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. *Cancer Cell*, *33*(4), 690-705.e9. https://doi.org/10.1016/j.ccell.2018.03.014

Berger, A. H., Brooks, A. N., Wu, X., Shrestha, Y., Chouinard, C., Piccioni, F., Bagul, M., Kamburov, A., Imielinski, M., Hogstrom, L., Zhu, C., Yang, X., Pantel, S., Sakai, R., Watson, J., Kaplan, N., Campbell, J. D., Singh, S., Root, D. E., … Boehm, J. S. (2016). High-throughput Phenotyping of Lung Cancer Somatic Mutations. *Cancer Cell*, *30*(2), 214–228. https://doi.org/10.1016/j.ccell.2016.06.022

Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J. S., Dobson, J., Urashima, M., Mc Henry, K. T., Pinchback, R. M., Ligon, A. H., Cho, Y.-J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M. S., … Meyerson, M. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature*, *463*(7283), 899–905. https://doi.org/10.1038/nature08822

Blomen, V. A., Májek, P., Jae, L. T., Bigenzahn, J. W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F. R., Olk, N., Stukalov, A., Marceau, C., Janssen, H., Carette, J. E., Bennett, K. L., Colinge, J., Superti-Furga, G., & Brummelkamp, T. R. (2015). Gene essentiality and synthetic lethality in haploid human cells. *Science*, *350*(6264), 1092–1096. https://doi.org/10.1126/science.aac7557

Boettcher, M., Tian, R., Blau, J. A., Markegard, E., Wagner, R. T., Wu, D., Mo, X., Biton, A., Zaitlen, N., Fu, H., McCormick, F., Kampmann, M., & McManus, M. T. (2018). Dual

gene activation and knockout screen reveals directional dependencies in genetic networks. *Nature Biotechnology*, *36*(2), 170–178. https://doi.org/10.1038/nbt.4062

Bommi-Reddy, A., Almeciga, I., Sawyer, J., Geisen, C., Li, W., Harlow, E., Kaelin, W. G., Jr, & Grueneberg, D. A. (2008). Kinase requirements in human cells: III. Altered kinase requirements in VHL-/- cancer cells detected in a pilot synthetic lethal screen. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(43), 16484–16489. https://doi.org/10.1073/pnas.0806574105

Boone, C., Bussey, H., & Andrews, B. J. (2007). Exploring genetic interactions and networks with yeast. *Nature Reviews. Genetics*, *8*(6), 437–449. https://doi.org/10.1038/nrg2085

Bracci, L., Schiavoni, G., Sistigu, A., & Belardelli, F. (2014). Immune-based mechanisms of cytotoxic chemotherapy: implications for the design of novel and rationale-based combined treatments against cancer. *Cell Death and Differentiation*, *21*(1), 15–25. https://doi.org/10.1038/cdd.2013.67

Brinkman, E. K., Chen, T., Amendola, M., & van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research*, *42*(22), e168. https://doi.org/10.1093/nar/gku936

Bryant, H. E., Schultz, N., Thomas, H. D., Parker, K. M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N. J., & Helleday, T. (2005). Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature*, *434*(7035), 913–917. https://doi.org/10.1038/nature03443

Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., & Stuart, J. M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics*, *45*(10), 1113–1120. https://doi.org/10.1038/ng.2764

Cao, J., Wu, L., Zhang, S.-M., Lu, M., Cheung, W. K. C., Cai, W., Gale, M., Xu, Q., & Yan, Q. (2016). An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic Acids Research*, *44*(19), e149. https://doi.org/10.1093/nar/gkw660

Caplen, N. J., Parrish, S., Imani, F., Fire, A., & Morgan, R. A. (2001). Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(17), 9742–9747. https://doi.org/10.1073/pnas.171251798

Ceccaldi, R., Liu, J. C., Amunugama, R., Hajdu, I., Primack, B., Petalcorin, M. I. R., O'Connor, K. W., Konstantinopoulos, P. A., Elledge, S. J., Boulton, S. J., Yusufzai, T., & D'Andrea, A. D. (2015). Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature*, *518*(7538), 258–262. https://doi.org/10.1038/nature14184

Chalmers, Z. R., Connelly, C. F., Fabrizio, D., Gay, L., Ali, S. M., Ennis, R., Schrock, A., Campbell, B., Shlien, A., Chmielecki, J., Huang, F., He, Y., Sun, J., Tabori, U., Kennedy, M., Lieber, D. S., Roels, S., White, J., Otto, G. A., … Frampton, G. M. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Medicine*, *9*(1), 34. https://doi.org/10.1186/s13073-017-0424-2

Chan, E. M., Shibue, T., McFarland, J. M., Gaeta, B., Ghandi, M., Dumont, N., Gonzalez, A., McPartlan, J. S., Li, T., Zhang, Y., Bin Liu, J., Lazaro, J.-B., Gu, P., Piett, C. G., Apffel, A., Ali, S. O., Deasy, R., Keskula, P., Ng, R. W. S., … Bass, A. J. (2019). WRN helicase is a synthetic lethal target in microsatellite unstable cancers. *Nature*, *568*(7753), 551–556. https://doi.org/10.1038/s41586-019-1102-x

Chandra, R. A., Keane, F. K., Voncken, F. E. M., & Thomas, C. R., Jr. (2021). Contemporary radiotherapy: present and future. *The Lancet*, *398*(10295), 171–184. https://doi.org/10.1016/S0140-6736(21)00233-6

Chargari, C., Magne, N., Guy, J.-B., Rancoule, C., Levy, A., Goodman, K. A., & Deutsch, E. (2016). Optimize and refine therapeutic index in radiation therapy: Overview of a century. *Cancer Treatment Reviews*, *45*, 58–67. https://doi.org/10.1016/j.ctrv.2016.03.001

Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., Sternberg, S. H., Joung, J. K., Yildiz, A., & Doudna, J. A. (2017). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature*, *550*(7676), 407–410. https://doi.org/10.1038/nature24268

Cheng, D. T., Mitchell, T. N., Zehir, A., Shah, R. H., Benayed, R., Syed, A., Chandramohan, R., Liu, Z. Y., Won, H. H., Scott, S. N., Brannon, A. R., O'Reilly, C., Sadowska, J., Casanova, J., Yannes, A., Hechtman, J. F., Yao, J., Song, W., Ross, D. S., … Berger, M. F. (2015). Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *The Journal of Molecular Diagnostics: JMD*, *17*(3), 251–264. https://doi.org/10.1016/j.jmoldx.2014.12.006

Cherniack, A. D., Shen, H., Walter, V., Stewart, C., Murray, B. A., Bowlby, R., Hu, X., Ling, S., Soslow, R. A., Broaddus, R. R., Zuna, R. E., Robertson, G., Laird, P. W., Kucherlapati, R., Mills, G. B., Cancer Genome Atlas Research Network, Weinstein, J. N., Zhang, J., Akbani, R., & Levine, D. A. (2017). Integrated Molecular Characterization of Uterine Carcinosarcoma. *Cancer Cell*, *31*(3), 411–423. https://doi.org/10.1016/j.ccell.2017.02.010

Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., Cole, M. A., Liu, D. R., Joung, J. K., Bauer, D. E., & Pinello, L. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology*, *37*(3), 224–226. https://doi.org/10.1038/s41587-019-0032-3

Collins, S. R., Schuldiner, M., Krogan, N. J., & Weissman, J. S. (2006). A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biology*, *7*(7), R63. https://doi.org/10.1186/gb-2006-7-7-r63

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143

Corsello, S. M., Bittker, J. A., Liu, Z., Gould, J., McCarren, P., Hirschman, J. E., Johnston, S. E., Vrcic, A., Wong, B., Khan, M., Asiedu, J., Narayan, R., Mader, C. C., Subramanian, A., & Golub, T. R. (2017). The Drug Repurposing Hub: a next-generation drug library and information resource. *Nature Medicine*, *23*(4), 405–408. https://doi.org/10.1038/nm.4306

Corsello, S. M., Nagari, R. T., Spangler, R. D., Rossen, J., Kocak, M., Bryan, J. G., Humeidi, R., Peck, D., Wu, X., Tang, A. A., Wang, V. M., Bender, S. A., Lemire, E., Narayan, R., Montgomery, P., Ben-David, U., Garvie, C. W., Chen, Y., Rees, M. G., … Golub, T. R. (2020). Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. *Nature Cancer*, *1*(2), 235–248. https://doi.org/10.1038/s43018-019-0018-6

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., Ding, H., Koh, J. L. Y., Toufighi, K., Mostafavi, S., Prinz, J., St Onge, R. P., VanderSluis, B.,

Makhnevych, T., Vizeacoumar, F. J., Alizadeh, S., Bahr, S., Brost, R. L., Chen, Y., … Boone, C. (2010). The genetic landscape of a cell. *Science*, *327*(5964), 425–431. https://doi.org/10.1126/science.1180823

Costanzo, M., Hou, J., Messier, V., Nelson, J., Rahman, M., VanderSluis, B., Wang, W., Pons, C., Ross, C., Ušaj, M., San Luis, B.-J., Shuteriqi, E., Koch, E. N., Aloy, P., Myers, C. L., Boone, C., & Andrews, B. (2021). Environmental robustness of the global yeast genetic interaction network. *Science*, *372*(6542). https://doi.org/10.1126/science.abf8424

Costanzo, M., VanderSluis, B., Koch, E. N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S. D., Pelechano, V., Styles, E. B., Billmann, M., van Leeuwen, J., van Dyk, N., Lin, Z.-Y., Kuzmin, E., Nelson, J., Piotrowski, J. S., … Boone, C. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science*, *353*(6306). https://doi.org/10.1126/science.aaf1420

Dagogo-Jack, I., & Shaw, A. T. (2018). Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews. Clinical Oncology*, *15*(2), 81–94. https://doi.org/10.1038/nrclinonc.2017.166

Dandage, R., & Landry, C. R. (2019). Paralog dependency indirectly affects the robustness of human cells. *Molecular Systems Biology*, *15*(9), e8871. https://doi.org/10.15252/msb.20198871

Dandage, R., & Landry, C. R. (2021). Identifying features of genome evolution to exploit cancer vulnerabilities [Review of *Identifying features of genome evolution to exploit cancer vulnerabilities*]. *Cell Systems*, *12*(12), 1127–1130. Elsevier. https://doi.org/10.1016/j.cels.2021.08.007

De Kegel, B., Quinn, N., Thompson, N. A., Adams, D. J., & Ryan, C. J. (2021). Comprehensive prediction of robust synthetic lethality between paralog pairs in cancer cell lines. *Cell Systems*. https://doi.org/10.1016/j.cels.2021.08.006

De Kegel, B., & Ryan, C. J. (2019). Paralog buffering contributes to the variable essentiality of genes in cancer cell lines. *PLoS Genetics*, *15*(10), e1008466. https://doi.org/10.1371/journal.pgen.1008466

Dean, E. J., Davis, J. C., Davis, R. W., & Petrov, D. A. (2008). Pervasive and persistent redundancy among duplicated genes in yeast. *PLoS Genetics*, *4*(7), e1000113. https://doi.org/10.1371/journal.pgen.1000113

Debela, D. T., Muzazu, S. G., Heraro, K. D., Ndalama, M. T., Mesele, B. W., Haile, D. C., Kitui, S. K., & Manyazewal, T. (2021). New approaches and procedures for cancer treatment: Current perspectives. *SAGE Open Medicine*, *9*, 20503121211034370. https://doi.org/10.1177/20503121211034366

Dede, M., Kim, E., & Hart, T. (2020). Biases and Blind-Spots in Genome-Wide CRISPR Knockout Screens. In *bioRxiv* (p. 2020.01.16.909606). https://doi.org/10.1101/2020.01.16.909606

Dede, M., McLaughlin, M., Kim, E., & Hart, T. (2020). Multiplex enCas12a screens detect functional buffering among paralogs otherwise masked in monogenic Cas9 knockout screens. *Genome Biology*, *21*(1), 262. https://doi.org/10.1186/s13059-020-02173-2

DeLuna, A., Vetsigian, K., Shoresh, N., Hegreness, M., Colón-González, M., Chao, S., & Kishony, R. (2008). Exposing the fitness contribution of duplicated genes. *Nature Genetics*, *40*(5), 676–681. https://doi.org/10.1038/ng.123

Dennis, M. Y., & Eichler, E. E. (2016). Human adaptation and evolution by segmental duplication. *Current Opinion in Genetics & Development*, *41*, 44–52. https://doi.org/10.1016/j.gde.2016.08.001

Deshpande, R., Asiedu, M. K., Klebig, M., Sutor, S., Kuzmin, E., Nelson, J., Piotrowski, J., Shin, S. H., Yoshida, M., Costanzo, M., Boone, C., Wigle, D. A., & Myers, C. L. (2013). A comparative genomic approach for identifying synthetic lethal interactions in human cancer. *Cancer Research*, *73*(20), 6128–6136. https://doi.org/10.1158/0008-5472.CAN-12-3956

DeWeirdt, P. C., McGee, A. V., Zheng, F., Nwolah, I., Hegde, M., & Doench, J. G. (2022). Accounting for small variations in the tracrRNA sequence improves sgRNA activity predictions for CRISPR screening. In *bioRxiv* (p. 2022.06.27.497780). https://doi.org/10.1101/2022.06.27.497780

DeWeirdt, P. C., Sanson, K. R., Sangree, A. K., Hegde, M., Hanna, R. E., Feeley, M. N., Griffith, A. L., Teng, T., Borys, S. M., Strand, C., Joung, J. K., Kleinstiver, B. P., Pan, X., Huang, A., & Doench, J. G. (2020). Optimization of AsCas12a for combinatorial genetic screens in human cells. *Nature Biotechnology*. https://doi.org/10.1038/s41587-020-0600-6

Dey, P., Baddour, J., Muller, F., Wu, C. C., Wang, H., Liao, W.-T., Lan, Z., Chen, A., Gutschner, T., Kang, Y., Fleming, J., Satani, N., Zhao, D., Achreja, A., Yang, L., Lee, J., Chang, E., Genovese, G., Viale, A., … DePinho, R. A. (2017). Genomic deletion of malic enzyme 2 confers collateral lethality in pancreatic cancer. *Nature*, *542*(7639), 119–123. https://doi.org/10.1038/nature21052

Diss, G., Gagnon-Arsenault, I., Dion-Coté, A.-M., Vignaud, H., Ascencio, D. I., Berger, C. M., & Landry, C. R. (2017). Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. *Science*, *355*(6325), 630–634. https://doi.org/10.1126/science.aai7685

Doench, J. G. (2018). Am I ready for CRISPR? A user's guide to genetic screens. *Nature Reviews. Genetics*, *19*(2), 67–80. https://doi.org/10.1038/nrg.2017.97

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, *34*(2), 184–191. https://doi.org/10.1038/nbt.3437

Druker, B. J., Sawyers, C. L., Kantarjian, H., Resta, D. J., Reese, S. F., Ford, J. M., Capdeville, R., & Talpaz, M. (2001). Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *The New England Journal of Medicine*, *344*(14), 1038–1042. https://doi.org/10.1056/NEJM200104053441402

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., & Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, *411*(6836), 494–498. https://doi.org/10.1038/35078107

Farmer, H., McCabe, N., Lord, C. J., Tutt, A. N. J., Johnson, D. A., Richardson, T. B., Santarosa, M., Dillon, K. J., Hickson, I., Knights, C., Martin, N. M. B., Jackson, S. P., Smith, G. C. M., & Ashworth, A. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, *434*(7035), 917–921. https://doi.org/10.1038/nature03445

Felsenstein, J. (1965). The effect of linkage on directional selection. *Genetics*, *52*(2), 349–363. https://www.ncbi.nlm.nih.gov/pubmed/5861564

Finan, C., Gaulton, A., Kruger, F. A., Lumbers, R. T., Shah, T., Engmann, J., Galver, L., Kelley, R., Karlsson, A., Santos, R., Overington, J. P., Hingorani, A. D., & Casas, J. P. (2017). The druggable genome and support for target identification and validation in drug development. *Science Translational Medicine*, *9*(383). https://doi.org/10.1126/scitranslmed.aag1166

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, *391*(6669), 806–811. https://doi.org/10.1038/35888

Fisher, R. A. (1919). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, *52*(2), 399–433. https://doi.org/10.1017/S0080456800012163

Gagnon, K. B., & Delpire, E. (2012). Molecular physiology of SPAK and OSR1: two Ste20-related protein kinases regulating ion transport. *Physiological Reviews*, *92*(4), 1577–1617. https://doi.org/10.1152/physrev.00009.2012

Gallo, D., Young, J. T. F., Fourtounis, J., Martino, G., Álvarez-Quilón, A., Bernier, C., Duffy, N. M., Papp, R., Roulston, A., Stocco, R., Szychowski, J., Veloso, A., Alam, H., Baruah, P. S., Fortin, A. B., Bowlan, J., Chaudhary, N., Desjardins, J., Dietrich, E., … Durocher, D. (2022). CCNE1 amplification is synthetic lethal with PKMYT1 kinase inhibition. *Nature*, *604*(7907), 749–756. https://doi.org/10.1038/s41586-022-04638-9

Garnett, M. J., Edelman, E. J., Heidorn, S. J., Greenman, C. D., Dastur, A., Lau, K. W., Greninger, P., Thompson, I. R., Luo, X., Soares, J., Liu, Q., Iorio, F., Surdez, D., Chen, L., Milano, R. J., Bignell, G. R., Tam, A. T., Davies, H., Stevenson, J. A., … Benes, C. H. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, *483*(7391), 570–575. https://doi.org/10.1038/nature11005

Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences*, *109*(39), E2579–E2586. https://doi.org/10.1073/pnas.1208507109

Gasperini, M., Findlay, G. M., McKenna, A., Milbank, J. H., Lee, C., Zhang, M. D., Cusanovich, D. A., & Shendure, J. (2017). CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. *American Journal of Human Genetics*, *101*(2), 192–205. https://doi.org/10.1016/j.ajhg.2017.06.010

Ghandi, M., Huang, F. W., Jané-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E. R., 3rd, Barretina, J., Gelfand, E. T., Bielski, C. M., Li, H., Hu, K., Andreev-Drakhlin, A. Y., Kim, J., Hess, J. M., Haas, B. J., Aguet, F., Weir, B. A., Rothberg, M. V., Paolella, B. R., … Sellers, W. R. (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*, *569*(7757), 503–508. https://doi.org/10.1038/s41586-019-1186-3

Gier, R. A., Budinich, K. A., Evitt, N. H., Cao, Z., Freilich, E. S., Chen, Q., Qi, J., Lan, Y., Kohli, R. M., & Shi, J. (2020). High-performance CRISPR-Cas12a genome editing for combinatorial genetic screening. *Nature Communications*, *11*(1), 3455. https://doi.org/10.1038/s41467-020-17209-1

Gonatopoulos-Pournatzis, T., Aregger, M., Brown, K. R., Farhangmehr, S., Braunschweig, U., Ward, H. N., Ha, K. C. H., Weiss, A., Billmann, M., Durbic, T., Myers, C. L., Blencowe, B. J., & Moffat, J. (2020). Genetic interaction mapping and exon-resolution functional

genomics with a hybrid Cas9–Cas12a platform. *Nature Biotechnology*, 1–11. https://doi.org/10.1038/s41587-020-0437-z

Gonçalves, E., Thomas, M., Behan, F. M., Picco, G., Pacini, C., Allen, F., Vinceti, A., Sharma, M., Jackson, D. A., Price, S., Beaver, C. M., Dovey, O., Parry-Smith, D., Iorio, F., Parts, L., Yusa, K., & Garnett, M. J. (2021). Minimal genome-wide human CRISPR-Cas9 library. *Genome Biology*, *22*(1), 40. https://doi.org/10.1186/s13059-021-02268-4

Grueneberg, D. A., Degot, S., Pearlberg, J., Li, W., Davies, J. E., Baldwin, A., Endege, W., Doench, J., Sawyer, J., Hu, Y., Boyce, F., Xian, J., Munger, K., & Harlow, E. (2008). Kinase requirements in human cells: I. Comparing kinase requirements across various cell types. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(43), 16472–16477. https://doi.org/10.1073/pnas.0808019105

Grueneberg, D. A., Li, W., Davies, J. E., Sawyer, J., Pearlberg, J., & Harlow, E. (2008). Kinase requirements in human cells: IV. Differential kinase requirements in cervical and renal human tumor cell lines. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(43), 16490–16495. https://doi.org/10.1073/pnas.0806578105

Guan, Y., Dunham, M. J., & Troyanskaya, O. G. (2007). Functional analysis of gene duplications in Saccharomyces cerevisiae. *Genetics*, *175*(2), 933–943. https://doi.org/10.1534/genetics.106.064329

Han, K., Jeng, E. E., Hess, G. T., Morgens, D. W., Li, A., & Bassik, M. C. (2017). Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nature Biotechnology*, *35*(5), 463–474. https://doi.org/10.1038/nbt.3834

Hanna, R. E., & Doench, J. G. (2020). Design and analysis of CRISPR-Cas experiments. *Nature Biotechnology*, *38*(7), 813–823. https://doi.org/10.1038/s41587-020-0490-7

Harland, R., & Weintraub, H. (1985). Translation of mRNA injected into Xenopus oocytes is specifically inhibited by antisense RNA. *The Journal of Cell Biology*, *101*(3), 1094–1099. https://doi.org/10.1083/jcb.101.3.1094

Harrison, R., Papp, B., Pál, C., Oliver, S. G., & Delneri, D. (2007). Plasticity of genetic interactions in metabolic networks of yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(7), 2307–2312. https://doi.org/10.1073/pnas.0607153104

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K. R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., Mero, P., Dirks, P., Sidhu, S., Roth, F. P., Rissland, O. S., Durocher, D., Angers, S., & Moffat, J. (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*, *163*(6), 1515–1526. https://doi.org/10.1016/j.cell.2015.11.015

Hart, T., & Moffat, J. (2016). BAGEL: a computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics*, *17*, 164. https://doi.org/10.1186/s12859-016-1015-8

Hart, T., Tong, A. H. Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A., Habsid, A., Sizova, O., Nedyalkova, L., Climie, R., Tworzyanski, L., Lawson, K., Sartori, M. A., Alibeh, S., Tieu, D., … Moffat, J. (2017). Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3* , *7*(8), 2719–2727. https://doi.org/10.1534/g3.117.041277

Hartwell, L. H., Szankasi, P., Roberts, C. J., Murray, A. W., & Friend, S. H. (1997). Integrating genetic approaches into the discovery of anticancer drugs. *Science*, *278*(5340), 1064–1068. https://doi.org/10.1126/science.278.5340.1064

Helming, K. C., Wang, X., Wilson, B. G., Vazquez, F., Haswell, J. R., Manchester, H. E., Kim, Y., Kryukov, G. V., Ghandi, M., Aguirre, A. J., Jagani, Z., Wang, Z., Garraway, L. A., Hahn, W. C., & Roberts, C. W. M. (2014). ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nature Medicine*, *20*, 251. https://doi.org/10.1038/nm.3480

Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., Shen, R., Taylor, A. M., Cherniack, A. D., Thorsson, V., Akbani, R., Bowlby, R., Wong, C. K., Wiznerowicz, M., Sanchez-Vega, F., Robertson, A. G., Schneider, B. G., Lawrence, M. S., Noushmehr, H., … Laird, P. W. (2018). Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell*, *173*(2), 291-304.e6. https://doi.org/10.1016/j.cell.2018.03.022

Hoffman, G. R., Rahal, R., Buxton, F., Xiang, K., McAllister, G., Frias, E., Bagdasarian, L., Huber, J., Lindeman, A., Chen, D., Romero, R., Ramadan, N., Phadke, T., Haas, K., Jaskelioff, M., Wilson, B. G., Meyer, M. J., Saenz-Vash, V., Zhai, H., … Jagani, Z. (2014). Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(8), 3128–3133. https://doi.org/10.1073/pnas.1316793111

Horlbeck, M. A., Xu, A., Wang, M., Bennett, N. K., Park, C. Y., Bogdanoff, D., Adamson, B., Chow, E. D., Kampmann, M., Peterson, T. R., Nakamura, K., Fischbach, M. A., Weissman, J. S., & Gilbert, L. A. (2018). Mapping the Genetic Landscape of Human Cells. *Cell*, *174*(4), 953-967.e22. https://doi.org/10.1016/j.cell.2018.06.010

Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., & Liu, D. R. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, *556*(7699), 57–63. https://doi.org/10.1038/nature26155

Huang, A., Garraway, L. A., Ashworth, A., & Weber, B. (2020). Synthetic lethality as an engine for cancer drug target discovery. *Nature Reviews. Drug Discovery*, *19*(1), 23–38. https://doi.org/10.1038/s41573-019-0046-z

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., … Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, *12*(2), 115–121. https://doi.org/10.1038/nmeth.3252

Iqbal, N., & Iqbal, N. (2014). Imatinib: a breakthrough of targeted therapy in cancer. *Chemotherapy Research and Practice*, *2014*, 357027. https://doi.org/10.1155/2014/357027

Ito, T., Young, M. J., Li, R., Jain, S., Wernitznig, A., Krill-Burger, J. M., Lemke, C. T., Monducci, D., Rodriguez, D. J., Chang, L., Dutta, S., Pal, D., Paolella, B. R., Rothberg, M. V., Root, D. E., Johannessen, C. M., Parida, L., Getz, G., Vazquez, F., … Sellers, W. R. (2021). Paralog knockout profiling identifies DUSP4 and DUSP6 as a digenic dependence in MAPK pathway-driven cancers. *Nature Genetics*, 1–9. https://doi.org/10.1038/s41588-021-00967-z

Izant, J. G., & Weintraub, H. (1984). Inhibition of thymidine kinase gene expression by anti-sense RNA: a molecular approach to genetic analysis. *Cell*, *36*(4), 1007–1015. https://doi.org/10.1016/0092-8674(84)90050-3

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *ELife*, *2*, e00471. https://doi.org/10.7554/eLife.00471

Joung, J., Konermann, S., Gootenberg, J. S., Abudayyeh, O. O., Platt, R. J., Brigham, M. D., Sanjana, N. E., & Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nature Protocols*, *12*(4), 828–863. https://doi.org/10.1038/nprot.2017.016

Kaelin, W. G., Jr. (1999). Choosing anticancer drug targets in the postgenomic era. *The Journal of Clinical Investigation*, *104*(11), 1503–1506. https://doi.org/10.1172/JCI8888

Kaelin, W. G., Jr. (2005). The concept of synthetic lethality in the context of anticancer therapy. *Nature Reviews. Cancer*, *5*(9), 689–698. https://doi.org/10.1038/nrc1691

Kim, W. Y., & Sharpless, N. E. (2006). The regulation of INK4/ARF in cancer and aging. *Cell*, *127*(2), 265–275. https://doi.org/10.1016/j.cell.2006.10.003

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., & Joung, J. K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, *529*(7587), 490–495. https://doi.org/10.1038/nature16526

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J.-R. J., Aryee, M. J., & Joung, J. K. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, *523*(7561), 481–485. https://doi.org/10.1038/nature14592

Kleinstiver, B. P., Sousa, A. A., Walton, R. T., Tak, Y. E., Hsu, J. Y., Clement, K., Welch, M. M., Horng, J. E., Malagon-Lopez, J., Scarfò, I., Maus, M. V., Pinello, L., Aryee, M. J., & Joung, J. K. (2019). Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nature Biotechnology*, *37*(3), 276–282. https://doi.org/10.1038/s41587-018-0011-0

Köferle, A., Schlattl, A., Hörmann, A., Thatikonda, V., Popa, A., Spreitzer, F., Ravichandran, M. C., Supper, V., Oberndorfer, S., Puchner, T., Wieshofer, C., Corcokovic, M., Reiser, C., Wöhrle, S., Popow, J., Pearson, M., Martinez, J., Weitzer, S., Mair, B., & Neumüller, R. A. (2022). Interrogation of cancer gene dependencies reveals paralog interactions of autosome and sex chromosome-encoded genes. *Cell Reports*, *39*(2), 110636. https://doi.org/10.1016/j.celrep.2022.110636

Kryukov, G. V., Wilson, F. H., Ruth, J. R., Paulk, J., Tsherniak, A., Marlow, S. E., Vazquez, F., Weir, B. A., Fitzgerald, M. E., Tanaka, M., Bielski, C. M., Scott, J. M., Dennis, C., Cowley, G. S., Boehm, J. S., Root, D. E., Golub, T. R., Clish, C. B., Bradner, J. E., … Garraway, L. A. (2016). *MTAP* deletion confers enhanced dependency on the PRMT5 arginine methyltransferase in cancer cells. *Science*, *351*(6278), 1214–1218. https://doi.org/10.1126/science.aad5214

Kuzmin, E., VanderSluis, B., Nguyen Ba, A. N., Wang, W., Koch, E. N., Usaj, M., Khmelinskii, A., Usaj, M. M., van Leeuwen, J., Kraus, O., Tresenrider, A., Pryszlak, M., Hu, M.-C., Varriano, B., Costanzo, M., Knop, M., Moses, A., Myers, C. L., Andrews, B. J., & Boone, C. (2020). Exploring whole-genome duplicate gene retention with complex genetic interaction analysis. *Science*, *368*(6498). https://doi.org/10.1126/science.aaz5667

Kuzmin, E., VanderSluis, B., Wang, W., Tan, G., Deshpande, R., Chen, Y., Usaj, M., Balint, A., Mattiazzi Usaj, M., van Leeuwen, J., Koch, E. N., Pons, C., Dagilis, A. J., Pryszlak, M.,

Wang, J. Z. Y., Hanchard, J., Riggi, M., Xu, K., Heydari, H., … Myers, C. L. (2018). Systematic analysis of complex genetic interactions. *Science*, *360*(6386). https://doi.org/10.1126/science.aao1729

Lan, X., & Pritchard, J. K. (2016). Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. *Science*, *352*(6288), 1009–1013. https://doi.org/10.1126/science.aad8411

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., … International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., & Qi, L. S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols*, *8*(11), 2180–2196. https://doi.org/10.1038/nprot.2013.132

Lavi, O. (2015). Redundancy: a critical obstacle to improving cancer therapy. *Cancer Research*, *75*(5), 808–812. https://doi.org/10.1158/0008-5472.CAN-14-3256

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, W., Xu, H., Xiao, T., Cong, L., Love, M. I., Zhang, F., Irizarry, R. A., Liu, J. S., Brown, M., & Liu, X. S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biology*, *15*(12), 554. https://doi.org/10.1186/s13059-014-0554-4

Li, Z., Fang, R., Fang, J., He, S., & Liu, T. (2018). Functional implications of Rab27 GTPases in Cancer. *Cell Communication and Signaling: CCS*, *16*(1), 44. https://doi.org/10.1186/s12964-018-0255-9

Lin, A., Giuliano, C. J., Palladino, A., John, K. M., Abramowicz, C., Yuan, M. L., Sausville, E. L., Lukow, D. A., Liu, L., Chait, A. R., Galluzzo, Z. C., Tucker, C., & Sheltzer, J. M. (2019). Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Science Translational Medicine*, *11*(509). https://doi.org/10.1126/scitranslmed.aaw8412

Lin, X., Morgan-Lappe, S., Huang, X., Li, L., Zakula, D. M., Vernetti, L. A., Fesik, S. W., & Shen, Y. (2007). "Seed" analysis of off-target siRNAs reveals an essential role of Mcl-1 in resistance to the small-molecule Bcl-2/Bcl-XL inhibitor ABT-737. *Oncogene*, *26*(27), 3972–3979. https://doi.org/10.1038/sj.onc.1210166

Lord, C. J., & Ashworth, A. (2017). PARP inhibitors: Synthetic lethality in the clinic. *Science*, *355*(6330), 1152–1158. https://doi.org/10.1126/science.aam7344

Lord, C. J., Quinn, N., & Ryan, C. J. (2020). Integrative analysis of large-scale loss-of-function screens identifies robust cancer-associated genetic interactions. *ELife*, *9*. https://doi.org/10.7554/eLife.58925

Loyer, P., & Trembley, J. H. (2020). Roles of CDK/Cyclin complexes in transcription and pre-mRNA splicing: Cyclins L and CDK11 at the cross-roads of cell cycle and regulation of

gene expression. *Seminars in Cell & Developmental Biology*, *107*, 36–45. https://doi.org/10.1016/j.semcdb.2020.04.016

Loyer, P., Trembley, J. H., Grenet, J. A., Busson, A., Corlu, A., Zhao, W., Kocak, M., Kidd, V. J., & Lahti, J. M. (2008). Characterization of Cyclin L1 and L2 Interactions with CDK11 and Splicing Factors: INFLUENCE OF CYCLIN L ISOFORMS ON SPLICE SITE SELECTION*. *The Journal of Biological Chemistry*, *283*(12), 7721–7732. https://doi.org/10.1074/jbc.M708188200

Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A., Brannigan, B. W., Harris, P. L., Haserlat, S. M., Supko, J. G., Haluska, F. G., Louis, D. N., Christiani, D. C., Settleman, J., & Haber, D. A. (2004). Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. *The New England Journal of Medicine*, *350*(21), 2129–2139. https://doi.org/10.1056/NEJMoa040938

Maffini, M., Denes, V., Sonnenschein, C., Soto, A., & Geck, P. (2008). APRIN is a unique Pds5 paralog with features of a chromatin regulator in hormonal differentiation. *The Journal of Steroid Biochemistry and Molecular Biology*, *108*(1–2), 32–43. https://doi.org/10.1016/j.jsbmb.2007.05.034

Mair, B., Tomic, J., Masud, S. N., Tonge, P., Weiss, A., Usaj, M., Tong, A. H. Y., Kwan, J. J., Brown, K. R., Titus, E., Atkins, M., Chan, K. S. K., Munsie, L., Habsid, A., Han, H., Kennedy, M., Cohen, B., Keller, G., & Moffat, J. (2019). Essential Gene Profiles for Human Pluripotent Stem Cells Identify Uncharacterized Genes and Substrate Dependencies. *Cell Reports*, *27*(2), 599-615.e12. https://doi.org/10.1016/j.celrep.2019.02.041

Mateos-Gomez, P. A., Gong, F., Nair, N., Miller, K. M., Lazzerini-Denchi, E., & Sfeir, A. (2015). Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature*, *518*(7538), 254–257. https://doi.org/10.1038/nature14157

Mavrakis, K. J., McDonald, E. R., 3rd, Schlabach, M. R., Billy, E., Hoffman, G. R., deWeck, A., Ruddy, D. A., Venkatesan, K., Yu, J., McAllister, G., Stump, M., deBeaumont, R., Ho, S., Yue, Y., Liu, Y., Yan-Neale, Y., Yang, G., Lin, F., Yin, H., … Sellers, W. R. (2016). Disordered methionine metabolism in MTAP/CDKN2A-deleted cancers leads to dependence on PRMT5. *Science*, *351*(6278), 1208–1213. https://doi.org/10.1126/science.aad5944

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., Goodale, A., Lee, Y., Ali, L. D., Jiang, G., Lubonja, R., Harrington, W. F., Strickland, M., Wu, T., Hawes, D. C., … Tsherniak, A. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nature Genetics*, *49*(12), 1779–1784. https://doi.org/10.1038/ng.3984

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). Sustainable data analysis with Snakemake. *F1000Research*, *10*, 33. https://doi.org/10.12688/f1000research.29032.2

Morgens, D. W., Wainberg, M., Boyle, E. A., Ursu, O., Araya, C. L., Tsui, C. K., Haney, M. S., Hess, G. T., Han, K., Jeng, E. E., Li, A., Snyder, M. P., Greenleaf, W. J., Kundaje, A., & Bassik, M. C. (2017). Genome-scale measurement of off-target activity using Cas9

toxicity in high-throughput screens. *Nature Communications*, *8*, 15178. https://doi.org/10.1038/ncomms15178

Mullard, A. (2022). What's next for the synthetic lethality drug discovery engine? *Nature Reviews. Drug Discovery*. https://doi.org/10.1038/d41573-022-00107-0

Najm, F. J., Strand, C., Donovan, K. F., Hegde, M., Sanson, K. R., Vaimberg, E. W., Sullender, M. E., Hartenian, E., Kalani, Z., Fusi, N., Listgarten, J., Younger, S. T., Bernstein, B. E., Root, D. E., & Doench, J. G. (2018). Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. *Nature Biotechnology*, *36*(2), 179–189. https://doi.org/10.1038/nbt.4048

Neggers, J. E., Paolella, B. R., Asfaw, A., Rothberg, M. V., Skipper, T. A., Yang, A., Kalekar, R. L., Krill-Burger, J. M., Dharia, N. V., Kugener, G., Kalfon, J., Yuan, C., Dumont, N., Gonzalez, A., Abdusamad, M., Li, Y. Y., Spurr, L. F., Wu, W. W., Durbin, A. D., … Aguirre, A. J. (2020). Synthetic Lethal Interaction between the ESCRT Paralog Enzymes VPS4A and VPS4B in Cancers Harboring Loss of Chromosome 18q or 16q. *Cell Reports*, *33*(11), 108493. https://doi.org/10.1016/j.celrep.2020.108493

Nellen, W., & Lichtenstein, C. (1993). What makes an mRNA anti-sense-itive? *Trends in Biochemical Sciences*, *18*(11), 419–423. https://doi.org/10.1016/0968-0004(93)90137-c

Ohno, S. (2013). *Evolution by Gene Duplication*. Springer Science & Business Media. https://play.google.com/store/books/details?id=5SjqCAAAQBAJ

O'Neil, N. J., Bailey, M. L., & Hieter, P. (2017). Synthetic lethality and cancer. *Nature Reviews. Genetics*, *18*(10), 613–623. https://doi.org/10.1038/nrg.2017.47

Pacini, C., Dempster, J. M., Boyle, I., Gonçalves, E., Najgebauer, H., Karakoc, E., van der Meer, D., Barthorpe, A., Lightfoot, H., Jaaks, P., McFarland, J. M., Garnett, M. J., Tsherniak, A., & Iorio, F. (2021). Integrated cross-study datasets of genetic dependencies in cancer. *Nature Communications*, *12*(1), 1661. https://doi.org/10.1038/s41467-021-21898-7

Paez, J. G., Jänne, P. A., Lee, J. C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F. J., Lindeman, N., Boggon, T. J., Naoki, K., Sasaki, H., Fujii, Y., Eck, M. J., Sellers, W. R., Johnson, B. E., & Meyerson, M. (2004). EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*, *304*(5676), 1497–1500. https://doi.org/10.1126/science.1099314

Pao, W., Miller, V., Zakowski, M., Doherty, J., Politi, K., Sarkaria, I., Singh, B., Heelan, R., Rusch, V., Fulton, L., Mardis, E., Kupfer, D., Wilson, R., Kris, M., & Varmus, H. (2004). EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(36), 13306–13311. https://doi.org/10.1073/pnas.0405220101

Parrish, P. C. R., Thomas, J. D., Gabel, A. M., Kamlapurkar, S., Bradley, R. K., & Berger, A. H. (2021). Discovery of synthetic lethal and tumor suppressor paralog pairs in the human genome. *Cell Reports*, *36*(9). https://doi.org/10.1016/j.celrep.2021.109597

Phillips, P. C. (1998). The language of gene interaction. *Genetics*, *149*(3), 1167–1171. https://www.ncbi.nlm.nih.gov/pubmed/9649511

Ryan, C. J., Mehta, I., Kebabci, N., & Adams, D. J. (2023). Targeting synthetic lethal paralogs in cancer. *Trends in Cancer Research*, *0*(0). https://doi.org/10.1016/j.trecan.2023.02.002

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., La, K. C., Dimitriadoy, S., Liu, D. L., Kantheti, H. S., Saghafinia, S., Chakravarty, D., Daian, F., Gao, Q., Bailey, M. H., Liang, W.-W., Foltz, S. M., Shmulevich, I., Ding, L., Heins, Z., … Schultz, N.

(2018). Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell*, *173*(2), 321-337.e10. https://doi.org/10.1016/j.cell.2018.03.035

Sanjana, N. E., Shalem, O., & Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods*, *11*(8), 783–784. https://doi.org/10.1038/nmeth.3047

Sanson, K. R., Hanna, R. E., Hegde, M., Donovan, K. F., Strand, C., Sullender, M. E., Vaimberg, E. W., Goodale, A., Root, D. E., Piccioni, F., & Doench, J. G. (2018). Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nature Communications*, *9*(1), 5416. https://doi.org/10.1038/s41467-018-07901-8

Schirrmacher, V. (2019). From chemotherapy to biological therapy: A review of novel concepts to reduce the side effects of systemic cancer treatment (Review). *International Journal of Oncology*, *54*(2), 407–419. https://doi.org/10.3892/ijo.2018.4661

Segrè, D., Deluna, A., Church, G. M., & Kishony, R. (2005). Modular epistasis in yeast metabolism. *Nature Genetics*, *37*(1), 77–83. https://doi.org/10.1038/ng1489

Setton, J., Zinda, M., Riaz, N., Durocher, D., Zimmermann, M., Koehler, M., Reis-Filho, J. S., & Powell, S. N. (2021). Synthetic Lethality in Cancer Therapeutics: The Next Generation. *Cancer Discovery*, *11*(7), 1626–1635. https://doi.org/10.1158/2159-8290.CD-20-1503

Shen, J. P., Zhao, D., Sasik, R., Luebeck, J., Birmingham, A., Bojorquez-Gomez, A., Licon, K., Klepper, K., Pekin, D., Beckett, A. N., Sanchez, K. S., Thomas, A., Kuo, C.-C., Du, D., Roguev, A., Lewis, N. E., Chang, A. N., Kreisberg, J. F., Krogan, N., … Mali, P. (2017). Combinatorial CRISPR–Cas9 screens for de novo mapping of genetic interactions. *Nature Methods*, *14*(6), 573–576. https://doi.org/10.1038/nmeth.4225

Singh, P. P., Affeldt, S., Cascone, I., Selimoglu, R., Camonis, J., & Isambert, H. (2012). On the Expansion of "Dangerous" Gene Repertoires by Whole-Genome Duplications in Early Vertebrates. *Cell Reports*, *2*(5), 1387–1398. https://doi.org/10.1016/j.celrep.2012.09.034

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science*, *351*(6268), 84–88. https://doi.org/10.1126/science.aad5227

Smith, I., Greenside, P. G., Natoli, T., Lahr, D. L., Wadden, D., Tirosh, I., Narayan, R., Root, D. E., Golub, T. R., Subramanian, A., & Doench, J. G. (2017). Evaluation of RNAi and CRISPR technologies by large-scale gene expression profiling in the Connectivity Map. *PLoS Biology*, *15*(11), e2003213. https://doi.org/10.1371/journal.pbio.2003213

Stewart, S. A., Dykxhoorn, D. M., Palliser, D., Mizuno, H., Yu, E. Y., An, D. S., Sabatini, D. M., Chen, I. S. Y., Hahn, W. C., Sharp, P. A., Weinberg, R. A., & Novina, C. D. (2003). Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* , *9*(4), 493–501. https://doi.org/10.1261/rna.2192803

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Szymańska, E., Nowak, P., Kolmus, K., Cybulska, M., Goryca, K., Derezińska-Wołek, E., Szumera-Ciećkiewicz, A., Brewińska-Olchowik, M., Grochowska, A., Piwocka, K., Prochorec-Sobieszek, M., Mikula, M., & Miączyńska, M. (2020). Synthetic lethality between VPS4A and VPS4B triggers an inflammatory response in colorectal cancer. *EMBO Molecular Medicine*, *12*(2), e10812. https://doi.org/10.15252/emmm.201910812

Tang, S., Wu, X., Liu, J., Zhang, Q., Wang, X., Shao, S., Gokbag, B., Fan, K., Liu, X., Li, F., Cheng, L., & Li, L. (2022). Generation of dual-gRNA library for combinatorial CRISPR

screening of synthetic lethal gene pairs. *STAR Protocols*, *3*(3), 101556. https://doi.org/10.1016/j.xpro.2022.101556

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S. C., Kok, C. Y., Noble, K., Ponting, L., Ramshaw, C. C., Rye, C. E., … Forbes, S. A. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*, *47*(D1), D941–D947. https://doi.org/10.1093/nar/gky1015

Taylor, A. M., Shih, J., Ha, G., Gao, G. F., Zhang, X., Berger, A. C., Schumacher, S. E., Wang, C., Hu, H., Liu, J., Lazar, A. J., Caesar-Johnson, S. J., Demchok, J. A., Felau, I., Kasapi, M., Ferguson, M. L., Hutter, C. M., Sofia, H. J., Tarnuzzer, R., … Mariamidze, A. (2018). Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell*, *0*(0). https://doi.org/10.1016/j.ccell.2018.03.007

TCGA. (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, *511*(7511), 543–550. https://doi.org/10.1038/nature13385

Thariat, J., Hannoun-Levi, J.-M., Sun Myint, A., Vuong, T., & Gérard, J.-P. (2013). Past, present, and future of radiotherapy for the benefit of patients. *Nature Reviews. Clinical Oncology*, *10*(1), 52–60. https://doi.org/10.1038/nrclinonc.2012.203

Thomas, J. D., Polaski, J. T., Feng, Q., De Neef, E. J., Hoppe, E. R., McSharry, M. V., Pangallo, J., Gabel, A. M., Belleville, A. E., Watson, J., Nkinsi, N. T., Berger, A. H., & Bradley, R. K. (2020). RNA isoform screens uncover the essentiality and tumor-suppressor activity of ultraconserved poison exons. *Nature Genetics*, 1–11. https://doi.org/10.1038/s41588-019-0555-z

Thompson, N. A., Ranzani, M., van der Weyden, L., Iyer, V., Offord, V., Droop, A., Behan, F., Gonçalves, E., Speak, A., Iorio, F., Hewinson, J., Harle, V., Robertson, H., Anderson, E., Fu, B., Yang, F., Zagnoli-Vieira, G., Chapman, P., Del Castillo Velasco-Herrera, M., … Adams, D. J. (2021). Combinatorial CRISPR screen identifies fitness effects of gene paralogues. *Nature Communications*, *12*(1), 1302. https://doi.org/10.1038/s41467-021-21478-9

Tilsed, C. M., Fisher, S. A., Nowak, A. K., Lake, R. A., & Lesterhuis, W. J. (2022). Cancer chemotherapy: insights into cellular and tumor microenvironmental mechanisms of action. *Frontiers in Oncology*, *12*, 960317. https://doi.org/10.3389/fonc.2022.960317

Tong, A. H. Y., Lesage, G., Bader, G. D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G. F., Brost, R. L., Chang, M., Chen, Y., Cheng, X., Chua, G., Friesen, H., Goldberg, D. S., Haynes, J., Humphries, C., He, G., Hussein, S., … Boone, C. (2004). Global mapping of the yeast genetic interaction network. *Science*, *303*(5659), 808–813. https://doi.org/10.1126/science.1091317

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., Meyers, R. M., Ali, L., Goodale, A., Lee, Y., Jiang, G., Hsiao, J., Gerath, W. F. J., Howell, S., Merkel, E., … Hahn, W. C. (2017). Defining a Cancer Dependency Map. *Cell*, *170*(3), 564-576.e16. https://doi.org/10.1016/j.cell.2017.06.010

van der Meer, D., Barthorpe, S., Yang, W., Lightfoot, H., Hall, C., Gilbert, J., Francies, H. E., & Garnett, M. J. (2019). Cell Model Passports-a hub for clinical, genetic and functional datasets of preclinical cancer models. *Nucleic Acids Research*, *47*(D1), D923–D929. https://doi.org/10.1093/nar/gky872

van Leeuwen, J., Boone, C., & Andrews, B. J. (2017). Mapping a diversity of genetic interactions in yeast. *Current Opinion in Systems Biology*, *6*, 14–21. https://doi.org/10.1016/j.coisb.2017.08.002

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., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., … Zhu, X. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304–1351. https://doi.org/10.1126/science.1058040

Vichas, A., Nkinsi, N. T., Riley, A., Parrish, P. C. R., Duke, F., Chen, J., Fung, I., Watson, J., Rees, M., Lee, J. K., Piccioni, F., Hatch, E. M., & Berger, A. H. (2020). An integrative oncogene-dependency map identifies unique vulnerabilities of oncogenic EGFR, KRAS, and RIT1 in lung cancer. In *BioRxiv* (p. 2020.07.03.187310). Cold Spring Harbor Laboratory. https://doi.org/10.1101/2020.07.03.187310

Vilella, A. J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., & Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Research*, *19*(2), 327–335. https://doi.org/10.1101/gr.073585.107

Viswanathan, S. R., Nogueira, M. F., Buss, C. G., Krill-Burger, J. M., Wawer, M. J., Malolepsza, E., Berger, A. C., Choi, P. S., Shih, J., Taylor, A. M., Tanenbaum, B., Pedamallu, C. S., Cherniack, A. D., Tamayo, P., Strathdee, C. A., Lage, K., Carr, S. A., Schenone, M., Bhatia, S. N., … Meyerson, M. (2018). Genome-scale analysis identifies paralog lethality as a vulnerability of chromosome 1p loss in cancer. *Nature Genetics*, *50*(7), 937–943. https://doi.org/10.1038/s41588-018-0155-3

Walton, R. T., Christie, K. A., Whittaker, M. N., & Kleinstiver, B. P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science*, *368*(6488), 290–296. https://doi.org/10.1126/science.aba8853

Wang, T., Birsoy, K., Hughes, N. W., Krupczak, K. M., Post, Y., Wei, J. J., Lander, E. S., & Sabatini, D. M. (2015). Identification and characterization of essential genes in the human genome. *Science*, *350*(6264), 1096–1101. https://doi.org/10.1126/science.aac7041

Wang, T., Yu, H., Hughes, N. W., Liu, B., Kendirli, A., Klein, K., Chen, W. W., Lander, E. S., & Sabatini, D. M. (2017). Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell*, *168*(5), 890-903.e15. https://doi.org/10.1016/j.cell.2017.01.013

Ward, H. N., Aregger, M., Gonatopoulos-Pournatzis, T., Billmann, M., Ohsumi, T. K., Brown, K. R., Blencowe, B. J., Moffat, J., & Myers, C. L. (2021). Analysis of combinatorial CRISPR screens with the Orthrus scoring pipeline. *Nature Protocols*, *16*(10), 4766–4798. https://doi.org/10.1038/s41596-021-00596-0

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wong, A. S. L., Choi, G. C. G., Cui, C. H., Pregernig, G., Milani, P., Adam, M., Perli, S. D., Kazer, S. W., Gaillard, A., Hermann, M., Shalek, A. K., Fraenkel, E., & Lu, T. K. (2016). Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(9), 2544–2549. https://doi.org/10.1073/pnas.1517883113

Wu, X., Yang, X., Xiong, Y., Li, R., Ito, T., Ahmed, T. A., Karoulia, Z., Adamopoulos, C., Wang, H., Wang, L., Xie, L., Liu, J., Ueberheide, B., Aaronson, S. A., Chen, X., Buchanan, S. G., Sellers, W. R., Jin, J., & Poulikakos, P. I. (2021). Distinct CDK6 complexes determine tumor cell response to CDK4/6 inhibitors and degraders. *Nature Cancer*, 1–15. https://doi.org/10.1038/s43018-021-00174-z

Xue, W., Kitzing, T., Roessler, S., Zuber, J., Krasnitz, A., Schultz, N., Revill, K., Weissmueller, S., Rappaport, A. R., Simon, J., Zhang, J., Luo, W., Hicks, J., Zender, L., Wang, X. W., Powers, S., Wigler, M., & Lowe, S. W. (2012). A cluster of cooperating tumor-suppressor gene candidates in chromosomal deletions. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(21), 8212–8217. https://doi.org/10.1073/pnas.1206062109

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J. A., Thompson, I. R., Ramaswamy, S., Futreal, P. A., Haber, D. A., Stratton, M. R., Benes, C., McDermott, U., & Garnett, M. J. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research*, *41*(Database issue), D955-61. https://doi.org/10.1093/nar/gks1111

Yoda, S., Dagogo-Jack, I., & Hata, A. N. (2019). Targeting oncogenic drivers in lung cancer: Recent progress, current challenges and future opportunities. *Pharmacology & Therapeutics*, *193*, 20–30. https://doi.org/10.1016/j.pharmthera.2018.08.007

Yoshida, K., Gowers, K. H. C., Lee-Six, H., Chandrasekharan, D. P., Coorens, T., Maughan, E. F., Beal, K., Menzies, A., Millar, F. R., Anderson, E., Clarke, S. E., Pennycuick, A., Thakrar, R. M., Butler, C. R., Kakiuchi, N., Hirano, T., Hynds, R. E., Stratton, M. R., Martincorena, I., … Campbell, P. J. (2020). Tobacco smoking and somatic mutations in human bronchial epithelium. *Nature*. https://doi.org/10.1038/s41586-020-1961-1

Zamanighomi, M., Jain, S. S., Ito, T., Pal, D., Daley, T. P., & Sellers, W. R. (2019). GEMINI: a variational Bayesian approach to identify genetic interactions from combinatorial CRISPR screens. *Genome Biology*, *20*(1), 137. https://doi.org/10.1186/s13059-019-1745-9

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., Koonin, E. V., & Zhang, F. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, *163*(3), 759–771. https://doi.org/10.1016/j.cell.2015.09.038

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., Winblad, N., Choudhury, S. R., Abudayyeh, O. O., Gootenberg, J. S., Wu, W. Y., Scott, D. A., Severinov, K., van der Oost, J., & Zhang, F. (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nature Biotechnology*, *35*(1), 31–34. https://doi.org/10.1038/nbt.3737

Zhang, W., Bouchard, G., Yu, A., Shafiq, M., Jamali, M., Shrager, J. B., Ayers, K., Bakr, S., Gentles, A. J., Diehn, M., Quon, A., West, R. B., Nair, V., van de Rijn, M., Napel, S., & Plevritis, S. K. (2018). GFPT2-Expressing Cancer-Associated Fibroblasts Mediate Metabolic Reprogramming in Human Lung Adenocarcinoma. *Cancer Research*, *78*(13), 3445–3457. https://doi.org/10.1158/0008-5472.CAN-17-2928

Zhou, H., Liu, Y., Zhu, R., Ding, F., Wan, Y., Li, Y., & Liu, Z. (2017). FBXO32 suppresses breast cancer tumorigenesis through targeting KLF4 to proteasomal degradation. *Oncogene*, *36*(23), 3312–3321. https://doi.org/10.1038/onc.2016.479

VITA

Phoebe Parrish grew up in Ithaca, New York. She received her Bachelor of Science degree in Biology *magna cum laude* with Honors from Davidson College in 2016. As an undergraduate, she worked in the labs of Dr. A. Malcolm Campbell studying synthetic biology in *Escherichia coli* and Dr. Alan Collmer studying the role of the bacterial toxin coronatine in plant disease in *Nicotiana benthamiana*. She completed a postbaccalaureate Intramural Research and Training Award fellowship in Dr. Beth Kozel's laboratory at the National Institutes of Health, where she identified genetic modifiers of cardiovascular disease severity in patients with Williams-Beuren syndrome. She started her PhD research in Genome Sciences at the University of Washington in 2018. She conducted her dissertation research in Dr. Alice Berger's laboratory in the Human Biology Division of the Fred Hutchinson Cancer Center, where she developed experimental and computational methods to identify new cancer drug targets. She is also passionate about promoting diversity, equity, and inclusion in STEM. Alongside her PhD research, she co-led the Genome Sciences Association for the Inclusion of Marginalized Students and served as a volunteer mentor for Genome Hackers and Coding for Cancer, two summer programs dedicated to helping local high school students from historically excluded backgrounds gain exposure to genetics and computational biology research. In her free time, she enjoys biking, reading, hiking with friends, and taking her dog Jack on walks around Seattle.