Genetic Screening Approaches to Cancer Driver Characterization and Synthetic Lethal Target Discovery

Citation

Mengwasser, Kristen Elizabeth. 2018. Genetic Screening Approaches to Cancer Driver Characterization and Synthetic Lethal Target Discovery. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:41121232

Terms of Use

Share Your Story

Accessibility

Genetic Screening Approaches to Cancer Driver Characterization

and Synthetic Lethal Target Discovery

A dissertation presented

by

Kristen Elizabeth Mengwasser

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

May 2018

Genetic Screening Approaches to Cancer Driver Characterization and Synthetic Lethal Target Discovery

Abstract

Advances in genetic screening technology have expanded the toolkit for systematic perturbation of gene function. While the CRISPR-Cas9 system robustly probes genetic loss-of-function in mammalian cells, a barcoded ORFeome library offers the opportunity to systematically study genetic gain-of-function. We employed these two screening tools for three purposes.

First, we performed shRNA and CRISPR-based screens for synthetic lethality with *BRCA2* deficiency, in two pairs of *BRCA2* isogenic cell lines. *BRCA2* mutation commonly drives hereditary breast and ovarian cancer, but also creates vulnerability to PARP inhibitor treatment. Among other genes, we found the AP endonuclease *APEX2* and the flap endonuclease *FEN1* to be synthetic lethal with *BRCA2* deficiency; the base excision repair (BER) pathway was synthetic lethal overall. We demonstrated that FEN1 plays a role in microhomology-mediated end joining (MMEJ) and that a FEN1 inhibitor selectively targets *BRCA*-deficient cells, offering therapeutic potential in the setting of PARP inhibitor resistance.

Second, we screened a barcoded ORFeome library for genes that either upregulate PD-L1 or interfere with IFNγ signaling. Tumor-expressed PD-L1 engages PD-1 on T cells, dampening T-cell based immune responses to tumors, while defects in

IFNγ signaling underlie resistance to immunotherapy. We found 12 GPCRs that significantly increase cell surface PD-L1 display, including three LPARs (*LPAR1/2/5*), and a family of RNA-binding proteins (*CELF3/4/5*) that upregulates PD-L1. Conversely, we showed that *CLK2* overexpression antagonizes IFNγ signaling; *CLK2* is known to phosphorylate *PTPN1* and is amplified in tumor types that respond poorly to immunotherapy. Thus, segregating patients based on *CLK2* amplification status or augmenting immunotherapy with CLK2 inhibitors may improve clinical responses.

Finally, we utilized CRISPR and ORF-based libraries targeting known and predicted tumor suppressor genes (TSGs) or oncogenes (OGs) to systematically screen cancer drivers for a variety of phenotypes. We found at least 20 genes not commonly annotated as drivers that enhanced proliferation, including *PAWR*, *AMBRA1,* and *USP28*. We screened OGs bearing tumor-associated point mutations and functionally validated the *MYCN* P44L mutation. We mapped drivers across multiple hallmark cancer phenotypes and found that *IRF6* knockout promotes anoikis bypass. Thus, functional studies can augment computational approaches in identifying cancer driver genes.

Table of Contents

List of Commonly Used Abbreviations

APC	Allophycocyanin (Red fluorescent phycobiliprotein)
BER	Base excision repair
CTLA-4	Cytotoxic T lymphocyte antigen-4
CRISPR	Clustered regularly interspaced short palindromic repeats
DAVID	Database for Annotation, Visualization, and Integrated Discovery
DDR	DNA Damage Response
DSB	Double-strand break
DsRed	*Discosoma* sp. red fluorescent protein
E2C	E2-Crimson
edgeR	Empirical analysis of DGE in R
FACS	Fluorescence-activated cell sorting
GFP	Green fluorescent protein
GPCR	G-protein coupled receptor
GSEA	Gene set enrichment analysis
HBOC	Hereditary breast and ovarian cancer
HR	Homologous Recombination
IFNγ	Interferon-gamma
IPA	Ingenuity Pathway Analysis
LPAR	Lysophosphatidic acid receptor
MAGeCK	Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout
MHC	Major histocompatibility complex
MMEJ	Microhomology-mediated end joining
MOI	Multiplicity of infection
NSCLC	Non-small-cell lung cancer
OG	Oncogene
ORF	Open reading frame
PD-1	Programmed cell death protein 1
PD-L1	Programmed death-ligand 1
ROS	Reactive Oxygen Species
rtTA	Reverse tetracycline-controlled transactivator
SCNA	Somatic copy number alteration
TCR	T cell receptor
TRE	Tetracycline Response Element
TSG	Tumor suppressor gene

Acknowledgements

Working in Steve Elledge's lab is an extraordinary privilege and an incredible learning experience for a developing scientist. Over the years, I've often thought of Steve's lab as a garden of ideas. I truly believe that nothing limits one's potential in Steve's lab except one's own imagination, energy, and passion for science. I want to thank Steve for giving me the opportunity to study with him, the chance to learn and grow in the incredible environment he cultivates, and for everything he has taught me during these formative years.

Steve attracts a group of unbelievably brilliant, talented, enthusiastic scientists who have taught me so much during my time in the Elledge lab. I want to start by thanking my rotation mentor Natasha Pavlova, who took me under her wing and shared some of her precious time with me toward the end of her graduate education. To have learned so many basic skills and concepts from such an intuitive, thoughtful scientist is a foundational aspect of my training for which I will be eternally grateful!

I also want to thank Britt Adamson, Lior Izhar, Alberto Ciccia, Chanhee Kang, Jay Guo, Mike Emanuele, Yifan Wang, Anna Burrows, Itay Koren, and Tim Martin for sharing bits of scientific insight and wisdom along the way. I thank Tony Anselmo for teaching me how to do screens with technical proficiency. Mamie Li is a cloning goddess, with an incredible bag of cloning tricks and secrets; I thank her for every insight she has shared with me, and every cloning snag she's helped me troubleshoot. Tony Liang has been my γ-TC-mate for many years; I thank him for his excellent

coaching in achieving record-breaking lentiviral titer, and for the innumerable competent cells he prepares for the lab.

Emma Watson, you are a hoot, and a completely brilliant scientist to boot! Thanks for the fun and the insights into cancer cell metabolism! Shawna Guillemete, it was a blast having you around and learning about BRCA2 from you! Sida Liao, thank you for your help with our attempted drug synergy study and for your fantastic Western blotting tips! Richard Adeyemi, thank you for patiently bearing with me as I learned about the DNA damage response, and for teaching me so many fundamentals. Andy Elia, thank you for your keen and insightful review of my manuscript and for your advice about the MD-PhD career track.

Tomasz Kula has been my γ-TC-mate (long time no T-*"C"*!) and graduate student compatriot for the past six years; I have been so lucky to learn and grow over the years with one of the most kind-hearted, brilliant, humble scientists I've ever had the privilege of meeting. Peter Bruno, thank you for your extremely clutch advice on a variety of occasions, particularly on MCA assay optimization! I can already see your mentoring skills developing, and I know you are going to be an incredible PI. Mo Dezfulian, it has been such a pleasure and privilege working with you, particularly in our recent collaboration on follow-up of the PD-L1 project!

Eric Wooten, thank you so much for all of your help over the years with computational challenges, alignment issues, and debugging. I also want to offer you a particularly massive heap of gratitude for your heroic tissue culture work on the TSG

and OG drug screens. Thank you for your work and your cheerful attitude, particularly during the disastrous media shortages and inconvenient timing that occurred around the holidays.

Qikai Xu, your passion for science continually amazes me, and your knowledge base is unparalleled. I've said it before and I'll say it again: five minutes with Qikai will solve any problem. Yumei Leng, your diligent, thorough work on the OG library produced a truly incredible reagent. Thank you for your thoughtful work on both this library and the BRCA2 site-directed mutagenesis project, and for your daily kindness in general. Daisy Choi, thank you for your incredible work with the TSG and OG library screens last summer, and for your diligent work on our BRCA2 CRISPR screens; it is always a pleasure to work with you!

Kamila Naxerova, I have learned so much from you. Thank you for being such a powerful example of personal strength, independence, and self-direction; I hope some of it rubbed off on me! Teresa Davoli joined the lab just after I arrived, and I am continually awed and inspired by her creativity, her work ethic, her resourcefulness, and her kindness. Teresa, thank you for sharing your time, expertise, and knowledge with me, and for being such an incredible role mode of a developing independent female scientist!

Laura Sack, how can begin to thank universe for bringing us together? You have been my lab sister/friend soulmate during my years in the Elledge lab, and during my thesis writing time you have been my primary lifeline of encouragement and support!

Chapter 1: Introduction

I.   Drivers of Tumorigenesis

Genomic Alterations Drive Cancer

Cancer is a complex, multi-stage disease process through which normal cells progressively acquire the capacity to grow abnormally. Histologically, through sequential stages, tumorigenic tissues increasingly lose normal architecture and function, while their constituent cells exhibit large nuclei with atypical mitotic figures and abnormal cellular structure. Interestingly, the predecessor stage of this process is often a stress-responsive increase in normal cell growth (hyperplasia) or transmutation of cell type (metaplasia), which can lead to the generation of abnormal cells (dysplasia), growth of *in situ* and benign tumors, progression to invasive malignancy, and eventually metastatic spread throughout the body.

These progressive histological stages reflect ongoing accumulation of genomic alterations, some of which actively drive abnormal proliferative behavior. Theodor Boveri first suggested this notion in 1914, when he postulated that an "abnormal combination of chromosomes" (aneuploidy) can give rise to the aberrant proliferative behavior that underlies tumor development (*3*). This concept gained considerable credibility when a particular chromosomal abnormality, the Philadelphia chromosome, was observed in a large proportion of chronic myelogenous leukemia (CML) cases (*4*). The causative link between such chromosomal aberrancies and tumor development

began to emerge from experimental demonstration that physical and chemical carcinogens such as coal tar (*5*), X-rays (*6*), and aflatoxin (*7*) actively mutagenize genetic content (*8*).

Definitive proof of this causative link arose from a revolutionary experiment with another class of carcinogens: tumor-causing viruses. Rous sarcoma virus (RSV), long known to be capable of causing chicken tumors *in vivo* (*9*) and transforming chicken embryonic fibroblasts *in vitro (10)*, contained what could be considered the first identified cancer driver gene. Utilizing the availability of reverse transcriptase, Peter Vogt generated a DNA probe that selectively bound to transformation-competent RSV but not transformation-incompetent RSV; this probe was in fact hybridizing to viral Src (*v*-Src). Surprisingly, this probe also hybridized to normal chicken DNA, highlighting that individual endogenous genetic elements, such as cellular Src (*c*-Src) in this case, are capable of transforming cells to a cancerous state (*11*).

This prescient experiment led to a concept that has remained central to cancer biology: some driver genes such as *c*-Src, when altered, are capable of actively contributing a selective growth advantage to tumorigenic cells. However, many genetic alterations that occur in the process of cancer progression are simply passengers in the tumorigenic process: their alteration does not contribute to tumor development, but rather occurs as a by-product of it. Distinguishing alterations that actively drive tumorigenesis from hitchhiking inactive passenger mutations remains a central challenge in the cancer biology field.
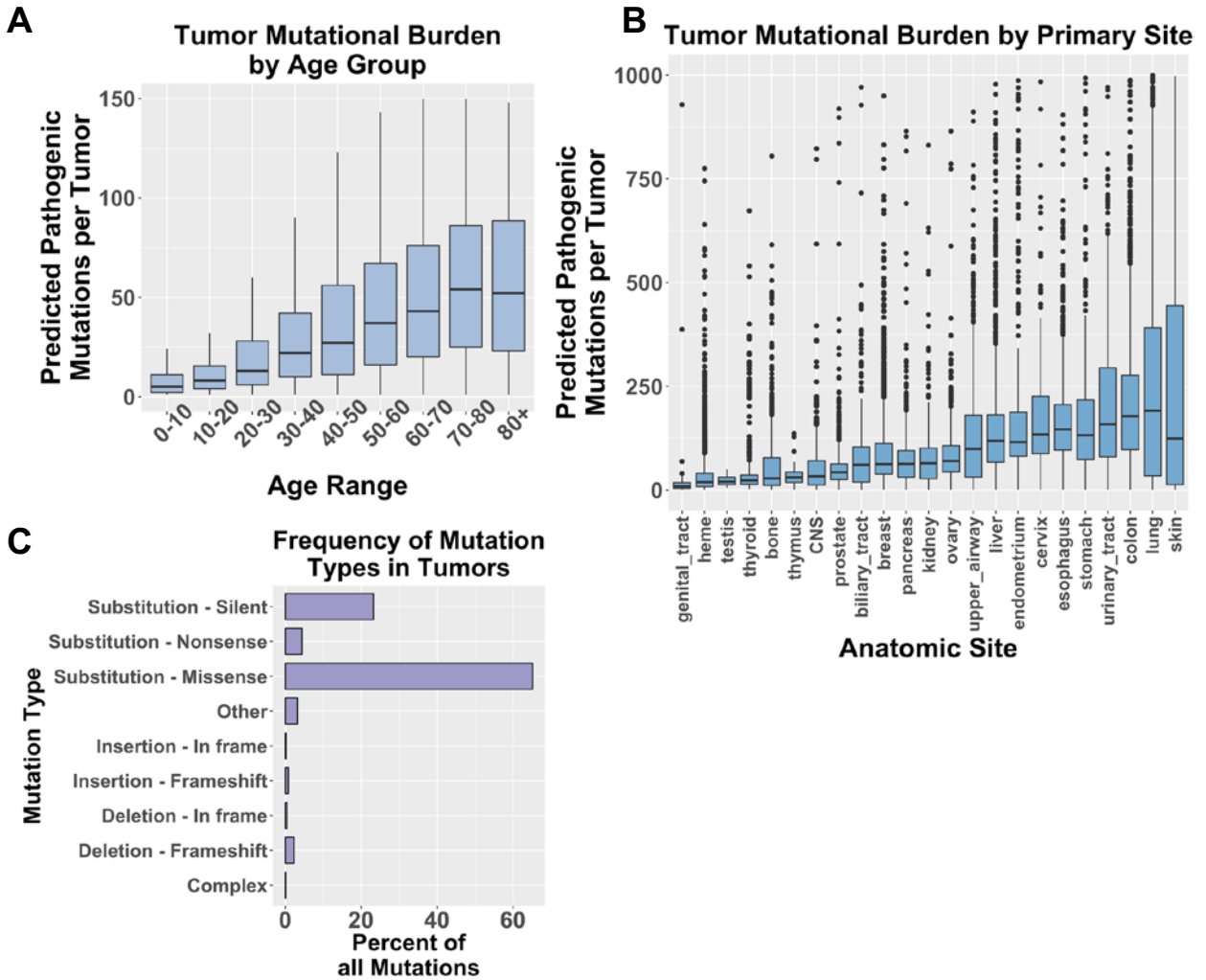
Figure 1-1. Analysis of Mutations Available in Cosmic Database. All mutations available from ~22K tumors in the Cosmic Database were downloaded and analyzed for the indicated properties. (**A**) Median-centered quartile box-and-whisker plot illustrating the number of mutations predicted to be pathogenic by Fathmm *(2)* per tumor for each of the indicated age groups. (**B**) Median-centered quartile box-and-whisker plot illustrating the number of mutations predicted to be pathogenic by Fathmm (*2*) for each of the indicated anatomic sites of the primary tumor; outliers above 1000 predicted pathogenic mutations per tumor are not shown. (**C**) The frequency of each type of mutation is plotted as a percent of all mutations in tumors.

In adult solid tumors, generally 33-66 genes possess mutations in their protein-coding sequence that would be expected to alter their function (*12*). If one examines all mutations in the ~22K tumors currently accessible through the Cosmic database, the

cumulative nature of genetic alterations in tumors is apparent, based on the clear correlation of tumor mutational burden with the age of the patient when tumor sampling was performed (Figure 1-1A). Half or more of these alterations are incidental passenger mutations that occur in the preneoplastic phase (*13*).

To complicate matters, tumor mutational burden also varies widely among tumor types (Figure 1-1B). Carcinogenic insults such as cigarette smoke in lung cancer and UV light in skin cancer create hypermutated tumors, as does loss of DNA mismatch-repair gene functionality that creates microsatellite instability (MSI) in colon cancer (*14*). The majority of observed mutations result in missense point substitutions (Figure 1-1C), a fraction of which impact protein function, and further subgroup of which actively drive tumorigenesis.

Taking a Census of Cancer Drivers

Cancer driver genes generally fall into one of two broad categories: oncogenes (OGs) and tumor suppressor genes (TSGs). Whereas OGs contribute to tumorigenesis when activated, inactivation of TSG function promotes oncogenesis. In tumors, this activation or inactivation can occur by a variety of genetic or epigenetic means. While OGs are commonly activated by missense mutation, translocation, overexpression, focal amplification, or chromosome arm-level amplification, TSGs can be inactivated by nonsense, missense, or splicing site point mutations, focal deletion, chromosome arm-level deletion, epigenetic silencing, or downregulation (*15*).

Of the myriad genetic and epigenetic alterations that accrue during tumor formation, what proportion of these changes functionally contribute to tumorigenesis? And relatedly, how many cancer driver genes exist? A classic epidemiologic study posited that tumors require alterations in 5-8 driver genes to develop (*16*), and it has been suggested that only three driver gene mutations are required to cause tumor formation (*17, 18*). An analysis of mutation data from 12 common cancer types suggested that most tumors harbor mutations in 2-6 genes that commonly act as drivers (*19*). However, an independent analysis found a mean number of ~1 OG mutation, ~3 TSG mutations, ~4 whole chromosome gains or losses, ~8 chromosome arm-level gains or losses, and ~23 focal somatic copy number alterations (SCNAs) per tumor (*15*). This latter study observed that SCNA patterns in tumors correlate with the presence of OGs or TSGs in regions of amplification or deletion respectively, implying selection for altered gene dosage of drivers in regions of SCNAs (*15*).

Though it is equally pertinent to study SCNA and mutation data when identifying driver genes, regions of amplification and deletion are often multigenic. In a comprehensive study of 3,131 tumors across various lineages, 76 common focal amplification peak regions and 82 common focal deletion peak regions were identified, containing a median of 6.5 and 7 genes, respectively (*20*). Revision of this statistical approach, termed GISTIC2.0, has further refined boundaries of focal SCNAs (*21*). This data can be integrated with whole-exome and whole-genome tumor sequencing, RNA-seq, and DNA methylation profiling to improve confidence in driver predictions.

Over time, efforts have converged to integrate this data and define a formal list of cancer driver genes. In 2004, the Cancer Gene Census (CGC) catalogued driver genes and found 291 drivers to be significantly implicated in a causal role in the tumorigenic process (*22*). Other resources have collected lists of 140 driver genes (*12*), or estimates of 200 genes (*17*). In April 2018, *Cell* released the Pan-Cancer Atlas, a series of 27 papers that constitute a meta-analysis of The Cancer Genome Atlas (TCGA) dataset (*23*). TCGA, a joint effort between the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI), collected a vast dataset encompassing 33 tumor types and 11,000 tumors from 2006 to 2016, and released 31 tumor-type specific marker papers over the course of a decade (*24, 25*). The Pan-Cancer Atlas includes a "final consensus list" of 299 driver genes (*26*).

The landscape of tumor mutation frequency has been described as hills, valleys, and mountains, with frequently mutated, highly penetrant drivers being "mountains" in this landscape (*12*). Less frequently mutated "hill" genes may provide a significant driving force in tumors when altered, but their overall frequency of mutation is lower. Despite the formalization of smaller lists of cancer drivers, some estimates of the number of "hills and mountains" range from 350 (*22*) to ~500 genes (*15, 22*).

Although frequency of alteration should correlate with the likelihood of any given gene being a *bona fide* cancer driver, there are many layers of complication to this assumption. For example, somatic mutation rates vary significantly across the genome, more than 100-fold between some regions, introducing inherent bias (*27*). In addition,

chromosomal linkage to neighboring genes that promote proliferation may bias a false positive driver gene toward amplification, while linkage to toxic genes may prevent amplification of a true drive gene. Genes that are located near or that encompass fragile sites may be deleted, despite the fact that they do not confer a selective growth advantage (22). These factors, among others, may complicate any attempt to thoroughly catalog the list of cancer drivers from a computational analysis of gene alteration frequencies in tumors.

Drivers of Hereditary Breast and Ovarian Cancer

The French physician Pierre Paul Broca is traditionally credited with observing the first case of hereditary breast cancer: his wife acquired breast cancer at an early age, and he documented four generations of her family with cases of breast cancer (28). The heritability of multiple familial cancer syndromes soon led to the conclusion that cancer is fundamentally a genetic disease (29). In 1971, Alfred Knudson formalized the observation that onset of unilateral, sporadic cases of retinoblastoma occurs at a later age than onset of bilateral, familial cases of retinoblastoma, and that the kinetics of onset implies inheritance of a predisposing mutation and development of a second, somatic mutation, the so-called "two-hit model" (30).

A broad collection of tumor susceptibility syndromes was observed and characterized over time, including Lynch syndrome, von Hippel Lindau disease, Li-Fraumeni-syndrome, Cowden's disease, and neurofibromatosis among many others (31).

Sequencing of tumors began to catalog the causative mutations of these familial disease, beginning with *TP53* mutations underlying Li-Fraumeni syndrome (*32*). Inherited defects in DNA repair genes were found to cause a prominent group of these syndromes including ataxia telangiectasia, Bloom's syndrome, Fanconi's anemia, and xeroderma pigmentosum (*33*).

For hereditary breast and ovarian cancer (HBOC), two major genes have been associated with susceptibility: breast cancer susceptibility gene 1 (*BRCA1*) and breast cancer susceptibility gene 2 (*BRCA2*) (*34*). Mutation in either of these genes confers a lifetime 60-85% risk of developing breast cancer and 15-50% risk of developing ovarian cancer for women (*35-37*), and 1.2% (*BRCA1*) or 6.8% (*BRCA2*) risk of developing breast cancer for men (*28*). Unlike *BRCA1*, *BRCA2* mutation also predisposes to pancreatic and prostate cancer (*38*). These two breast cancer susceptibility genes share a crucial role in repair of DNA double-strand breaks (DSBs) by the high-fidelity repair process homologous recombination (HR).

Studies have estimated that approximately 25% of HBOC cases can be attributed to *BRCA1* or *BRCA2* mutation (*39*). Over 25 additional genes have shown less prevalent associations with HBOC, but many of these genes function in DNA repair and genome maintenance pathways related to *BRCA1* and *BRCA2* function. For example, among these genes are direct *BRCA1* and *BRCA2*-interacting proteins BRCA1-associated RING domain protein 1 (*BARD1*) and Partner and localizer of BRCA2 (*PALB2*), both of which participate in HR (*40, 41*). Also associated with HBOC risk

are *MRE11* and *NBN*, two components of the MRN complex, which resects DNA at DSBs before repair by HR can ensue (*42-44*). Finally, three paralogs of the RAD51 recombinase (*RAD51B*, *RAD51C*, and *RAD51D*) that mediate strand invasion during HR have also been associated with risk of HBOC (*45, 46*).

Recently, Knijnenburg *et al.* found that, across 33 tumor types, the majority of cancer predisposition genes affect genome integrity, with 64% being core members of the DNA damage response (DDR) (*47*). When germline and somatic mutations are pooled across all tumors, the most frequently mutated cancer predisposition genes are *BRCA1* and *BRCA2* (*48*). Unsurprisingly, patients who harbor germline *BRCA1* or *BRCA2* mutations develop tumors at a younger age than patients who develop mutations in these genes somatically. Across the pan-cancer cohort in this study, the mean age of diagnosis in patients with germline mutations was 54.4±13.0 years, compared to 62.3±13.4 years for cases with only somatic mutations (*48*).

Also released in the Pan-Cancer Atlas compendium of analyses, Huang *et al.* performed the largest systematic investigation of germline predisposition variants conducted to date, from 10,389 cases across 33 cancer types (*49*). 8% of cancers were found to carry a probable pathogenic germline variant, and 22.9% of cases from the Pediatric Cancer Genome Project contained one such variant. The majority of these germline variants were in TSGs, but rare variants were discovered in five OGs. Pathogenic and likely pathogenic germline variants of *BRCA1* and *BRCA2* were frequently discovered in ovarian and breast cancer, as expected, and germline

pathogenic variants of *BRCA2* were also found to be significantly enriched in pancreatic adenocarcinoma.

Although Knudon's original paper did not directly propose that the two "hits" of familial tumors occur in the same gene, his two-hit model has frequently been interpreted in this way. An alternative model proposes that most sporadic TSGs are haploinsufficient, and that loss-of-function alteration in a single allele of most TSGs would likely impart a significant phenotype (*15*). To speak to this point, Huang *et al.* also assessed the second allele in 154 cases with germline *BRCA1* or *BRCA2* mutation. Deletion of the wild-type allele was observed in ~38% of cases, and in 3 cases with germline *BRCA* mutations, distinct somatic mutations in the second allele were found (*49*).

II.   Hallmarks of Cancer

Hallmark Properties and Stress Phenotypes of Cancer

Despite the fact that tumors can arise from almost any tissue in the body, diverse forms of cancer ultimately grow to share certain phenotypic characteristics, collectively termed the "hallmarks of cancer" by Hanahan and Weinberg (*50*). They proposed the notion that most tumors acquire this shared set of hallmark features, and that the commonality of these phenotypic features implies a need for tumors to obtain these

capabilities during the course of their evolution. In a classic review, they described six

phenotypes that enable tumor growth and metastatic dissemination of cancer (*50*).

Perhaps the most foundational of these phenotypes is sustained proliferative

signaling. Cancer cells acquire a license to proliferate even in the absence of signals

which normally instruct cells to grow and divide. This self-sufficient license to sustain

chronic proliferation contributes to the bulk mass of cells that ultimately forms the

tumor; a palpable ~1cm^3 tumor consists of 10^9 cells and has undergone 32 population

doublings (for cells that are ~10μm in diameter) (*51*). One example of a genetic lesion

that licenses cells to proliferate in the absence of growth factors stimulation is

amplification or activating mutation of epidermal growth factor receptor (EGFR).

Second, in addition to sustained proliferative signaling, tumors also acquire the

ability to evade growth suppressive signals, including powerful growth control

mechanisms within the cell such as the p53 pathway, which ordinarily limits

proliferation in the presence of cellular stress. Third, tumors learn to evade apoptotic

signals, sometimes including evasion of a special form of apoptosis known as anoikis.

Anoikis is an apoptotic pathway that activates upon disruption of contacts between

epithelial cells and their underlying extracellular matrix (ECM), through which they

receive integrin-mediated growth signals (*52*).

Fourth, cancer cells must subvert the issue of telomere shortening which limits

their replicative potential, either through expression of telomerase or the alternative

lengthening of telomeres (ALT) pathway (*53*). Fifth, tumors acquire a heightened

ability to induce neovasculature through increased angiogenesis, supplying them with oxygen and nutrients to support their enhanced proliferative growth. And finally, cancer cells invade neighboring tissue and eventually spawn showers of micro-emboli that can disseminate in the circulation and colonize foreign tissue, forming metastases. The latter capability is responsible for 90% of cancer-associated mortality (*54*).

This original set of six phenotypic hallmarks has expanded to include the phenomenon of immune evasion, discussed in detail below (*55*). In addition, an additional set of five "stress phenotypes" have been proposed as hallmark characteristics of cancer (*56*). In contrast to the six classical hallmark phenotypes, which are prerequisite skills acquired by the tumor to promote oncogenic development, the five stress phenotypes are intrinsic forms of stress that are endemic to the tumorigenic state. They include mitotic, oxidative, proteotoxic, metabolic, and DNA damage stress, and result from both the demands and consequences of unbridled proliferation. Phenotypes such as aneuploidy, increased reactive oxygen species (ROS), and hypoxia create these stresses, which require tolerance through the enlistment of stress support pathways.

Genomic Instability Fuels Tumor Evolution

When outlining the six original hallmark features of cancer, Hanahan and Weinberg noted that these phenotypes represent acquired capabilities, but there is also one shared characteristic of tumors that instead *enables* tumorigenesis: increased genomic instability (*50*). Enhanced mutagenic capacity continuously provides a

background of greater genetic and epigenetic heterogeneity from which clones with increased fitness can arise. From this landscape of alterations, mutations that confer a selective advantage spawn expansion of clones, resulting in bursts of clonal growth that occur in a step-wise process, in the context of a Darwinian-like evolutionary setting (*57*).

Tumors acquire increased genomic instability through alteration of so-called "caretaker" genes: tumor suppressor genes that safeguard the genome from alteration (*58-60*). Caretaker genes encode proteins that detect DNA damage and activate an appropriate repair response, that actively participate in repairing DNA, that maintain telomeres, and that intercept mutagenic molecules with the potential to damage DNA (*61-63*). A similar concept, the classic "mutator hypothesis," postulates that loss of DNA repair genes leads to genomic instability, which enhances mutation rates and promotes cellular transformation (*64, 65*). Though these proposals predated the accumulation of abundant tumor sequencing data, current evidence supports their predictions, as mutation in a core DNA damage repair gene correlates with a significantly higher tumor mutational burden (*48*). Evidence suggests that alteration of genes involved in the DNA repair pathway occurs early and frequently tumorigenesis, with up to 40-50% of many cancer types exhibiting alteration in one or more of these genes (*66*).

In addition, mutations in DNA repair proteins have been linked to a variety of hereditary cancer predisposition syndromes, including *BRCA1* and *BRCA2* mutations that frequently occur in hereditary breast and ovarian cancer, as discussed above. In

addition, germline mutations in a number of other DNA repair genes have been linked

to familial cancer syndromes (*29*); for example, mutation in genes involved in the

nucleotide excision repair pathway (*XPA-G*) cause xeroderma pigementosum (*67*).

Mutations in genes that aid in repair of DNA double-strand breaks (DSBs) and

intrastrand cross-links, including Werner syndrome helicase (*WRN*), Bloom syndrome

helicase (*BLM*), RecQ protein-like 4 (*RECQL4*), and the Fanconi anemia genes, have

been linked to the development various cancers including leukemias and lymphomas (*68-*

70).

Similarly, germline mutations in genes encoding members of the mismatch repair

(MMR) pathway (*MSH2, MSH6, MLH1*) predispose to hereditary nonpolyposis

colorectal cancer (HNPCC), also known as Lynch Syndrome, and lead to a specific type

of genomic instability known as microsatellite instability (MSI, also known as MIN) (*71,*

72). Microsatellite instability refers to a DNA sequence feature apparent in the genomes

of MSI tumors: expansion or contraction of the number of oligonucleotide repeats

present in microsatellite sequences (*71, 72*). This feature of MSI-tumors highlights a

broader theme: alteration in DNA repair genes can create signature patterns of genomic

alteration that can be discerned from tumor sequencing data. For example, tumors with

increased APOBEC activity display frequent cytidine conversion at TpC dinucleotides

(*73*). Alternatively, the genomes of *BRCA1* and *BRCA2*-mutant tumors exhibit larger

and more frequent insertions and deletions (indels), with microhomology regions on

either side of the deletion breakpoint region (*74*); this feature may reflect the use of

microhomology-mediated repair (MMEJ), an alternate DSB repair pathway to

homologous recombination (HR) discussed below.

The DNA Damage Response (DDR)

One of the five stress phenotypes of tumors is genotoxic stress, which threatens

the integrity and informational content of DNA. The deoxyribonucleic acid (DNA)

macromolecule is a polymer of nucleotides: organic compounds consisting of five-carbon

deoxyribose sugars, aromatic nitrogenous bases, and phosphate groups. The order in

which these nucleotides self-assemble in a polymeric format encodes the information

required to coordinate cellular activities and generate life from gametes. Thus, it is vital

that this information is preserved with high fidelity as it is replicated during cellular

mitosis and meiosis.

Despite the inherent structural stability in its design, DNA is subject to a

constant onslaught of both endogenous and exogenous damage, from a variety of

sources. Endogenous damage can result in non-enzymatic hydrolytic depurination,

dNTP misincorporation during DNA replication, cytosine deamination, alkylation of

DNA bases, and oxidation of DNA bases by reactive oxygen species (ROS) (*75*).

Alternatively, exogenous sources of damage can produce a broad range of additional

insults. For example, ultraviolet (UV) light from the sun can induce pyrimidine dimers

and 6-4 phosphoproducts, at an estimated rate of 10^5 DNA lesions per cell per day (*76*).

Figure 1-2. Sources of Damage, DNA lesions, and Repair Pathways. DNA damage is overall categorized as either single-strand damage (orange) or double-strand damage (green). Within these categories, types of lesions are listed along with possible exogenous and endogenous sources of damage that cause them. Lesions are depicted graphically and the pathways utilized to repair the relevant lesions are listed.

Cigarette smoke causes bulky aromatic DNA adducts (*77*), and ionizing radiation (IR) from cosmic radiation and X-rays can generate single-strand breaks (SSBs) and double-strand breaks (DSBs) in DNA. Lastly, chemotherapeutic agents can induce either kind of DNA break (camptothecin and etoposide), alkylate DNA (methyl methanesulfonate and temozolomide), or chemically crosslink DNA (mitomycin C) (*61*).

Fortunately, a battery of coordinated cellular programs exists to recognize, repair, and respond to each of these types of lesions, broadly called the DNA damage response (DDR). The DDR encompasses a number of repair pathways, and the type of

DNA damage that occurs prompts activation of the appropriate pathway response

(Figure 1-2). For example, misincorporation of incorrect dNTPs during replication

results in base-base mismatches. The mismatch repair (MMR) system, mutated in MSI-

tumors as discussed, recognizes base-base mismatches (usually in the post-replicative

setting) and repairs this specific type of DNA lesion (*78*).

Small lesions confined to a single base, such as alkylated or oxidized bases,

activate the base excision repair (BER) pathway. An appropriate glycosylase (such as

OGG1, *UNG*, *NEIL1-3* or *NTHL1*) excises the damaged base, generating an apurinic

(AP) site that stimulates several steps of downstream processing, terminating in

resolution by ligation of a single-stranded nick (*79, 80*). If the lesion encompasses

several bases, such as in pyrimidine dimers or intrastrand cross-link lesions, the related

nucleotide excision repair (NER) pathway removes an oligonucleotide of approximately

30 base pairs containing the damaged lesion (*81*).

All of the repair mechanisms mentioned above can utilize the paired undamaged

DNA strand as a faithful template for repair, since they involve damage to a single

strand of the DNA molecule. Thus, repair of double-stranded DNA lesions is

intrinsically more complicated; it must involve pairing with the undamaged sister

chromatid in order to retain information from the damaged lesion with high fidelity.

Though the downstream steps of this process are more involved for high-fidelity repair

processes, repair of double-strand lesions operates within the same logical framework as

repair of single-strand lesions: pathway steps involve sensing of the DNA lesion,

activation of master-regulators and signal transducers, and recruitment of effectors. Central transducers of the DDR in mammalian cells include three members of the phosphatidylinositol-3-kinase-like protein kinase (PIKKs) family: ATM (ataxia-telangiectasia mutated), ATR (ATM- and Rad3-Related), and DNA-PKcs (DNA-dependent protein kinase catalytic subunit) (*82-84*).

Of the many lesions that can threaten the integrity of the genome, the double-strand break (DSB) is one of the most detrimental. Repair of DSBs in mammalian cells can proceed through several pathways, including classical non-homologous end-joining (c-NHEJ), microhomology-mediated end joining (MMEJ), single-strand annealing (SSA) or homologous recombination (HR) (*61, 85*). While c-NHEJ is a template-free process that involves minimal end processing of DSBs and potentially inaccurate religation, HR coordinates pairing of the DSB-bearing strand with its intact sister chromatid, precisely restoring the original genomic sequence. The hereditary breast and ovarian cancer susceptibility genes *BRCA1* and *BRCA2*, mentioned above, play a crucial role in HR: they facilitate exchange of the ssDNA-coating heterotrimer RPA for RAD51 recombinase.

Multiple layers of regulation influence the method of DSB repair selected by the cell; for example, pathway choice is tightly coupled to cell cycle, because resection of double-strand breaks by the MRN complex (MRE11-RAD50-NBS1) is restricted to S/G2 phase (*86, 87*). Similarly, 53BP1 negatively regulates HR by antagonizing resection *(88)*. While c-NHEJ acts to rapidly rejoin non-resected DSBs and is favored

in G1, HR and MMEJ act on resected DSBs that are generated in S/G2 phase (*89*). Of these pathway alternatives, MMEJ is the least fully characterized pathway; it involves direct annealing of regions of microhomology on either side of the DSB with the aid of polymerase theta (POLƟ) (*90*). Interestingly, MMEJ and HR are competitive pathways that exhibit synthetic lethality when co-mutated; not only do they share the same starting substrate, but POLƟ has also been shown to directly suppress HR (*91*).

Tumor Neoantigens and the Antitumor Immune Response

Although mutations in DDR genes and concomitant genomic instability generally promote tumorigenic development, they do impart one advantageous corollary. The increased mutational burden associated with DDR gene mutation has been shown to increase the diversity of neoantigens available to the immune system, which promotes an antitumor immune response (*92*). In the extreme, MSI-tumors with particularly high mutational burdens respond well to immunotherapy (*93*). How does an increased complement of neoantigens provoke a tumor-targeted response from the immune system?

Activated CD8[+] T-cells can identify and destroy aberrant cellular targets, such as virally-infected cells, by engaging the target cell's MHC I-antigen complex with their T cell receptors (TCRs). A complex set of co-stimulatory and co-inhibitory inputs determines the output of MHC-TCR engagement, resulting in either T-cell mediated killing of the target cell (via perforin-delivered granzyme) or survival of the target.

Accurate interpretation of this balance of costimulatory and coinhibitory signals is crucial to the maintenance of self-tolerance and the efficacy of T-cell based immunity.

The atypical, aberrant cellular targets that CD8+ T cells can become licensed to destroy include not only virally-infected cells, but also cancer cells. The latter concept was unclear for many years, because athymic (nude) mice do not exhibit an increased incidence of spontaneous or carcinogen-induced tumors (*94*). However, athymic mice retain many components of innate immune function, and improved mouse models of immunodeficiency revealed that the immune system not only controls growth, but also shapes the immunogenicity of tumors (*95-99*).

Despite the potential for T-cell based immunity to control tumor growth, tumors frequently evade immunosurveillance (*100*); this behavior was proposed as the seventh hallmark feature of cancer by Kroemer and Pouyssegur (*55*). Among other tactics, tumors downregulate antigen presentation machinery and upregulate cell surface-expressed molecules that deliver coinhibitory signals to T cells. With some spatial and temporal variation, T cells receive these coinhibitory signals through their cell surface-expressed molecules CTLA-4 (cytotoxic T lymphocyte antigen-4) and PD-1 (programmed cell death protein 1); appropriate delivery of these signals is critical for self-tolerance (*94*).

CTLA-4 binding to CD80 and CD86, which are expressed mainly in lymphoid tissues, results in critical anti-inflammatory effects during initial T cell priming in lymphoid organs. Similarly, T cells upregulate PD-1 after activation and maintain its

expression through repeated exposure to antigen and T cell exhaustion. PD-1 interacts with two main ligands: PD-L1 (Programmed death ligand 1; *CD274*) and PD-L2 (Programmed cell death 1 ligand 2; *CD273*) (*101*). While PD-L2 expression is primarily restricted to myeloid cells such as dendritic cells, macrophages, and mast cells, PD-L1 expression is broad, ranging from lymphoid and myeloid cells to several types of non-hematopoietic healthy tissue (*102*), and it appears to be more ubiquitously expressed in tumors than PD-L2 (*103, 104*). Upregulation of immune evasion molecules such as PD-L1 is a mechanistic example of the seventh hallmark behavior of cancer: the ability to evade detection and destruction by the immune system.

III.   Challenges and Paradigms in Cancer Therapeutics

Opening a Therapeutic Window

It is a relatively simple task to identify a small organic molecule capable of killing cancer cells. In fact, 0.1-1% of small molecules present in a common pharmaceutical compound library are toxic to cancer cells within the concentration ranges typically tested in high-throughput screening (*105*). Rather, the key challenge in chemotherapeutic development is the task of selectively killing cancer cells while sparing the healthy, wild-type counterpart cells from which they are derived.

Classic cytotoxic cancer treatments, such as IR, DNA alkylating agents, and mitotic poisons, differentiate between cancerous and normal tissue by the crude measure

of proliferative rate. This distinction is very poor, and inevitably leads to severe side-effects in all of the rapidly proliferative normal tissues of the body (hair, skin, bone marrow, and gastrointestinal tissue). Targeting rapidly proliferative cells is a strategy that offers a narrow therapeutic window (the concentration range over which therapeutic effects would be expected) and represents a low therapeutic index (the dose required for toxic effects divided by the dose required for therapeutic effects) (*106*).

Early examples of chemotherapy targeted toward specific tumor types utilized clinical evidence to guide therapeutic strategies. For example, sex steroids were found to drive proliferation of reproductive tissue; thus, hormone-ablation therapy evolved for breast cancer, first by oophorectomy, then through administration of estrogen receptor antagonists such as tamoxifen (*107*). Similarly, Sidney Farber observed that children with acute lymphoblastic leukemia exhibit low serum folic acid levels, leading him to postulate that their leukemia cells may be particularly reliant on folic acid (*108*). This hypothesis led to the design and implementation of antifolate therapy, predating the discovery of either the target (dihydrofolate reductase, DHFR) or mechanism of the drug (*109*).

Discovery and mechanistic understanding of cancer driver genes significantly broadened the potential therapeutic window of chemotherapeutic reagents, by offering the opportunity to target individual genetic lesions that are distinct to tumor cells. Less than two decades after the discovery of the Philadelphia chromosome (a chromosomal translocation event that produces a fusion between the genes *BCR* and *ABL1*) (*110*),

the c-Abl kinase inhibitor imatinib (Gleevec) became the first successful example of gene-specific targeted therapy (*111*). Gleevec produced striking, stable remissions in 90% of BCR-ABL+ CML patients in initial trials (*111*); however, a proportion of patients became resistant to the drug, most commonly through development of secondary mutations in the Abl kinase that rendered them resistant (frequently the T315I mutation) (*111*). Nonetheless, long-term outcomes of imatinib treatment for CML remain very good, with an estimated overall survival rate at 10 years of 83.3% and fewer side-effects than traditional chemotherapeutic treatments (*112*).

Acquired Resistance to Targeted Therapies

With anticipation of equal degrees of success, this principle was applied to develop therapies targeted toward driver genes in other tumor types. For example, Genentech developed a humanized monoclonal antibody to HER2, trastuzumab (Herceptin), to target HER2+ breast cancer (*113*). Similarly, because the epidermal growth factor receptor (EGFR) is overexpressed in a large proportion of non-small-cell lung cancer (NSCLC), efforts led to the development of a small molecule ATP-competitive EGFR tyrosine kinase inhibitor (TKI), gefitinib (*114*). Disappointingly, on the whole, gefitinib did not provide a substantial benefit to EGFR+ NSCLC patients (*115, 116*). However, about 10% of patients exhibited a rapid, dramatic clinical response; this subsection of patients harbored an activating mutation in EGFR, rendering them more sensitive to the inhibitor (*117*).

Despite this key insight, patients with activating EGFR mutations who did benefit from targeted therapy almost inevitably underwent relapse within one year of treatment, and sometimes within mere months (*118, 119*). Most often, this resistance occurred due to secondary mutation in EGFR or maintenance of signaling through the pathway through other means (commonly through *MET* or *AXL* amplification) (*120*). Although EGFR-mutant cells that respond to EGFR TKIs have become reliant on EGFR signaling for growth, a phenomenon known as "oncogene addiction," there is sufficient intrinsic heterogeneity within most tumors that drug-resistant clonal evolution occurs rapidly in the setting of targeted therapy (*121, 122*). Upon reflection, perhaps the fact that CML generally exhibits lower intratumoral heterogeneity underlies the singular success of Gleevec, which has not been recapitulated in any solid tumor setting; perhaps the tumor, rather than the target, explains the special success in this case.

The inevitable development of resistance to targeted therapies appears to be a pervasive theme. Another example is the BRAF inhibitor vemurafenib, which produced stunning initial responses, but fell victim to the same frequency of relapse and resistance as EGFR TKIs (*123*). For patients who relapse on targeted TKI treatment, second-line therapies have been developed to target common mechanisms of resistance. For example, for EGFR, second- and third-line therapy includes TKIs targeting the most common gefitinib-resistant mutant form of EGFR, the so-called "gatekeeper" T790M mutation (*124, 125*). It has been suggested that "rational upfront polytherapy,"

concurrent targeting of multiple critical signaling pathways, may be the only way to prevent or delay resistance to targeted therapy (*126*).

Instead of targeting oncogenic lesions that are tumor-specific, an alternative therapeutic strategy is to target tumor-specific stress pathways that are not themselves altered or mutated in the tumor. This strategy refers to targeting non-oncogene addiction (*127*): tumors become specifically dependent on certain cellular activities and stress support pathways, which can serve as drug targets (*56*). Either overloading the stress pathways by increasing tumor-specific stress, or inhibiting stress support mechanisms, represent viable non-oncogene addiction therapeutic targeting strategies. For example, bortezomib (Velcade®) targets the proteasome in multiple myeloma; this antibody-secreting plasma cell tumor intrinsically experiences significant proteotoxic stress and requires proteasome function for homeostasis and survival (*128*).

Targeting non-oncogene addiction exploits the principle of, to borrow a term from yeast and fly genetics, synthetic lethality. Two genes are "synthetic lethal" if mutation of either gene is viable, but mutation of both genes simultaneously is lethal (*106*). This principle can be extended to target not only non-oncogene addiction pathways, but also "undruggable" oncogenic lesions. For example, the transcription factor Myc is among the most highly mutated oncogenes in tumors, but inhibition of its activity with small molecules has proved challenging (*129*). Thus, genetic screening efforts have attempted to identify drug targets that are synthetic lethal with Myc: inactivation of these genes is

25

lethal in the context of mutated Myc, but viable in the setting of wild-type Myc (*130*).

Similar efforts have been performed for *KRAS* (*131, 132*) and *PIK3CA* (*133*).

In a clinical setting, this therapeutic strategy has been most successfully applied to the treatment of *BRCA1* and *BRCA2*-deficient breast and ovarian cancer patients with inhibitors of poly(ADP-ribose) polymerase (PARP). Though it was long believed that PARP inhibitors exert their synthetic lethal effects in *BRCA*-deficient cells through inhibition of PARP enzymatic function, recent work has shown that PARP inhibitors physically trap PARP onto DNA, forming toxic complexes that stall replication forks and require HR for resolution (*134-136*). Five PARP inhibitors are in clinical trials for *BRCA*-mutant or ovarian cancer and have shown clinical benefit in these settings (*137-139*), but as with targeted therapies, patients commonly acquire resistance to PARP inhibitors (*66, 140*). Often these resistance mechanisms include restoration of the HR pathway, for example through reversion mutations in *BRCA* that restore its functionality, or decreased expression of 53BP1 (*141, 142*).

The Promise of Immunotherapy

In contrast to all of the aforementioned therapeutic strategies, immunotherapy represents an entirely different mechanism of promoting tumor clearance. Rather than directly exerting toxicity in tumor cells, immunotherapeutic agents instead harness the immune system's natural potential to impart that toxicity in a highly specific manner. This highly promising group of therapeutics encompasses a broad variety of technologies

including cancer vaccines, engineered T cells, and antibodies that block coinhibitory

pathways which inhibit T cell-mediated immune responses to tumors.

In the latter category, monoclonal antibodies that block either CTLA-4 or the

PD-1-PD-L1 signaling axis have shown remarkable potential as therapeutic agents.

Examples of success in murine models led first to the clinical development of anti-

CTLA-4 blocking antibodies as cancer therapeutics (*143, 144*). Ipilimumab, an anti-

CTLA-4 monoclonal antibody, produced a significant overall survival advantage in

metastatic melanoma clinical trials when administered with a peptide vaccine,

generating great excitement about this treatment modality (*145*). Later, CTLA-4

exhibited efficacy as a single agent in non-small cell lung cancer (NSCLC) (*146*),

hepatocellular carcinoma (HCC) (*147*), colorectal cancer (*148*), and mesothelioma (*149*).

The most striking property of anti-CTLA4 blockade in clinical trials has been the

proportion of patients who achieve curative durable remission on this treatment

regimen; among patients with non-resectable or metastatic melanoma, 22% responded

durably, with their complete response extending for the ten years of data available in a

long-term survival study (*150*).

Similarly, blockade of PD-1-PD-L1 interaction with monoclonal antibodies to

either PD-1 or PD-L1 has produced remarkable success in a variety of cancer types,

leading to over 1500 clinical trials and FDA-approval of its treatment for NSCLC (*151-*

157), melanoma (*158-164*), MMR-mutated (MSI) colorectal cancer (*165-167*), RCC

(*168*), Hodgkin's lymphoma (*169, 170*), head and neck squamous cell carcinoma

(HNSCC) (*171, 172*), Merkel-cell carcinoma (*173*), HCC (*174*), urothelial cancer (*175-181*), and gastric cancer (*182*). Anti-PD-1 and anti-PD-L1 monoclonal antibodies produce fewer adverse inflammatory reactions than anti-CTLA-4 antibodies (*183, 184*), and thus have been approved for first-line use in NSCLC (*185*), melanoma, urothelial cancer, and Merkel cell carcinoma (*186, 187*). Addition of pembrolizumab to standard chemotherapy as first-line therapy augments both overall and progression-free survival compared to standard chemotherapy alone in NSCLC (*188*). However, the clinical activity of these antibodies does not extend to all tumor types, with responses in breast, ovarian, non-MSI colorectal, and prostate cancer being limited (*189, 190*).

Within tumor types that do respond to anti-PD-1 or anti-PD-L1 therapy, the frequency of complete response (CR), partial response (PR), and overall response rate (ORR) varies widely, with the latter ranging from 14.3 to 69% (*101*). This observed heterogeneity in response rates has prompted significant efforts to determine predictive biomarkers of response to PD-1-PD-L1 axis blockade. From these efforts, several themes have emerged: determinants of response include "foreignness" of the tumor (i.e. the diversity of tumor neo-antigens) (*191, 192*), the presence of a tumor T cell infiltrate (*193*), the presence of intact antigen presentation machinery in the tumor (*194*), and the expression of PD-L1 on tumor cells (*152, 154, 161-163, 180*). In some contexts, PD-L1 expression on infiltrating T cells themselves also possesses predictive value (*195-197*).

Immunohistochemical staining for PD-L1 expression (PD-L1 tumor proportion score, TPS) appears to be the strongest of these predictive biomarkers for many tumor

types, though it is an imperfect marker: positive responses can occur in the absence of PD-L1 staining, and in other cases PD-L1-positive tumors fail to respond to immunotherapy (*101, 198*). There is some evidence that secondary immunosuppressive effects in the tumor microenvironment, such as the effect imposed by TGFβ, may explain PD-L1+ unresponsive tumors (*199*). And for unclear reasons, immunohistochemistry (IHC) for PD-L1 expression is not strongly associated with clinical benefit in HNSCC or RCC (*168, 200*).

Efforts to improve understanding of discrepancies between PD-L1 tumor staining and clinical response to immune checkpoint blockade will likely improve the efficacy of this forefront therapy. In addition, a more comprehensive understanding of mechanisms of acquired and innate resistance to immunotherapy will aid in patient stratification and in supplementation with adjunct therapies. In a study of melanoma patients who developed resistance to PD-1-blocking antibodies, the most commonly occurring mutations found to promote resistance occurred in components of the IFNγ signaling pathway (JAK2, STAT1) and a component of antigen presentation machinery, β2-microglobulin (*201*). Despite these cases of acquired resistance and the diversity in responses observed across various tumor types, immunotherapy currently represents the promising avenue toward development of potentially curative chemotherapeutic treatments.

Chapter 2: Genetic Screens Reveal APEX2 and FEN1 as BRCA2 Synthetic Lethal Targets

Attributions

Work presented herein is currently under review:

Kristen Mengwasser[1], Yumei Leng[1], Mei Yuk Choi[1], Connor Clairmont[2], Alan D'Andrea[2] and Stephen J. Elledge[1]

[1]Howard Hughes Medical Institute, Department of Genetics, Harvard Medical School, Division of Genetics, Brigham and Women's Hospital, Boston, MA

[2]Department of Radiation Oncology, Dana-Farber Cancer Institute, Boston, MA 02215

I. Introduction

Damage to human genetic material frequently contributes to oncogenic transformation, and it threatens faithful transmission of genetic information across generations. In response to the wide variety of lesions in DNA caused by endogenous and exogenous sources of damage, cells have evolved a complex signal transduction pathway known as the DNA damage response (DDR). The DDR senses DNA damage, recruits a coordinated set of factors to repair the lesion appropriately, and executes a coordinated cellular response which may include immune system activation, cell cycle arrest, cellular senescence or cell death (*61, 202*).

Among the many lesions that occur in DNA, the double-strand break (DSB) is one of the most detrimental. Failure to properly repair DSBs can lead to a variety of adverse outcomes such as replication fork collapse, cell death, oncogenic translocation, or loss of telomeric DNA. Three main pathways exist to repair DSB lesions: homologous recombination (HR), classical non-homologous end-joining (NHEJ), and microhomology-mediated end-joining (MMEJ). Whereas HR utilizes information from a paired sister chromatid to seamlessly repair a DSB with high fidelity, NHEJ and MMEJ can result in sequence alterations at the repaired DSB site. However, unlike NHEJ, HR and MMEJ both act on resected DSBs, and these two pathways both complement and directly compete with one another (*91, 203*).

Several genes that play a crucial role in DSB repair are frequently mutated in cancer. Germline mutations in *BRCA1* or *BRCA2*, both important players in the HR

31

pathway, account for the majority of hereditary breast and ovarian cancer (*204-206*). Women carrying mutations in one of these two tumor suppressor genes have up to an 80% risk of developing breast cancer and up to a 50% risk of developing ovarian cancer (*206, 207*). Overall, up to 50% of all high grade serous ovarian cancers (HGSOC) have detectable germline and somatic inactivation of HR genes, with 30% of these alterations being *BRCA1* or *BRCA2* (*140*). And unlike *BRCA1*, *BRCA2* mutation also predisposes to pancreatic and prostate cancer (*38*).

Although BRCA1 and BRCA2 share a common function in facilitating HR, BRCA2 is known to play several additional roles in safeguarding genomic integrity. By loading Rad51 onto DNA breaks and gaps, BRCA2 prevents Mre11-dependent degradation of nascent DNA at stalled replication forks (*208-211*), and with *BRCA1* promotes HR-mediated resolution of fork stalling (*211*). Also, BRCA2 protects telomere integrity (*212*) and prevents accumulation of R-loops, which can lead to replication fork stalling and interference with transcriptional elongation (*213*).

In part due to a synthetic lethal (SL) relationship, *BRCA1* and *BRCA2*-deficient cells are exquisitely sensitive to treatment with inhibitors of poly(ADP-ribose) polymerase (PARP) (*214, 215*). Synthetic lethality refers to a concept borrowed from classical genetics; it describes a condition in which mutation of either of two genes is viable, but simultaneous mutation of both genes is lethal (*106*). In this case, inactivation of *BRCA1* or *BRCA2* renders cells sensitive to PARP inhibition. PARP inhibitors have recently gained FDA approval for their success in a Phase III clinical

trial (*137, 138*) for patients with metastatic breast cancer and a germline *BRCA* mutation, and are FDA-approved for the treatment of recurrent HGSOC (*139, 140*).

However, dual depletion of *PARP* and *BRCA2* by siRNA does not recapitulate the potent lethality observed upon chemical inhibition of PARP (*214*). Rather than solely exploiting a genetic SL relationship, PARP inhibitors also cause lethality by physically trapping PARP onto single-strand break (SSB) intermediates, obstructing progression of replication forks (*134-136*), and in that sense behaving more like classical DNA damage agents to which *BRCA2*-mutant tumors are also sensitive. Indeed, the toxic PARP-DNA complexes formed as a result of PARP trapping have been shown to be more cytotoxic than unresolved SSBs (*136*).

Targeting the intrinsic genetic vulnerabilities in *BRCA1* and *BRCA2*-deficient cells offers the opportunity to meet an urgent clinical need in the treatment of refractory or metastatic breast cancer in patients with a germline *BRCA* mutation (*216*) and recurrent HGSOC more broadly (*139, 217*). Despite recent success in clinical trials, PARP inhibitor efficacy appears to be limited by inherent and acquired resistance, underscoring the urgent need for identification of synergistic and alternative targets (*66*). Because the sensitivity of *BRCA*-deficient cells to PARP inhibition is not due solely to synthetic lethality, we sought to explore if additional genetic synthetic lethal relationships exist with *BRCA2* deficiency. We chose *BRCA2* for this study because of its myriad important roles in protecting genomic integrity beyond its crucial role in HR.

To uncover novel *BRCA2* synthetic lethal genes (B2SLs), we used a genetic screening approach, studying both shRNA and CRISPR-based genetic libraries in a pooled screening format, in two pairs of isogenic cell lines. We found *BRCA2* mutant (B2MUT) cells to be more dependent than their wild-type counterparts (B2WT) on several pathways including base excision repair (BER), ATR activation, and MMEJ. We identified *APEX2* and *FEN1* as novel B2SL targets, and we showed through the use of a novel cell-based reporter that *FEN1* participates in MMEJ.

II.   Results

shRNA and CRISPR screens identify B2SL Candidates

To identify novel B2SL candidates, we began by establishing a pair of cell lines that are isogenic except for the presence or absence of a *BRCA2* mutation. We obtained a modified DLD-1 colon cancer cell line with a homozygous deletion of BRC repeat 6 in exon 11 that also introduces a loxP site and a stop codon between BRC repeats 5 and 6, resulting in a biallelic *BRCA2* premature truncation mutation (*218*). To this *BRCA2* mutant (B2MUT) cell line, we introduced a full-length *BRCA2* mammalian expression construct through transfection and selection for stable integrants. These add back *BRCA2* wild-type cells are a closer isogenic comparison to B2MUT cells

Figure 2-1. Establishment of isogenic cell line systems for B2SL screening. (A) Extracts from the indicated cell lines untreated or treated with the indicated siRNAs were immunoblotted with antibodies to BRCA2 and GAPDH. **(B)** Immunofluorescence with antibodies to γH2AX and Rad51 protein was performed on cells of the indicated genotypes to evaluate foci formation after 10Gy IR. **(C)** The indicated cell lines were passaged in the presence of the indicated dose of olaparib or DMSO; after three days cell survival was quantified using CellTiter-Blue. **(D)** Schematic diagram depicting the experimental procedure for B2SL screening in isogenic BRCA2 cell lines. **(E)** Schematic of wild-type BRCA2 structure, depicted with its functional domains and sites of interaction with key binding partners. BRCA2 truncation mutant proteins possessed by the colonic and ovarian BRCA2 MUT cell lines used in this study are shown for comparison. Type and location of all *BRCA2* mutations observed in the METABRIC breast cancer study are diagrammed.
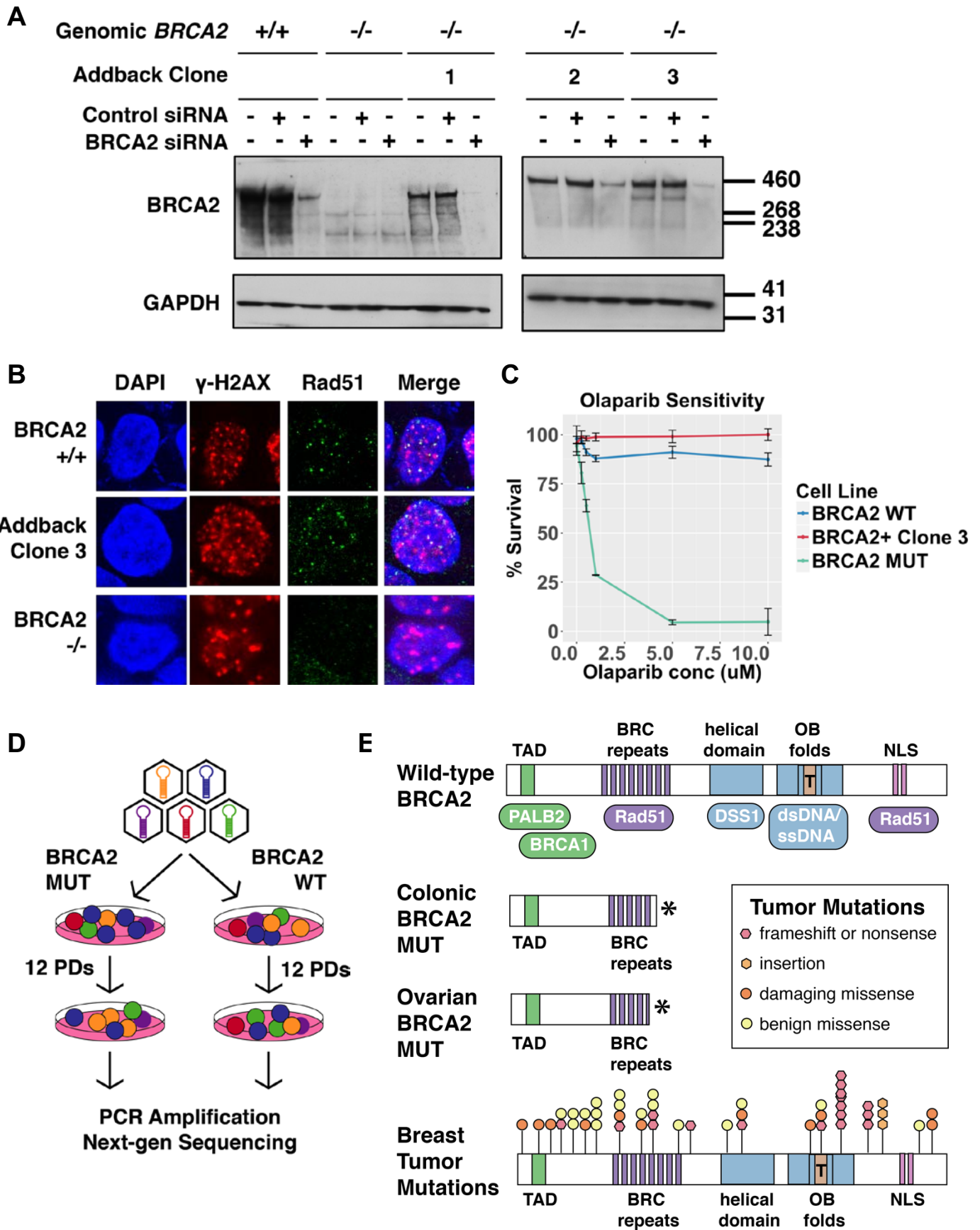
Figure 2-1 (Continued). Establishment of isogenic cell line systems for B2SL screening.

than the parental DLD-1 line, due to the genetic drift that occurs in this mismatch repair (MMR)-deficient background.

We isolated individual clones from these *BRCA2* wild-type cells (B2WT) and characterized several clones to demonstrate restoration of functional BRCA2 expression. We confirmed full-length BRCA2 protein expression by Western blotting, utilizing BRCA2 siRNA to confirm the identity of the protein (Figure 2-1A). We observed that expression of full-length BRCA2 enhanced the growth rate of B2MUT cells (Supplemental Figure S2-1A) and that expression of BRCA2 in our add back clones restored resistance to the PARP inhibitor olaparib (Figure 2-1C). Finally, we confirmed that adding back full-length BRCA2 to B2MUT cells restored their ability to form Rad51 foci in response to ionizing radiation (IR) (Figure 2-1B).

Using this pair of isogenic cell lines, we performed a series of genetic screens designed to identify novel B2SL genes. We screened a targeted library of 380 genes with a known or suspected role in the DDR; inclusion of genes in this library was based on either a role in DNA repair described in the literature or performance of the gene in prior screens for sensitivity to DNA damaging agents (unpublished data). In each screen, B2MUT and B2WT lines were separately transduced with either an shRNA-based or CRISPR-based library targeting these DDR genes, at a low multiplicity of infection (MOI) in triplicate (Figure 2-1D). Cell pellets were collected before and after passaging for 12 population doublings (PDs), and the relative change of shRNAs or

gRNAs during culture was determined by next-generation sequencing (NGS) of shRNA half-hairpins (*219*) or gRNAs PCR-amplified from PD0 and PD12 cell pellets.

To explore the generalizability of our results, we also obtained a second previously characterized isogenic cell line pair: ovarian PEO1 B2MUT cells, which contain a hemizygous truncation mutation in BRC repeat 5 of exon 11, and an *in vitro*-derived cisplatin-selected clone C4-2 which contains a reversion mutation that restores full-length BRCA2 expression (B2WT) (*220*). We note that the colonic DLD-1 B2MUT cell line and the ovarian PEO1 B2MUT cell line express similar truncation mutants of BRCA2 (*218, 220*), both prematurely terminated directly in or immediately after BRC repeat 5 in exon 11 (Figure 2-1E). This truncation mutant retains the N-terminal transactivation domain of BRCA2 (TAD) and the first four BRC repeats, which are known to bind free RAD51 (*221*), but lacks the DNA binding domains of BRCA2 and its C-terminal Rad51 binding domain (Figure 2-1E). We examined all BRCA2 mutations reported from 2,433 tumors in the METABRIC breast cancer targeted exome sequencing study (*222*) and found that mutations in the BRC repeat region of exon 11 are not uncommon, with about 16% of functionally impactful mutations occurring in this region (Figure 2-1E). We screened our CRISPR-based library, but not our shRNA-based library, in this ovarian isogenic cell line pair (Supplemental Figure S2-1B).

We analyzed the results of our screens by MAGeCK (*223*) and edgeR (*224*), relying on edgeR analysis to calculate reported false discovery rates (FDRs) (Figures 2-2A-F). B2SL genes drop out in the B2MUT cell line more than the B2WT cell line,

Figure 2-2 BRCA2 Synthetic Lethal Screen Results. (A-C) Volcano plots for the shRNA colonic, CRISPR colonic, and CRISPR ovarian B2SL screens. Significance (-$\log_{10}$FDR) is plotted against the genetic interaction (GI) score (average log2-fold-change for each gene in BRCA2 MUT versus BRCA2 WT cells). Genes that met a significance threshold of -$\log_{10}$FDR>1 are color-coded as green for relative dropout or red for relative enrichment in BRCA2 MUT vs BRCA2 WT cells. **(D-F)** Results from the colonic shRNA, colonic CRISPR, and ovarian CRISPR B2SL screens, plotted as the log2-fold-change in BRCA2 MUT cells against the log2-fold-change in BRCA2 WT cells. **(G)** Comparison of GI score (average log2-fold-change for each gene in BRCA2 MUT versus BRCA2 WT cells) in the ovarian CRISPR screen versus the colonic CRISPR screen. **(H)** MCA assays were performed after infection with the indicated gRNAs in the indicated isogenic cell line pairs. GFP-labeled BRCA2 MUT cells were mixed with DsRed-labeled BRCA2 WT cells and relative growth was measured as the change in GFP+ cells by FACS after 12 days, normalized to negative control gRNA-expressing cells. The BRCA2 MUT line of one isogenic cell line pair (colonic) possesses a BRCA2 truncation mutant that terminates after BRC repeat 5, while the BRCA2 MUT line of a second isogenic cell line pair (pancreatic) possesses a longer mutant protein that terminates in BRC repeat 8.
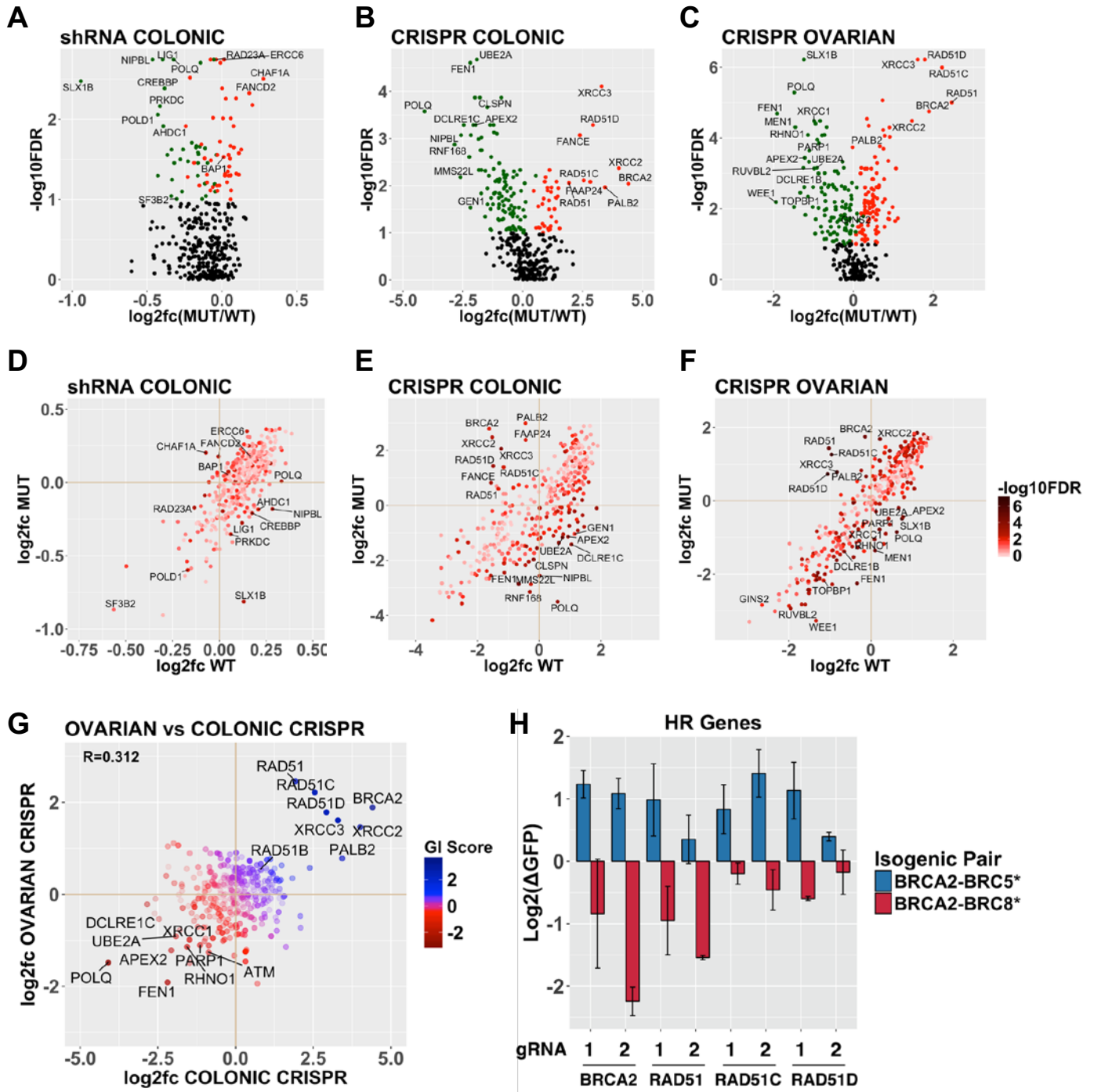
Figure 2-2 (Continued). BRCA2 Synthetic Lethal Screen Results.

and thus the fold-change in the B2MUT line is lower than the fold-change in the B2WT line, yielding a B2MUT/B2WT ratio that is negative on a log2 scale (GI score) (Figures 2-2A-C); this value can also be negative for genes that promote growth of B2WT cells more than B2MUT cells. Overall, the magnitude of effects we observed was more potent with our CRISPR-based than our shRNA-based library, and the CRISPR library produced larger fold-changes in the diploid colonic background than the aneuploid ovarian background.

Analysis of our primary shRNA screen revealed *POLQ* and *SF3B2* as B2SL candidates. *POLQ* encodes polymerase theta, a large A-family DNA polymerase that plays an important role in MMEJ and is known to be synthetic lethal with the HR pathway (*91, 203*). Its helicase domain promotes unwinding and exposure of microhomology regions on either side of a DSB, while its polymerase domain extends DNA from a region of microhomology base pairing (*225*). Thus, *POLQ* serves as a positive control for *BRCA2* synthetic lethality.

SF3B2 is part of the U2 snRNP that assembles with other snRNP components to form the spliceosome (*226*). While *SF3B2* loss is toxic to both B2MUT and B2WT cells, B2MUT cells appear to be more reliant on its function than B2WT cells. When we re-screened 50 hits from this library in a secondary shRNA screen, we observed that *SF3B2* and *POLQ* validated as B2SL candidates (Supplemental Figures S2-2A and S2-2B), and that our secondary shRNA screen sublibrary enriched for B2SL candidates overall (Supplemental Figures S2-1C-F).

Screening our CRISPR library in the colonic DLD-1 isogenic cell line pair confirmed *POLQ* as a B2SL gene and also found *FEN1*, *APEX2*, *UBE2A*, *CLSPN*, and *DCLRE1C* to behave as B2SLs (Figures 2-2B and 2-2E). Screening this same library in the ovarian PEO1 isogenic cell line pair revealed *POLQ*, *FEN1*, *XRCC1*, and *RHNO1* to be among the strongest B2SL hits (Figures 2-2C and 2-2F). Because we are most interested in genes that generalize as SL broadly, we compared ovarian and colonic cancer cell lines to find *POLQ*, *FEN1*, and *APEX2* as strong, consistent B2SL in both contexts (Figure 2-2G). While *POLQ* serves as a positive control, *FEN1* and *APEX2* represent novel B2SL genes and novel potential drug targets in *BRCA*-deficient tumors.

Notably, several of our B2SL hits demonstrated stronger synthetic lethality than *PARP1* in both ovarian and colonic cells, including *FEN1*, *APEX2*, *POLQ*, *XRCC1*, and *UBE2A*. This observation underscores the well-established finding that PARP inhibitor effectiveness in *BRCA2*-deficient cells is due not only to a mild synthetic lethal interaction between *PARP1* and *BRCA2*, but also because of the phenomenon of PARP trapping (*134, 136*).

Loss of early HR components promotes growth of cells with *BRCA2* truncation mutations

From our comparison of CRISPR screens in the ovarian and colonic cell line pairs, we noticed a striking phenomenon: every RAD51 paralog (*RAD51*, *RAD51B*, *RAD51C*, *RAD51D*, *XRCC2* and *XRCC3*) exhibited a strongly positive genetic

interaction (GI) score in both cell line contexts (Figure 2-2G, Supplemental Figure S2-2C). This positive GI score (the average log2-fold-change in B2MUT versus B2WT cells) reflects not only a detrimental effect in B2WT cells from the loss of these HR components, but also a surprising enhancement of B2MUT cell growth (Figures 2-2E and 2-2F). We hypothesized that the residual *BRCA2* truncation mutant present in both ovarian and colonic B2MUT cells may somehow be exerting a toxic neomorphic effect, involving its remaining intact interaction with free Rad51. When we examined our data for the performance of *BRCA2* itself we found that, like Rad51 and its paralogs, CRISPR-based knockout is detrimental to B2WT cells but also enhances growth of B2MUT cells, yielding a positive GI score. All of the gRNAs targeting *BRCA2* in our CRISPR library cut upstream of exon 11, except for a single gRNA that does not score in this fashion. These observations support our hypothesis that the residual *BRCA2* truncation mutant in complex with free Rad51 paralogs exerts a toxic effect in the B2MUT cells.

To validate this result, we employed a multicolor competition assay (MCA): as previously described (*227*), we mixed GFP-labeled B2MUT cells with DsRed-labeled B2WT cells, infected the mixture with individual gRNAs, and monitored the relative change in percent GFP$^+$ cells by FACS (Supplemental Figure S2-2D). We individually tested the relative effect of two gRNAs to *RAD51*, *RAD51C*, *RAD51D*, and the N-terminal portion of *BRCA2* in B2MUT versus B2WT cells, utilizing our colonic isogenic cell line pair. Consistent with our screen results, all of the gRNAs tested resulted in an

enhanced ratio of B2MUT to B2WT cells, relative to negative control gRNAs that cut intergenic regions with low predicted off-target frequencies (Figure 2-2H).

To test our hypothesis that the growth enhancement seen upon depletion of *RAD51* or its paralogs is specific to *BRCA2* mutants truncated near BRC repeat 5, we performed MCA assays with our gRNAs in an isogenic cell line pair that possesses a different *BRCA2* truncation mutant: the pancreatic CAPAN-1 cell line and its *in vitro*-derived cisplatin-resistant clone C2-5 (*228*). CAPAN-1 *BRCA2* mutant cells have a single base pair deletion in one copy of *BRCA2* that produces a frameshift and prematurely truncates the protein in BRC repeat 8 (c.[6174delT*fsTer6431]), while the second allele of *BRCA2* is deleted. C2-5 cells possess a second small deletion within the mutated copy of *BRCA2* that restores its correct reading frame. In an MCA assay with this second isogenic cell line pair, the same individual gRNAs targeting *RAD51*, *RAD51C*, *RAD51D*, and the N-terminal portion of *BRCA2* no longer result in an enhancement of B2MUT cell growth relative to B2WT cells, but create the opposite effect, possibly due to suppression of residual HR in the mutant cell line (Figure 2-2H). This result suggests that the toxic effect of Rad51 depletion in cells expressing BRCA2 mutant truncation protein terminating early in BRC repeat 5 does not extend to cells that possesses a BRCA2 mutant protein with a nearly intact BRC repeat region.
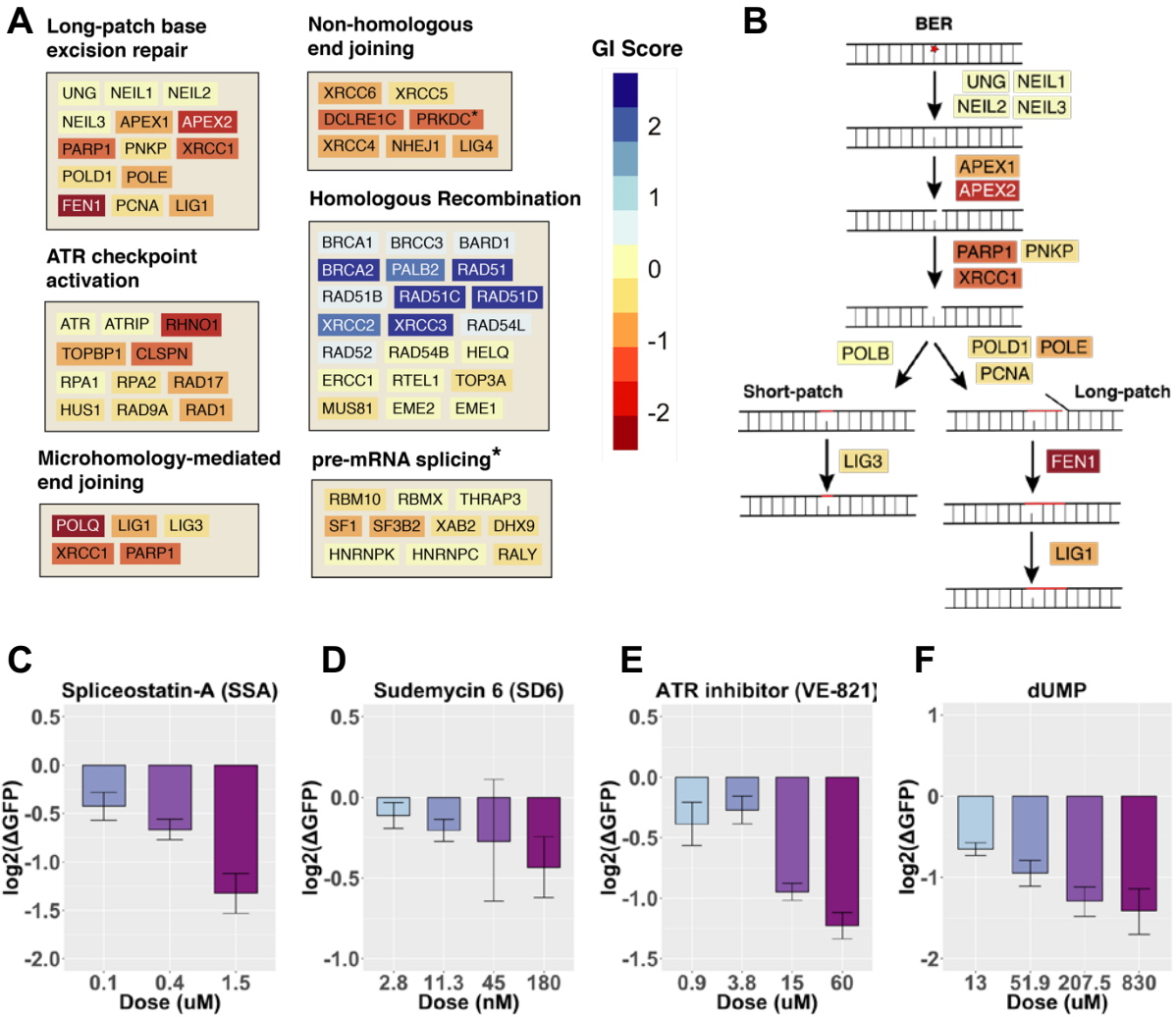
Figure 2-3. Pathway analysis of BRCA2 Synthetic Lethality. **(A)** Performance of genes in several key pathways, plotted on a color scale for genetic interaction (GI) score: the normalized average log2-fold-change across both colonic and ovarian CRISPR screens. The asterisk indicates reporting of GI score from the shRNA screen instead of combined CRISPR screens. **(B)** Schematic of the base excision repair (BER) pathway, showing the strength of the GI score for each gene in the pathway, plotted on the same color scale as in (A). **(C-F)** MCA assays in which colonic GFP-labeled BRCA2 MUT and DsRed-labeled BRCA2 WT cells were mixed and co-treated with the indicated drugs. Change in percent GFP+ cells was measured by FACS after 12 days and normalized to vehicle control.

BRCA2 mutant cells rely on ATR activation and base excision repair

From our analysis of screens in both the ovarian and colonic isogenic cell line pairs, several pathways emerged as general themes. Overall, long-patch base excision

repair (BER) scores potently as SL overall (Figure 2-3A). Indeed, almost every member of this pathway scores as SL except for the partially redundant glycosylases which initiate repair through this pathway by removing damaged sugars; their absence would not be expected to lead to nicks or DSBs that could result in an increased need for HR or AP sites that may produce replication stress (Figure 2-3B). Similarly, several members of the ATR/Chk1 signaling pathway were B2SLs, except for some of the more essential components of the ATR/ATRIP pathway such as *ATR/ATRIP* themselves and *RPA1*, which did not score using the CRISPR-based library (Figure 2-3A).

Known components of MMEJ scored more strongly as B2SL than NHEJ components, perhaps implying that the majority of detrimental unresolved DSBs addressed by these auxiliary pathways occurs during replication. The strikingly potent positive GI score we observe for *BRCA2* and Rad51 paralogs does not extend to later steps of HR, but focally affects the components of HR involved in Rad51 loading. Finally, although *SF3B2* performs as a B2SL using our shRNA library, and *SF1* does as well to a lesser extent (Figure 2-3A), these essential factors do not score with the more penetrant CRISPR-based library.

To further validate some of these pathway synthetic lethalities, we employed a multicolor competition assay (MCA). As described above, we mixed GFP-labeled B2MUT cells with DsRed-labeled B2WT cells, subjected the mixture to various drug, gRNA or control treatments, and monitored the relative change in percent GFP$^+$ cells. We tested two spliceosome inhibitors, spliceostatin-A and sudemycin D6, both of which

inhibit the U2 component SF3B1 (*229*). Both of these drugs caused a dose-dependent depletion of B2MUT GFP$^+$ cells relative to B2WT DsRed$^+$ cells, supporting the hypothesis that inhibition of the U2 spliceosome component is SL with *BRCA2* loss of function (Figures 2-3C and 2-3D).

Similarly, because several components required for ATR activation emerged as B2SL from our CRISPR screens, including *RHNO1*, *CLSPN*, and to a lesser degree the more essential *TOPBP1*, we tested the ATR inhibitor VE-821 for synthetic lethality in our MCA assay. This ATP-competitive inhibitor of ATR also exhibited dose-dependent selective inhibition of B2MUT cell growth versus B2WT cells (Figure 2-3E), confirming that B2MUT cells are more dependent on ATR activation than B2WT cells. Finally, we tested our hypothesis that B2MUT cells are more dependent on competent function of the BER pathway by increasing the load of damage that must be repaired by BER. We introduced deoxyuridine monophosphate (dUMP) directly into our cell culture medium, which has been shown to increase misincorporation of uracil into DNA (*230*). dUMP also caused a dose-dependent selective depletion of B2MUT versus B2WT cells in our MCA assay, supporting our hypothesis that B2MUT cells are more dependent on BER than B2WT cells (Figure 2-3F).

AP endonuclease _APEX2_ is synthetic lethal with _BRCA1_ and _BRCA2_ loss-of-function

In our CRISPR-based screens, we identified two novel, potent, and generalizable B2SL hits: _APEX2_ and _FEN1_. _APEX2_ encodes Ape2, an AP endonuclease responsible for the second step of BER: after base removal by a lesion-appropriate glycosylase, Ape2 hydrolyzes the phosphodiester backbone immediately 5' to the AP site to create a single-strand break (SSB). In addition to its AP endonuclease activity, Ape2 also possesses 3' phosphodiesterase activity and 3'-5' exonuclease activity, executed from the same ExoIII-like nuclease domain.

Human cells rely on two ExoIII-family type II endonucleases to generate a nick upstream of AP sites during BER, encoded by _APEX1_ and _APEX2_. We were curious as to why _APEX2_ scored strongly as a B2SL gene, while _APEX1_ scored weakly, given that Ape1 possesses stronger _in vitro_ AP endonuclease activity than Ape2 (_231_). Ape1 and Ape2 both contain an ExoIII-like endonuclease domain with 29% sequence identity and high sequence similarity (_232_), but Ape1 also contains a separate, distinct N-terminal redox domain. A critical cysteine in this redox domain reduces the heterodimeric transcription factor AP-1 (c-Jun/c-Fos), enhancing its DNA-binding activity (_233_).

To validate _APEX2_ as a B2SL hit, we tested three individual _APEX2_ gRNAs in an MCA assay (Figure 2-4A) compared to non-cutting negative control gRNAs and negative control gRNAs that cut intergenic regions with low predicted off-target cutting

Figure 2-4. APEX2 is synthetic lethal with BRCA2. (A-B) MCA assays in which ovarian GFP-labeled BRCA2 MUT cells and DsRed-labeled BRCA2 WT cells were mixed and co-infected with 3 individual lentiviruses expressing gRNAs to *APEX2* or *APEX1.* Change in percent GFP$^+$ cells was measured by FACS after 7 days and normalized to gRNA-expressing cells. **(C-D)** MCA assays in which GFP-labeled BRCA1 MUT cells and DsRed-labeled BRCA1 WT cells were mixed and co-infected with 3 individual lentiviruses expressing gRNAs to *APEX2* or *APEX1.* Change in percent GFP cells was measured by FACS after 12 days and normalized to gRNA-expressing cells. **(E)** List of ORFs tested for complementation in cells expressing Cas9 and a gRNA to *APEX2.* **(F)** Examination of the ability of the ORFs from (F) to rescue the growth defect caused by expression of Cas9 and an *APEX2* gRNA in ovarian BRCA2 MUT cells. Cells were co-infected with lentivirus expressing an *APEX2* gRNA and the indicated gRNA-resistant APEX2 ORF or negative control peptide. After selection and growth for 8 days, survival was quantified with a FACS-based cell counting method. **(G)** MCA assay (*as described above*) in which colonic BRCA2 MUT and BRCA2 WT cells were mixed and treated for 12 days with the Ape nuclease inhibitor APEIII. **(H)** MCA assay (*as described above*) in which GFP-labeled colonic BRCA2 MUT and DsRed-labeled colonic BRCA2 WT cells were mixed and treated for 12 days with the Ape1 redox inhibitor E3330, which inhibits the redox functionality of Ape1 but not its nuclease domain.

#	ORF	Mutation	Effect
1	APEX2 WT	na	Wild-type APEX2
2	APEX2 ENDO	APEX2 D277A	Ablates APEX2 endonuclease activity
3	APEX1 WT	na	Wild-type APEX1
4	APEX1 ENDO	APEX1 Y171F	Ablates APEX1 endonuclease activity
5	NC	na	Small negative control peptide
6	DsRed-E2C	na	Red fluorescent protein E2-Crimson
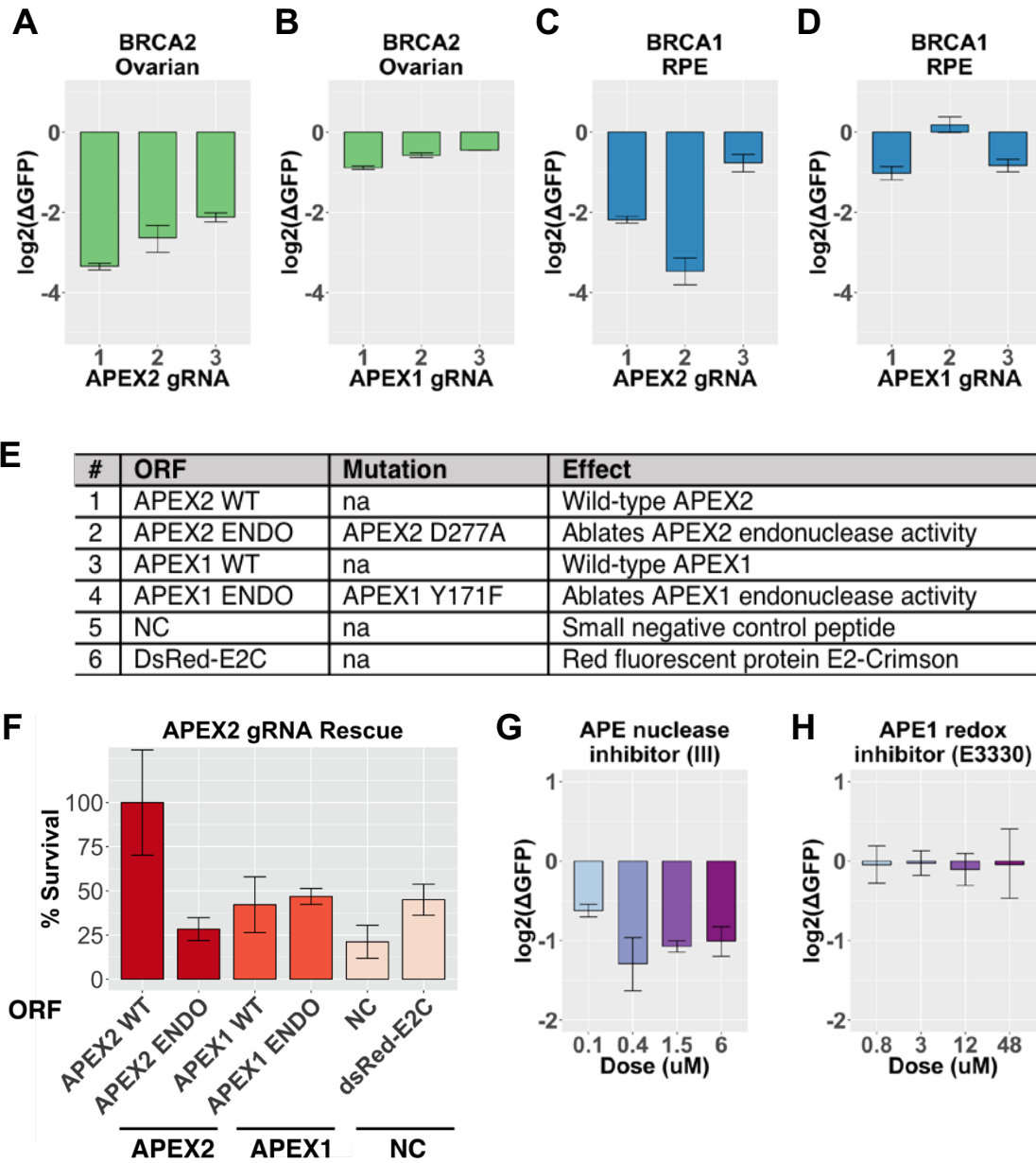
Figure 2-4 (Continued). APEX2 is synthetic lethal with BRCA2.

(Supplemental Figure S2-3A). Consistent with our screen results, three individual gRNAs targeting *APEX2* validated strongly as B2SL in our ovarian isogenic cell line pair, while three individual gRNAs targeting *APEX1* validated weakly as B2SL (Figures 2-4A and 2-4B). We found the same result to be true in our colonic isogenic cell line pair (Supplemental Figures S2-3H and S2-3I). In addition, we asked whether inactivation of *APEX1* or *APEX2* exhibits synthetic lethality with *BRCA1* by testing these same gRNAs in an MCA assay utilizing a *BRCA1* isogenic cell line pair in the RPE background. In a similar fashion to deficiency of *BRCA2*, knockout of *APEX2* is strongly SL with *BRCA1* deficiency, while knockout of *APEX1* exhibits synthetic lethality to a weaker degree (Figures 2-4C, 2-4D).

We further assessed the synthetic lethality of *APEX1* by testing two chemical inhibitors of Ape1: APE Inhibitor III and E3330. APE1 Inhibitor III enzymatically inhibits Ape1's nuclease domain (*234*), though its potential to exert off-target effects on the structurally similar nuclease domain of Ape2 has not yet been characterized. This small molecule caused selective growth inhibition of B2MUT cells relative to B2WT cells, through enzymatic inhibition of the nuclease domain of Ape1 and possibly also Ape2 (Figure 2-4G). In contrast, the small molecule inhibitor E3330 blocks the redox activity of Ape1 but does not impede its DNA repair function (*235*). This inhibitor did not exhibit the same SL effect, suggesting that the nuclease domain, rather than the redox domain, of Ape1 contributes to its B2SL phenotype (Figure 2-4H).

51

To verify that the B2SL phenotype exhibited by *APEX2* gRNAs is due to cutting of *APEX2* by Cas9, rather than off-target effects, we performed a rescue experiment (Figures 2-4E and 2-4F). We demonstrated that expression of gRNA-resistant wild-type Ape2 was able to rescue the growth defect in B2MUT cells caused by *APEX2* gRNA cutting, relative to two negative control peptides (Figure 2-4F). When we introduced a point mutation in the gRNA-resistant Ape2 ORF that inactivated its nuclease domain (D277A), we found that the nuclease-deficient Ape2 mutant was no longer capable of rescuing the growth deficit caused by the *APEX2* gRNA (Figure 2-4F). This observation suggests that the nuclease domain of Ape2 is required for the B2SL phenotype. Similarly, expression of the wild-type Ape1 ORF (which is not cut by our *APEX2* gRNA), or its nuclease-deficient point mutant (Y171F) was not able to rescue B2MUT cell growth. The fact that Ape1 cannot complement loss of Ape2 further suggests that Ape2 possesses a distinct feature responsible for its stronger SL phenotype.

Figure 2-5. FEN1 is a novel B2SL gene that promotes MMEJ. (A) MCA assay in which ovarian GFP-labeled BRCA2 MUT cells and DsRed-labeled BRCA2 WT cells were mixed and co-infected with 3 individual gRNAs to *FEN1*. Change in percent GFP$^+$ cells was measured by FACS after 7 days and normalized to negative control gRNA-expressing cells. **(B)** MCA assay in which colonic GFP-labeled BRCA2 MUT cells and DsRed-labeled BRCA2 WT cells were mixed and co-infected with 3 individual gRNAs to *FEN1*. Change in percent GFP$^+$ cells was measured by FACS after 7 days and normalized to negative control gRNA-expressing cells. **(C)** MCA assay in which colonic GFP-labeled BRCA2 MUT cells and DsRed-labeled BRCA2 WT cells were mixed and co-treated with the indicated doses of a FEN1 inhibitor. The percent of GFP$^+$ cells was quantified by FACS after 7 days and normalized to DMSO. **(D)** Dependency of *BRCA1/2* MUT or *BRCA1/2* WT cell lines on *FEN1*, determined by the CERES computational method from genome-scale CRISPR-Cas9 essentiality screens across 324 cancer cell lines from the Cancer Cell Line Encyclopedia (CCLE) (*236*). **(E)** Extracts from BRCA2 MUT ovarian cells expressing Cas9 and the indicated gRNAs were immunoblotted with the indicated antibodies; both panels were run on the same gel. **(F)** Examination of the ability of the ORFs listed in (G) to rescue the growth defect caused by expression of Cas9 and a *FEN1* gRNA. BRCA2 MUT ovarian cells were co-infected with lentivirus expressing a *FEN1* gRNA and the indicated gRNA-resistant FEN1 ORF or negative control peptide. After selection and growth for 8 days, survival was quantified with a FACS-based cell counting method. **(G)** List of gRNA-resistant ORFs tested for complementation in BRCA2 MUT cells expressing Cas9 and a FEN1 gRNA. **(H)** Extracts from *BRCA2* WT cells expressing Cas9, a validated FEN1 gRNA, and the indicated gRNA-resistant ORF were immunoblotted with the indicated antibodies. **(I)** Cells transfected with negative control (NC) or FEN1 siRNAs were stained with DAPI and an antibody to γH2AX to examine foci formation by immunofluorescence. **(J)** Schematic of the Dual EJ reporter. **(K)** Effect of siRNA knockdown of *PARP1*, *FEN1*, and *POLQ* on MMEJ and NHEJ repair as measured by the Dual EJ Reporter in U2OS cells. The presence of an asterisk indicates statistical significance (p<0.05 by t-test) for the corresponding siRNA versus negative control siRNA.

A BRCA2 Ovarian

B BRCA2 Colonic

C FEN1 inhibitor

D CCLE

E

F FEN1 gRNA Rescue

G

#	ORF	Mutation	Effect
1	WT	na	Wild-type
2	EXO	R47A	Ablates exonuclease activity
3	EXO-ENDO	R47A+R70A	Ablates exonuclease and reduces endonuclease activity
5	LOC1	S187A	Increases nucleolar subcellular localization
6	ENDO	D181A	Ablates endonuclease activity
7	NC	na	Small negative control peptide
8	DsRed-E2C	na	Red fluorescent protein (DsRed derivative E2-Crimson)

H

I

J Dual EJ Reporter Schematic

K Dual EJ Reporter

Figure 2-5 (Continued). FEN1 is a novel B2SL gene that promotes MMEJ.

54

Flap endonuclease FEN1 is a novel B2SL that plays a role in MMEJ

Our top novel B2SL hit, *FEN1*, performed strongly in both the ovarian and colonic isogenic *BRCA2* mutant cell line pairs. FEN1 is a structure-specific endonuclease that recognizes and cleaves 5' ended single-stranded flaps (*235*). Several essential processes in DNA repair and metabolism require this enzymatic function, such as the penultimate step of BER: FEN1 processes the 5' flap created by Polβ-mediated strand-displacement synthesis. Similarly, FEN1 resolves 5' flaps created in lagging strand synthesis during replication; while extending an Okazaki fragment, Polδ encounters the subsequent 3' fragment and proceeds by strand-displacement synthesis to displace the downstream RNA-DNA primer, creating a 5' flap structure. In addition to this important flap endonuclease activity, FEN1's sole active site also possesses gap endonuclease activity and 5'-3' exonuclease activity, both of which are weaker than its flap endonuclease activity *in vitro* (*237*).

We first validated that *FEN1* is a B2SL gene using the MCA assay described above. Three individual gRNAs targeting *FEN1* resulted in selective growth inhibition of B2MUT versus B2WT cells, in both the ovarian and colonic backgrounds (Figures 2-5A and 2-5B). To ask if this drug would induce synthetic lethality in *BRCA1*-mutant as well as *BRCA2*-mutant tumor lines, we tested three individual gRNAs to *FEN1* in an MCA assay utilizing a *BRCA1* isogenic cell line pair in the RPE background. Each of the gRNAs selectively inhibit growth of *BRCA1* MUT cells relative to *BRCA1* WT cells, indicating that *FEN1* is a target applicable to patients with either a germline *BRCA1* or

BRCA2 mutation (Supplemental Figure S2-3N). We verified that these gRNAs deplete the protein level of FEN1 by Western blotting (Figure 2-5E), and we further saw an increase in γH2AX phosphorylation and Chk1 phosphorylation at S317 in the presence of FEN1 gRNAs, possibly reflecting an increase in DSB load (Figure 2-5E). From this result, we cannot distinguish whether the increase in γH2AX and Chk1 phosphorylation is a result of gRNA cutting or of FEN1 depletion. Thus, we utilized a pooled siRNA mixture to deplete FEN1, and indeed we observed an increase in γH2AX foci by immunofluorescence upon FEN1 knockdown (Figure 2-5I).

To assess whether FEN1 may represent a novel drug target in *BRCA*-deficient tumors, we tested a small molecule inhibitor of FEN1. This inhibitor sterically occupies the active site of FEN1, and thus impedes both its exonuclease and endonuclease functionalities (*238, 239*). In an MCA assay, we found that this FEN1 inhibitor selectively impairs growth of B2MUT cells relative to B2WT cells, confirming that FEN1 is a viable novel B2SL drug target (Figure 2-5C). We also asked if we could validate *FEN1* as broadly *BRCA* SL by utilizing the publicly available Avana CRISPR-Cas9 genome-scale dataset, analyzed by the CERES algorithm (*236*). We separated 391 cell lines based on *BRCA*ness (by mutation status of *BRCA1* and *BRCA2*) and saw that the *BRCA*-mutant lines are more dependent on *FEN1* than the *BRCA*-wildtype lines (Figure 2-5D). This analysis provides independent confirmation of the strength of *FEN1* as a B2SL hit.

Next, we asked whether the B2SL phenotype induced by our *FEN1* gRNAs is directly due to on-target cutting of *FEN1* by CRISPR/Cas9. We rescued one of our validated *FEN1* gRNAs with either a gRNA-resistant wild-type FEN1 ORF or negative control peptides; we found that full-length gRNA-resistant wild-type FEN1 rescues the growth defect imparted by the *FEN1* gRNA, confirming an on-target effect of our gRNA (Figures 2-5F and 2-5G). To investigate which enzymatic function of FEN1 is important for its B2SL phenotype, we made a series of point mutants, including one mutant that inactivates only the exonuclease activity of FEN1, and two different point mutants that inactivate both its endonuclease and exonuclease activity. There is no single point mutation that inactivates the endonuclease activity of FEN1 while leaving its exonuclease activity intact. We found that the exonuclease-deficient FEN1 mutant was able to rescue the effect of our *FEN1* gRNA, while both endonuclease mutants were unable to rescue this effect, suggesting that *BRCA2*-deficient cells are selectively dependent upon the endonuclease but not the exonuclease activity of FEN1 (Figures 2-5F and 2-5G). We verified that all of these FEN1 wild-type and mutant ORFs were expressed in the presence of the chosen FEN1 gRNA by Western blotting (Figure 2-5H).

Finally, we considered why the flap endonuclease activity of *FEN1* is particularly important in *BRCA2*-deficient cells. We hypothesized that FEN1 may be involved in processing the 5' flaps created by *POLQ* during MMEJ repair (*240*). To test whether FEN1 plays a role in MMEJ flap resolution in mammalian cells, we designed a novel cell-based reporter construct: the Dual EJ Reporter (Figure 2-5J). This reporter is a

modified fusion protein between a DsRed derivative (E2-Crimson) and GFP that yields

information about both MMEJ and NHEJ-mediated repair.

We inserted an I-SceI cut site followed by a transcriptional stop sequence into a

structurally and functionally critical region of DsRed. We flanked both sides of this I-

SceI cut site with the same 8bp region of microhomology, so that repair by MMEJ

perfectly restores the original sequence of DsRed. To this modified DsRed protein, we

fused GFP with an intervening glycine and serine-rich flexible linker. Thus, MMEJ

repair of an I-SceI-induced DSB yields expression of both correctly-repaired DsRed and

GFP. In contrast, NHEJ repair of the I-SceI cut site does not restore the sequence of

this functionally important part of DsRed, but one-third of the time resolves into in-

frame expression of GFP. Thus, single positive GFP$^+$ cells reflect NHEJ repair, while

double-positive DsRed-GFP$^+$ cells reflect MMEJ repair.

We introduced our Dual EJ Reporter at low MOI into U2OS cells and tested the

effect of pooled siRNA-mediated knockdown of *PARP1*, *FEN1*, and *POLQ* versus

negative control siRNA on the reporter. Both *POLQ* and *PARP1* are known to have a

role in MMEJ (*241*); as expected, knockdown of both genes resulted in decreased

expression of the double-positive DsRed-GFP$^+$ fusion protein (Figure 2-5K).

Knockdown of *FEN1* had a similar effect, confirming our hypothesis that *FEN1*

participates in MMEJ in mammalian cells (Figures 2-5J and 2-5K). In contrast, none of

these genes are expected to play a role in classical NHEJ, and our reporter confirmed

that knockdown of these three genes did not affect classical NHEJ repair. Thus,

screening for B2SL genes led us to identify a *FEN1* as a component of the MMEJ

pathway, highlighting how dependent B2MUT cells are upon MMEJ in the absence of

functional HR.

III.   Discussion

It has become clear that PARP inhibitor effectiveness in *BRCA*-deficient cells is

partially due to the phenomenon of PARP trapping, in addition to exploitation of an

underlying synthetic lethal genetic interaction between *PARP* and *BRCA1/2*

deficiencies. However, a systematic analysis of *BRCA2*-mutant interactions with

deficiencies of other DNA repair functions has not been reported. We reasoned that

unexplored SL relationships with *BRCA*-deficiency may exist and these might represent

valuable drug targets for metastatic or refractory *BRCA*-deficient breast tumors,

recurrent HGSOC, and PARP inhibitor-resistant *BRCA*-deficient tumors. Thus, we

designed a series of genetic screens to identify novel B2SL targets. We show here that

there are several B2SL genes and pathways whose inactivation exhibits stronger

synthetic lethality with *BRCA2* inactivation than *PARP1* loss. Both *APEX2* and *FEN1*

are previously unreported, potent, and generalizable B2SL hits that represent novel drug

targets.

Inactivation of *APEX2* hampers the BER pathway, which overall is more

important to survival in the context of *BRCA2* inactivation than in counterpart *BRCA2*

wild-type cells. It has previously been assumed that PARP involvement in BER is SL

59

with *BRCA* inactivation because unresolved SSBs create an increased DSB burden during replication. However, inactivation of *APEX2* would not necessarily lead to an increase in SSBs, since Ape1 and Ape2 endonucleases are responsible for the nick-generating step of BER. Rather, in the absence of Ape2 endonuclease function, residual AP sites would persist during S phase, resulting in replication fork stalling. We hypothesize that this increase in replication stress is more toxic in the absence of functional *BRCA2*, since BRCA2 has been shown to safeguard forks from resection (*242*). This hypothesis is consistent with our observation that B2MUT cells are more dependent on ATR activation, given the master regulatory role of ATR activation in replication fork protection (*243*).

Despite the fact that Ape1 is commonly believed to be the dominant BER AP endonuclease, we found that *APEX2* is a stronger *BRCA1* and *BRCA2* SL hit than *APEX1*. The fact that we cannot complement *APEX2* loss with overexpression of the wild-type APEX1 ORF suggests that there is a distinct feature of APEX2 that is responsible for its stronger SL phenotype. In contrast to Ape1, Ape2 binds PCNA (*244, 245*), which suggests that Ape2 may participate in BER during replication, such as when a replication fork encounters an abasic lesion. We hypothesize that the co-localization of Ape2 with PCNA during replication is responsible for its stronger *BRCA1* and *BRCA2* SL effect.

Surprisingly, we found that depletion of every human Rad51 paralog, as well as *PALB2* and *BRCA2* itself, enhances growth of our B2MUT cell lines. We note that

both the ovarian and colonic B2MUT lines possess similar BRCA2 truncation mutants, present in about ~16% of breast tumors. This truncation mutant is capable of binding Rad51; thus, we propose that exon 11 BRCA2 truncation mutants may exert a toxic, neomorphic function involving Rad51. We did not see this same effect when we tested B2MUT cells possessing a longer truncation mutant, terminating at BRC repeat 8, suggesting that this toxic function involving Rad51 may be unique to BRCA2 truncation mutants that terminate near BRC repeat 5. This result suggests that inhibiting HR in the context of a BRCA2 exon 11 truncation mutant may actually accelerate tumor growth, implying that certain classes of HR inhibitors might exacerbate cancers if given to patients with exon 11 BRCA2 truncation mutants.

A few essential genes scored as SL in our shRNA screen, but not CRISPR-based screens, including the splicing factors *SF3B2* and *SF1*. We validated that two inhibitors of the SF3B1 component of the U2 snRNP selectively impair growth of B2MUT cells. Since inactivation of *BRCA2* leads to increased R-loop formation (*213*), dual inactivation of *BRCA2* and *SF3B1* may compound R-loop induced DNA damage to a toxic level. Alternatively, this effect may be explained by previously reported results that the U2 snRNP maintains protein levels of essential DDR components and prevents R-loop induced DNA damage (*246*). These two explanations for the effect of SF3B1 inhibition on *BRCA2* mutant cells are not mutually exclusive.

Our strongest novel B2SL candidate is the flap endonuclease *FEN1*. We validated *FEN1* as a *BRCA1* and *BRCA2* SL target and showed that its endonuclease

but not its exonuclease enzymatic activity is required to rescue the effect of *FEN1* gRNA knockout in *BRCA2* MUT cells. We showed using a novel cell-based reporter that *FEN1* is an essential component of the MMEJ pathway. The fact that we found this phenotype through a B2SL screen underscores how important the MMEJ pathway becomes in the absence of functional HR. However, the SL phenotype of *FEN1* knockout in B2MUT cells is likely multifactorial: its loss impairs BER, creates replication stress by hindering processing of Okazaki fragments, and also impairs MMEJ-mediated DSB repair. Overall, MMEJ appears to be more important in complementing loss of HR than classical NHEJ, perhaps implying that the predominant load of DSBs left unresolved by HR deficiency occurs during replication.

Finally, we showed that an existing inhibitor to FEN1 exhibits synthetic lethality *in vitro*, suggesting it could possibly be utilized in the treatment of *BRCA*-deficient tumors, subsequently or in combination with PARP inhibitors. Importantly, the toxic PARP-SSB intermediates that occur upon PARP trapping are thought to be resolved largely by BER; thus, because FEN1 executes the penultimate step of BER, it may exhibit a synergistic effect with PARP inhibition. Not only would its inhibition be SL in *BRCA*-deficient tumors due to the role of FEN1 in MMEJ, replication, and BER, but its inhibition should also leave toxic PARP-SSB complexes unresolved, potentially augmenting a synergistic lethality.

We are currently optimizing a drug synergy study between FEN1 and PARP inhibitors. In addition, we have generated a gRNA-resistant ORF encoding a PIP-box

mutant of Ape2 that should fail to bind PCNA; we are testing our hypothesis that Ape2 synthetic lethality is stronger than Ape1 synthetic lethality due to its ability to bind PCNA. Importantly, we are characterizing our Dual EJ reporter in depth: we FACS-sorted cells that repaired the reporter to express both DsRed-GFP or GFP alone, and deep sequenced the sorted populations to confirm their repair by the expected pathway. We are also repeating this assay with siRNA to known components of the NHEJ pathway including *XRCC5/6* as additional controls. Ideally, we feel that re-cloning the reporter to include a P2A-based DsRed-GFP construct instead of a fusion protein would yield better stability and likely increased signal.

IV.   Methods

Cell culture

Human DLD-1 cells and all clones derived from this cell line were maintained in RPMI-1640 Medium (ATCC modification) supplemented with 10% (v/v) fetal bovine serum (FBS), 100 units/mL penicillin, and 100 µg/mL streptomycin. Human PEO1 cells and their *BRCA2* revertant clone C4-2 were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS, 1% Pen-Strep, and 2mM glutamine. A pair of *BRCA1* isogenic RPE1 cell lines was generously shared by Connor Clairmont and Alan D'Andrea; this isogenic cell line pair was created by CRISPR-based knockout of *TP53* with or without CRISPR-based knockout of *BRCA1* in the RPE1 background.

These cells were grown in DMEM/F12 medium supplemented with 10% FBS, 100 units/mL penicillin, and 100 μg/mL streptomycin. HEK293T cells were maintained in DMEM supplemented with 10% FBS, 100 units/mL penicillin, and 100 μg/mL streptomycin.

Lentivirus and retrovirus production and titering

To produce lentivirus, HEK293T cells were seeded in tissue culture dishes at a density equivalent to $6x10^5$ cells per 0.9 cm^2 surface area. Plasmid DNA was diluted into serum-free medium with a lentiviral packaging plasmid mixture of SV40 VSVg, Gag/Pol, Tat, and Rev, and transfected with PolyJet or MIRUS Trans-IT. After 48 h, the supernatant was harvested, filtered through a low-protein binding HT Tuffryn® membrane with 0.45 μm pores (Pall cat. #4184), and stored at -80°C. Lentiviral titer was determined by transducing the cell line of interest plated at clonogenic density with serial dilutions of virus in the presence of 4-8 μg/mL polybrene. After selecting with puromycin or NAT, colonies were counted to determine viral titer. The same process was utilized to package retrovirus when applicable, except that a retroviral packaging plasmid mixture of 1:1 Gag/Pol and VSVG was co-transfected with retroviral constructs. All transductions were performed in the presence of 4-8 μg/mL polybrene.

Generation of isogenic cell lines

We obtained a *BRCA2* mutant cell line (*BRCA2$^{-/-}$* DLD-1) from Horizon Pharma. This cell line was originally generated in the colonic pseudodiploid DLD-1 background through two successive rounds of homologous recombination introduced by adeno-associated virus, each of which created a small deletion in exon 11, leaving a residual loxP site and premature stop codon (*218*). We received a pcDNA3.1-neo construct for full-length BRCA2 expression as a generous gift from Ralph Scully. We transformed this plasmid into DH5α *E. coli* and maintained growth of bacteria at 30°C in the presence of Ampicillin. The amplified plasmid was linearized, treated with calf alkaline phosphatase, and transfected into mutant *BRCA2$^{-/-}$* DLD-1cell line with MIRUS. Neomycin selection was maintained for several weeks to identify stable integrants, and individual clones were isolated. Restoration of full-length BRCA2 expression was confirmed by Western blotting, and selected clones were shown to exhibit olaparib resistance and Rad51 foci formation after 10Gy IR.

shRNA screens

We designed a targeted sublibrary of 380 genes with either a known or suspected role in the DNA damage response. Inclusion of genes in this DNA damage sublibrary was based either on a known role in DNA repair described in the literature, or performance of the gene in prior screens for sensitivity to DNA damaging agents (unpublished data). We began by performing a primary shRNA screen to these 380

genes, with an shRNA library targeting each gene with 50 different shRNAs (~19K total). Pooled oligonucleotides encoding these shRNAs were cloned into the mir30 backbone context in the retroviral hairpin-expressing vector MSCV (*247*). The library was packaged into retrovirus along with 10 negative control shRNAs in a pooled format, and titer was measured by a colony formation assay.

A pooled shRNA screen was then performed by transducing this retroviral library into both colonic DLD-1 B2MUT and B2WT cell lines at a low multiplicity of infection (0.5) in triplicate at a representation of 1000 cellular integrations per shRNA. After selection in 4 µg/mL puromycin for 3 days, cells were passaged for 12 population doubling (PDs), and cell pellets were collected both after puromycin selection (PD0) and after passaging cells for 12 population doublings (PD12). Genomic DNA was isolated from these cell pellets by phenol/chloroform extraction, and the relative representation of library reagents in each sample was determined by Illumina sequencing of PCR-amplified half-hairpins (*219*).

Fifty genes that performed well in the primary shRNA screen were selected to comprise a secondary screen sublibrary. Ten hairpin sequences were designed to each of these fifty genes (~500 total) and synthesized along with 772 negative control shRNAs. Pooled oligonucleotides encoding these hairpins were cloned into the mir-E shRNA context in the lentiviral vector pHAGE pInducer10 (*248*). A pooled shRNA screen was performed similarly to the primary shRNA screen in both colonic DLD-1 B2MUT and B2WT cell lines, with the lentiviral library pool being transduced at low multiplicity of

infection (0.2) in triplicate at a representation of 1000. After selection with 4 μg/mL puromycin for 3 days, cells were passaged for 12 PDs, and cell pellets were collected from both starting PD0 and final PD12 cell populations. Again, genomic DNA was isolated from these cell pellets by phenol-chloroform extraction, and the relative representation of library reagents was determined by Illumina sequencing of PCR-amplified half-hairpins (*219*).

For both the primary and shRNA screens, NGS sequencing reads were aligned to the library using Bowtie (*249*) and counts were obtained for each shRNA. MAGeCK (*223*), MAGeCK-VISPR (*250*) and edgeR (*224*) were used to calculate gene rank lists, false discovery rates (FDRs), and log2-fold-changes for each gene

CRISPR screens

gRNAs were designed to 357 of the 380 DNA damage sublibrary genes targeted by the primary shRNA library. Ten gRNAs were designed to each gene, yielding ~3500 gRNAs total. A negative control library targeted to the *E. coli* genome with minimal predicted off-target cutting sites in the human genome was designed in parallel. These oligonucleotides were cloned in a pooled format into the LentiCRISPRv2 plasmid, which contains EFS-driven Cas9 (*251*). The DNA damage sublibrary and *E. coli*-targeted negative control library were packaged into lentivirus and titered separately but pooled before transduction into colonic or ovarian B2MUT or B2WT cells in triplicate at a

MOI of 0.2. After puromycin selection (4 μg/mL for colonic B2MUT and B2WT lines and 1.5μg/mL for ovarian B2MUT and B2WT lines), cells were passaged for 12 PDs, and cell pellets were collected at both PD0 and PD12. Genomic DNA was isolated from cell pellets using phenol-chloroform extraction, and gRNAs were PCR-amplified from genomic DNA and adapted for Illumina sequencing. Sequencing reads were aligned to the starting libraries, and read counts were analyzed by MAGeCK (*223*), MAGeCK-VISPR (*250*) and edgeR (*224*) to calculate gene rank lists, FDRs, and log2-fold-changes for each gene.

Multicolor Competition Assay (MCA)

Individual gRNAs or small molecules were tested for synthetic lethality using a multi-color competition assay (MCA) (*132*). In this assay, GFP-labeled B2MUT cells are mixed with B2WT cells labeled with the DsRed-derivative red fluorescent protein E2-Crimson (E2C), at a 2:1 ratio. The percent of GFP$^+$ cells was monitored over time using fluorescence-activated cell sorting (FACS), and unmixed GFP-labeled B2MUT and E2C-labeled B2WT cells were maintained in parallel to assure the purity of each population.

For MCA assays evaluating the effect of gRNAs, individual gRNAs were cloned into the LentiCRISPRv2 plasmid containing the previously described F+E modifications to the tracrRNA sequence (*252*). After sequence verification of the cloned gRNA, each

gRNA to be tested was packaged separately into lentivirus (*251, 253*). Negative control

gRNAs included three sequences targeted to the *E. coli* genome that are predicted to

have no off-target cutting sites in the human genome, and several intergenic gRNA

sequences predicted to have a single cutting site in the human genome, in a DNAse

hypersensitive region, on chromosome 2, 15, or 16. The 2:1 mixture of pre-labeled

B2MUT and B2WT cells was infected with each gRNA to be tested, selected with

puromycin, and the percent GFP$^+$ cells after selection was measured as the starting

percent GFP. After passaging in culture for 1 week, the percent of GFP$^+$ cells from

each sample was measured again. The change in percent GFP after passaging cells was

normalized to the negative control gRNAs and adjusted for differing relative growth

rates as previously described (*254*), according to the following formula:

$$\text{Log2}(\Delta\text{GFP}) = \log_2 \frac{(\text{GFP}_{\text{treated}} - \text{GFP}_{\text{control}} * \text{GFP}_{\text{treated}})}{(\text{GFP}_{\text{control}} - \text{GFP}_{\text{control}} * \text{GFP}_{\text{treated}})}$$

For MCA assays evaluating the effect of small molecules, serial dilutions of each

drug were prepared in DMSO, with the final concentration of DMSO in culture never

exceeding 0.45%. We verified that the addition of 0.45% DMSO had no effect on the

change in percent GFP$^+$ cells over time. We mixed GFP-labeled B2MUT with E2C-

labeled B2WT cells as described above, and then added drug to the mixture every 3-4

days for 1 week of passaging in culture. We measured the percent of GFP$^+$ positive

cells by FACS and normalized to DMSO-treated cells using the formula above. The only exception to this procedure was for 2′-deoxyuridine 5′-monophosphate disodium salt (dUMP), which was solubilized in H_2O. For this compound, change in percent GFP$^+$ cells was normalized to the change in percent GFP$^+$ cells with the addition of the same volume of H_2O.

The drugs test in this study include: APE Inhibitor III (Calbiochem), ATR inhibitor VE-821 (Axon 1893), dUMP (Sigma D3876), E3330 (Sigma E8534), Olaparib (Selleck Chemicals S1060), and SSA (Adooq Bioscience A12700). Sudemycin 6 was a generous gift from Dr. Thomas Westbrook and FEN1 inhibitor was generously provided by Dr. Stephen Durant from AstraZeneca (*238*).

Western blotting

Cells were trypsinized, washed in PBS, counted in order to standardize gel loading, and lysed by resuspension in 2X NuPAGE LDS Sample Buffer (Thermo) containing Halt protease and phosphatase inhibitors (Thermo). Each sample was sonicated 4 times for 15 second intervals, with at least 15 seconds rest on ice in between successive sonication periods, before being boiled for 5 minutes at 95°C. Protein lysates were separated on 4-12% Bis-Tris gels (Thermo) and transferred to nitrocellulose membranes (Bio-Rad), which were blocked with 0.45μm-filtered 5% BSA in TBST (Cell Signaling Technology). The following primary antibodies and dilutions were used for immunoblotting: vinculin (1:1000, Sigma V9131), GAPDH (1:1000, Santa Cruz, sc-

70

25778), BRCA2 (1:1000, Millipore OP95), phospho-Chk1 S317 (1:500, Cell Signaling 2344), Chk1 (1:1000, Cell Signaling 2360), phospho-H2AX S139 (1:1000, Millipore 05-636), and FEN1 (1:500, Thermo 4E7). Goat anti-mouse (1:5000, Jackson 115-035-003) or goat anti-rabbit (1:5000, Jackson 111-035-003) HRP-conjugated secondary antibodies were used for detection with enhanced chemiluminescence (ECL) (Perkin-Elmer NEL104001EA).

Immunofluorescence

Cells grown on coverslips were fixed for 10 minutes at room temperature in 3.7% formaldehyde, freshly diluted in PBS. Cells were then permeabilized for 10 minutes at room temperature in 0.5% Triton-X in PBS and blocked for 20 minutes in 0.5% BSA and 0.2% gelatin from cold water fish skin (Sigma) in PBS. Three PBS washes occurred before each of these steps. Primary and secondary antibodies were diluted in PBS with 0.5% BSA and 0.2% gelatin from cold water fish skin (Sigma). Unless otherwise stated, primary antibody incubations were performed for 2 hours at room temperature in a dark, humidified chamber followed by a 1 h secondary antibody incubation in a dark, humidified chamber. The following antibodies and dilutions were used in this study: phospho-H2AX S139 (1:500, Millipore 05-636), Rad51 (1:200, 45 min, 37°C, Santa Cruz sc8349). Secondary antibodies included Alexa Fluor® 488 goat anti-mouse IgG (1:1000, Invitrogen A-11001), Alexa Fluor® 594 goat anti-mouse IgG (1:1000, Invitrogen A-11005), Alexa Fluor® 647 goat anti-mouse IgG (1:1000, Invitrogen A-21236), and Alexa

Fluor® 488 goat anti-rabbit IgG (1:1000, Invitrogen A-11008). Coverslips were

mounted on slides using Vectashield mounting medium with DAPI (Vector

Laboratories), and imaging was performed on an Olympus Fluoview FV1000 Confocal

Microscope.

Rescue Experiment

The ORF encoding full-length APEX2 was obtained from the human ORFeome

8.1 collection, and a stop codon was introduced after the last residue of the *APEX2*

coding sequence. A gRNA-resistant version of this construct was constructed by site-

directed mutagenesis (Agilent) of the PAM and several bases in the seed sequence of the

corresponding to a pre-validated *APEX2* gRNA (*APEX2* gRNA #3: 5'-

GGTAGCATTGTCCTTACAGA-3'). This gRNA-resistant construct was used as the

substrate to make several *APEX2* point mutants.

The full-length ORF encoding FEN1 terminated by its natural stop codon was

obtained from the Ultimate ORF collection (Thermo Fisher). A gRNA-resistant version

of this construct was cloned by site-directed mutagenesis (Agilent) of the PAM and

several bases in the seed sequence of a pre-validated gRNA to *FEN1* (*FEN1* gRNA #3:

5'- GGCTGGCAAAGTCTATGCTG-3'). This gRNA-resistant construct was used as

the substrate to make several *FEN1* point mutants.

All resulting constructs were cloned into a lentiviral destination vector by a

Gateway LR reaction. A small, 20 amino-acid negative control peptide and the red

fluorescent protein E2C were cloned in parallel to serve as negative controls. Pre-validated gRNAs to *APEX2* or *FEN1* were transduced into ovarian B2MUT cells or ovarian B2WT cells and selected with 1.5 µg/mL puromycin. Subsequently, ORF constructs were transduced and selected with 200µg/mL nourseothricin sulfate (Gold Biotechnology N-500-1). Each gRNA and ORF combination was counted individually after selection and plated evenly at 20K cells/well in 24-well tissue culture plates, in triplicate. After growth for 1 week, cells were trypsinized and counted using CountBright Absolute counting beads (Thermo Fisher) and propidium iodide as a viability stain with a flow cytometer. Statistical significance was calculated using a one-sided t-test.

Dual EJ Reporter Assay

The BstXI site of the DsRed derivative red fluorescent protein E2-Crimson (E2C) was used to introduce an I-SceI cut site followed by a stop codon and flanked by 8bp microhomology regions. A fusion protein containing E2C, a glycine-rich linker, and GFP was assembled using a sewing PCR reaction, with PCR primers terminating with Gateway attB sites. This PCR product was cloned by a BP reaction into pDONR221, and subsequently transferred to a lentiviral destination vector (pHAGE CMV DEST-Blast) by a Gateway LR reaction. The resulting lentiviral construct was sequence-verified and packaged into lentivirus as described above. To create the reporter cell line, U2-OS osteosarcoma cells were infected with this lentiviral construct at a very low

73

MOI of 0.1 and selected with 20 µg/mL blasticidin. I-SceI was cloned into a lentiviral construct expressing BFP, and this construct was packaged into lentivirus to transduce reporter cells. Ninety-six hours after transduction with I-SceI, reporter cells were analyzed by FACS for BFP, GFP, and E2C expression, in the presence of 1nM bortezomib (Selleck Chemicals). The percent of GFP and E2C cells were quantified from gated BFP$^+$ cells that received the I-SceI virus. MMEJ was quantified as the percent of double-positive GFP$^+$ E2C$^+$ cells, and NHEJ was quantified as the percent of single-positive GFP$^+$ cells.

Chapter 3: Genome-Scale ORFeome Screen Identifies Regulators of PD-L1 Expression and IFNγ Signaling

Attributions

I performed all experiments described herein, analyzed the data, and wrote the following chapter. The barcoded ORFeome library, generated by the lab through cumulative effort over many years, has recently been published (*255*). Steve Elledge and I conceived the study, with helpful input from Tomasz Kula.

I.   Introduction

Upon encountering MHC bearing a foreign antigen, T cells must interpret a balance of costimulatory and coinhibitory signals to determine the outcome of TCR-MHC engagement. Coinhibitory signals, delivered primarily through T-cell surface-expressed CTLA-4 and PD-1, dampen the T cell-mediated immune response and promote tolerance. Tumors frequently upregulate a ligand of PD-1, PD-L1, as a method of immune evasion. PD-1 can also interact with PD-L2, but PD-L2 expression is restricted to myeloid-derived cells (*102*), while PD-L1 expression is far more ubiquitous (*103, 104*) and more functionally important in tumor settings (*256*).

Antibodies that block CTLA-4, PD-1, or PD-L1, thus relieving the coinhibitory signals that prevent T cells from targeting tumor cells, have exhibited stunning

performances in clinical trials (*150, 152, 155, 156*). 22% of melanoma patients achieved

impressively durable long-term remissions from anti-CTLA-4 therapy (*150*), and the

anti-PD-1 antibody pembrolizumab has replaced standard chemotherapy as first-line
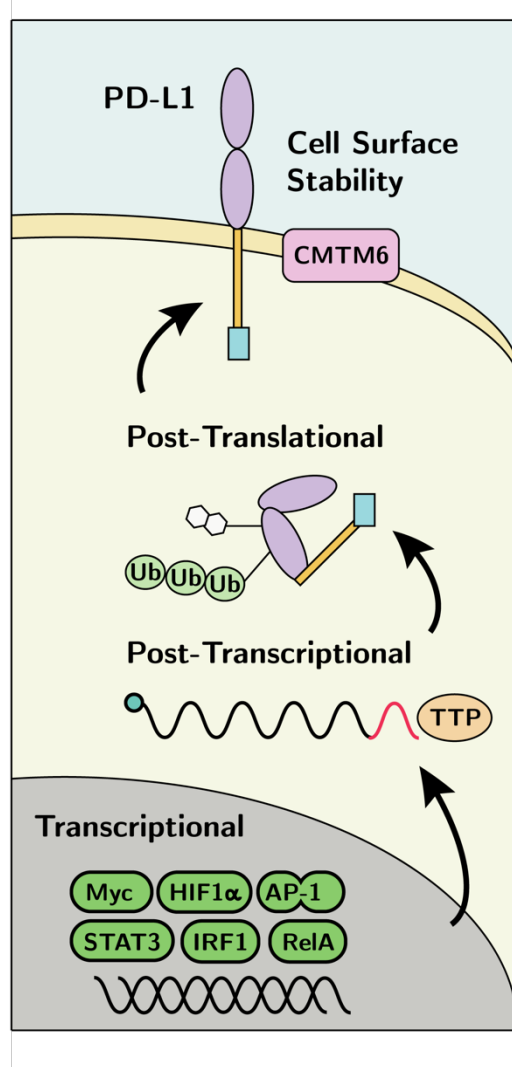
Figure 3-1. Layers of PD-L1 Regulation. Mechanisms of PD-L1 regulation can be broadly categorized as transcriptional, post-transcriptional, and post-translational. These layers of regulation are depicted, along with graphical representations of examples for each regulatory category.

therapy for NSCLC patients with a PD-L1 tumor proportion score (TPS) >50% (*188*). The clinical success of these agents has prompted over 1500 trials and accelerated FDA approval in many settings (*257*).

Although some patients respond durably and favorably to PD-1-PD-L1 blockade, clinical response rates vary within and among tumor types. To maximize the utility of these newly available treatments, it is clinically imperative to understand the parameters that predict clinical response. Several biomarkers have emerged as predictive of positive response to anti-PD-1-PD-L1 treatment, including the diversity of tumor neoantigens, the extent of T cell infiltrate in the tumor, the immune status of the patient, and expression of PD-L1 in the tumor (*101*). Of these prognostic parameters, the latter appears to correlate most directly with clinical outcome (*152, 154, 161-163, 180*).

Thus, understanding the mechanisms of PD-L1 regulation and expression on tumor cells is of major clinical relevance. Several layers of regulation are known to underlie expression of PD-L1 at the cell surface (Figure 3-1). Inflammatory signaling, such as IFN-γ-mediated signaling through the JAK-STAT1 pathway, is a dominant method of PD-L1 transcriptional upregulation (*258*) that also paradoxically upregulates components of the antigen presentation machinery (*259*). PD-L1 can also be transcriptionally up-regulated by Myc, NFκB, HIF1α, and signaling through the PI3K-Akt and MEK-ERK pathways (*101*).

Post-transcriptionally, loss of the PD-L1 3'UTR increases PD-L1 protein levels (*260*). This effect may be partly due to microRNA-mediated degradation of the PD-L1 transcript, which can occur through miR-513, among other microRNA species (*261*). Alternatively, K-Ras signaling promotes stabilization of the PD-L1 transcript through phosphorylation of TTP (tristetraprolin), a protein that binds to the AU-rich elements in the 3'UTR of PD-L1 and promotes its degradation.

Two whole-genome loss-of-function screens uncovered the same positive regulator of PD-L1 protein stability: CMTM6. CMTM6 directly interacts with PD-L1 at the plasma membrane, facilitating its correct endocytic recycling and protecting it from lysosomal degradation (*262*). CMTM4, a family member of CMTM6, can also facilitate PD-L1 cell surface stability (*263*). Finally, PD-L1 post-translational stability can be affected by the cell cycle, as Cdk4/6 phosphorylation of SPOP promotes degradation of PD-L1, mediated by the E3 ligase Cul3 (*264*).

These two fruitful whole-genome loss-of-function screens yielded a shared, interesting mechanism of positive PD-L1 regulation. Yet no systematic whole-genome gain-of-function screen for PD-L1 regulation has been performed to date. Our laboratory recently established a modular, Gateway-compatible barcoded ORFeome library that can be used for systematic gain-of-function genetic screens (*255*). Thus, we sought to apply this unique tool to search for mechanisms of PD-L1 regulation that may not have emerged from loss-of-function screening systems.

78

In addition, we realized that our screen design offered the opportunity to ask a second clinically-relevant question in parallel, with minimal expenditure of additional materials. Separate from the question of predicting intrinsic response to immunotherapy, several groups have addressed the issue of acquired resistance to anti-PD-1 blockade. Among melanoma patients who initially responded, but ultimately relapsed on PD-1 antibody treatment, two types of mutations have been found: defects in antigen presentation (such as β2-microglobulin, *B2M*), and defects in the IFNγ-signaling pathway (*201*). Of the two, IFNγ pathway mutations appear to be predominant, and were confirmed to cause resistance to immunotherapy in both *in vitro* (*265*) and *in vivo* (*266*) screens.

Thus, we screened our barcoded whole-genome ORFeome library both ORFs that promote PD-L1 expression, and ORFs that interfere with IFNγ signaling, using PD-L1 expression as a readout. A more complete understanding of modifiers of the IFNγ signaling pathway, beyond the downstream components of its JAK-STAT-IRF1 axis, may be of prognostic value and may represent alternative forms of intrinsic or acquired resistance to anti-PD-1-PD-L1 blockade.
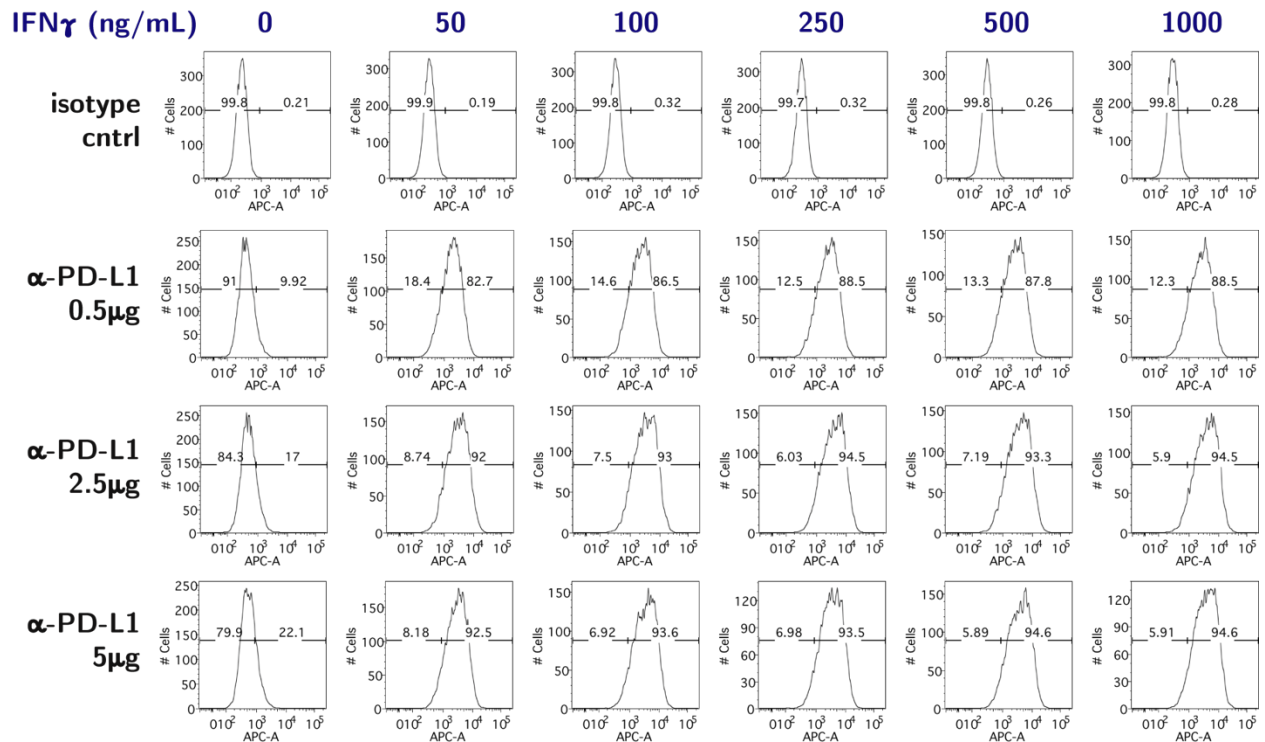
Figure 3-2. IFNγ induces cell surface expression of PD-L1 in H1299 cells. H1299 cells were treated with the indicated doses of IFNγ, and after 48 hours cells were harvested and stained with either monoclonal anti-PD-L1 antibody or isotype control antibody, as indicated. Both antibodies are conjugated to allophycocyanin (APC), and staining was quantified by FACS.

II.   Results

Optimization and design of two parallel genome-scale ORFeome screens

We began by establishing an optimal system in which to screen for induction of

PD-L1 cell surface expression. In particular, we chose a clinically-relevant cell line

background (NSCLC) with low cell surface PD-L1 expression but high inducibility of

PD-L1 in the presence of IFNγ. We tested the human NSCLC cell line H1299 for PD-

L1 cell surface expression through FACS staining, in the absence or presence of IFNγ

treatment. We found that 100ng/mL IFNγ or above induced PD-L1 cell surface expression on 93% of cells, whereas 84% of cells were PD-L1-negative in the absence of IFNγ (Figure 3-2).

From this parental H1299 population, we established an optimal clonal background to serve as the basis for screening. Namely, we selected a clone with low background PD-L1 expression, high inducibility of PD-L1 cell surface expression upon IFNγ treatment, and optimal expression of a reverse tetracycline-controlled transactivator (rtTA) to drive ORF expression from a Tet Response Element (TRE) promoter (Figure 3-3). To select this clone, we transduced parental H1299 cells with a neomycin-selectable rtTA construct at a very low multiplicity of infection (MOI). As part of a tetracycline-on system, addition of doxycycline to cells expressing this rtTA
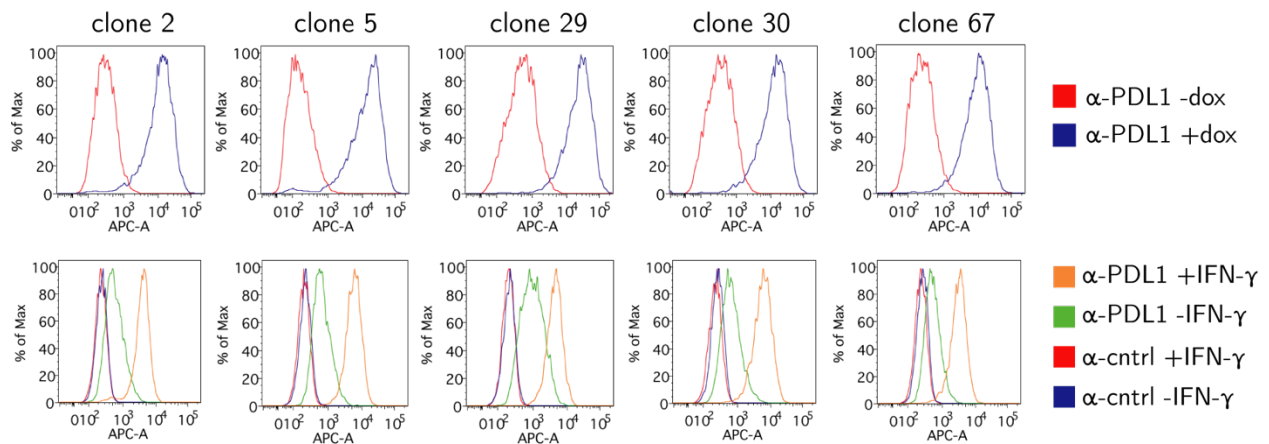
Figure 3-3. Optimization of a clonal background for screening. 96 rtTA-expressing clones were isolated from H1299 cells and tested for several properties relevant to screening. **(Upper)** Clones were infected with a test construct expressing a red fluorophore (E2C) under TRE promoter control. E2C expression was measured by FACS in the APC channel to select clones with high inducibility of TRE-driven E2C expression and minimal leakiness from the promoter. **(Lower)** Clones were stained with APC-conjugated anti-PD-L1 antibody or isotype control in the presence or absence of 100ng/mL IFNγ.

will induce transcription from TRE promoters. After neomycin selection, we plated these cells at low clonogenic density and isolated 96 clones to test for optimal screening properties. Ultimately, we chose clone 2 as having the lowest background and best inducibility of PD-L1, with robust and non-leaky rtTA expression (Figure 3-3).

After expansion, we transduced clone 2 with a 3' barcoded whole-genome ORFeome library under TRE promoter control at an MOI of 0.1 and a representation of 1000 cellular integrations per ORF, in triplicate. We selected the cells with puromycin and afterwards split them into two equivalent populations to perform two FACS-based screens in parallel. We induced ORF expression for 48 hours with doxycycline, stained the cells with a monoclonal antibody recognizing PD-L1, and FACS sorted the cells into PD-L1^{high}, PD-L1^{med}, and PD-L1^{low}-staining populations (Figure 3-4A). One screen, performed in the absence of IFNγ, was designed to identify ORFs that induce PD-L1 cell surface expression. The second parallel screen, performed in the presence of 100ng/mL IFNγ, was designed to identify ORFs that disrupt IFNγ-mediated induction of PD-L1 cell surface expression. In principle, failure to bind α-PD-L1 antibody could result from abrogation of any step in the IFNγ-JAK-STAT signaling pathway or negative transcriptional, post-transcriptional, or post-translational regulation of PD-L1.

ORF barcodes were PCR-amplified from the genomic DNA of all FACS-sorted populations and adapted for next-generation sequencing (Figure 3-4A). After alignment, read counts were analyzed by edgeR camera gene set enrichment analysis to calculate p-values for gene enrichment in each sorted population. Enriched populations were

normalized to the starting pre-sorted population and to the bulk (PD-L1^{med} staining) sorted population. For the PD-L1 Induction screen, either normalization produced an almost identical result, while normalization to the bulk (PD-L1^{med} staining) sorted population was statistically preferable for the IFNγ Response screen.

A genome-scale ORFeome screen identifies positive regulators of PD-L1 cell surface expression

Comparison of p-values for the hits from both screens (Figure 3-4B) demonstrates very little overlap between the two. Unsurprisingly, a top hit for the PD-L1 Induction screen is PD-L1 itself (Figures 3-4B and 3-5A-C), serving as an intrinsic positive control. In addition, several groups of genes exhibited a striking pattern of performance. First, almost every type I interferon included in the ORFeome library induced PD-L1 cell surface expression with a p-value of <0.05 (*IFNA1, IFNA2, IFNA4, IFNA5, IFNA6, IFNA8, IFNA10, IFNA13, IFNA14, IFNA17, IFNB1, IFNL1 IFNW1*), and 10 of these were among the top 30 hits with a p-value of <0.0013.

Type I interferons have been shown to induce PD-L1 expression in melanoma cells, endothelial cells, dendritic cells, and monocytes (*258, 267, 268*). Adding to the tissue diversity in which this phenomenon has been observed, our results show that overexpression of type I interferons can induce PD-L1 in lung cells. Notably, type I interferons have been shown to induce PD-L2 more potently than PD-L1 in melanoma
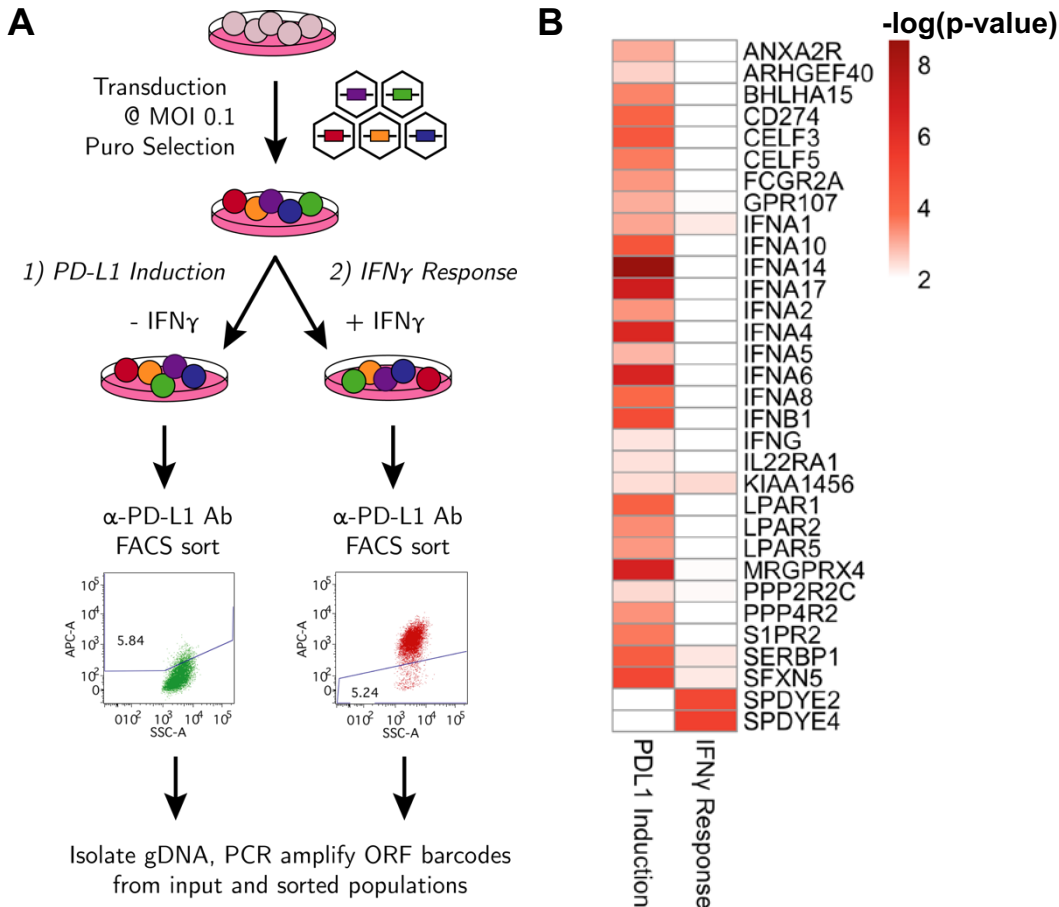
Figure 3-4. Design and results of two parallel whole-genome ORFeome screens.
(A) Schematic depicting the design of two parallel whole-genome ORFeome screens: (1) a
PD-L1 induction screen performed in the absence of IFNγ and (2) a screen for responsiveness
to IFNγ performed in the presence of 100ng/mL IFNγ. After staining with α-PD-L1
antibody, cells were FACS-sorted into high, medium, and low-staining populations. The
population of most interest for each screen is highighted, with ~5% of cells falling into the
indicated gate. **(B)** A heatmap of p-values for the top hits from the PD-L1 Induction and
IFNγ Response screens, plotted as color-coded values according to the adjacent -log$_{10}$ scale.

cells; in three melanoma cell lines, exposure to IFNγ more potently induces PD-L1 than

exposure to the same dose of IFNα or IFNβ (*258*). Our findings demonstrate that, when

overexpressed, type I interferons can induce PD-L1 as strongly as IFNγ in human

NSCLC cells.

Second, we found that multiple members of two gene families scored as strong inducers of PD-L1 cell surface expression (Table 3-1). From the CELF family of RNA-binding proteins, CELF3 and CELF5 scored as hits #9 and #16, respectively. We considered that, even in the presence of IFNγ, genes that strongly induce PD-L1 expression may be enriched in the corresponding PD-L1^{high} cells. When we examined the ORFs that induced the highest levels of PD-L1 in the presence of IFNγ, we saw that CELF3, CELF4, and CELF5 performed as hits #1, #17, and #2, respectively (Table 3-2, Figure 3-5D). In addition to the CELFs, three members of the LPAR gene family are among the top 25 hits in the PD-L1 induction screen: *LPAR1*, *LPAR2*, and *LPAR5* (Table 3-1). These genes belong to a family of G-protein coupled receptors (GPCRs) that signal in response to the small phospholipid lysophosphatidic acid (LPA) (*269*).

Among the top PD-L1-inducing ORFs, we observed several additional genes of note. For example, Fcγ Receptor IIA (*FCGR2A*) enriched in the PD-L1^{high} population (Figures 3-5A-C). In a prior whole-genome antibody-based FACS screen for cell surface regulation (*data not shown*), we observed very strong enrichment of *FCGR2A*, presumably due to its binding of the Fc region of the primary antibody used for staining and FACS sorting. In this screen, to mitigate noise introduced by this potential effect, we utilized a primary antibody that is directly conjugated to the red fluorophore APC through its Fc region, which should sterically obstruct binding of the Fcγ Receptor. This strategy appears to have partially, but not entirely, dampened unintentional binding of the Fcγ Receptor to our primary antibody.
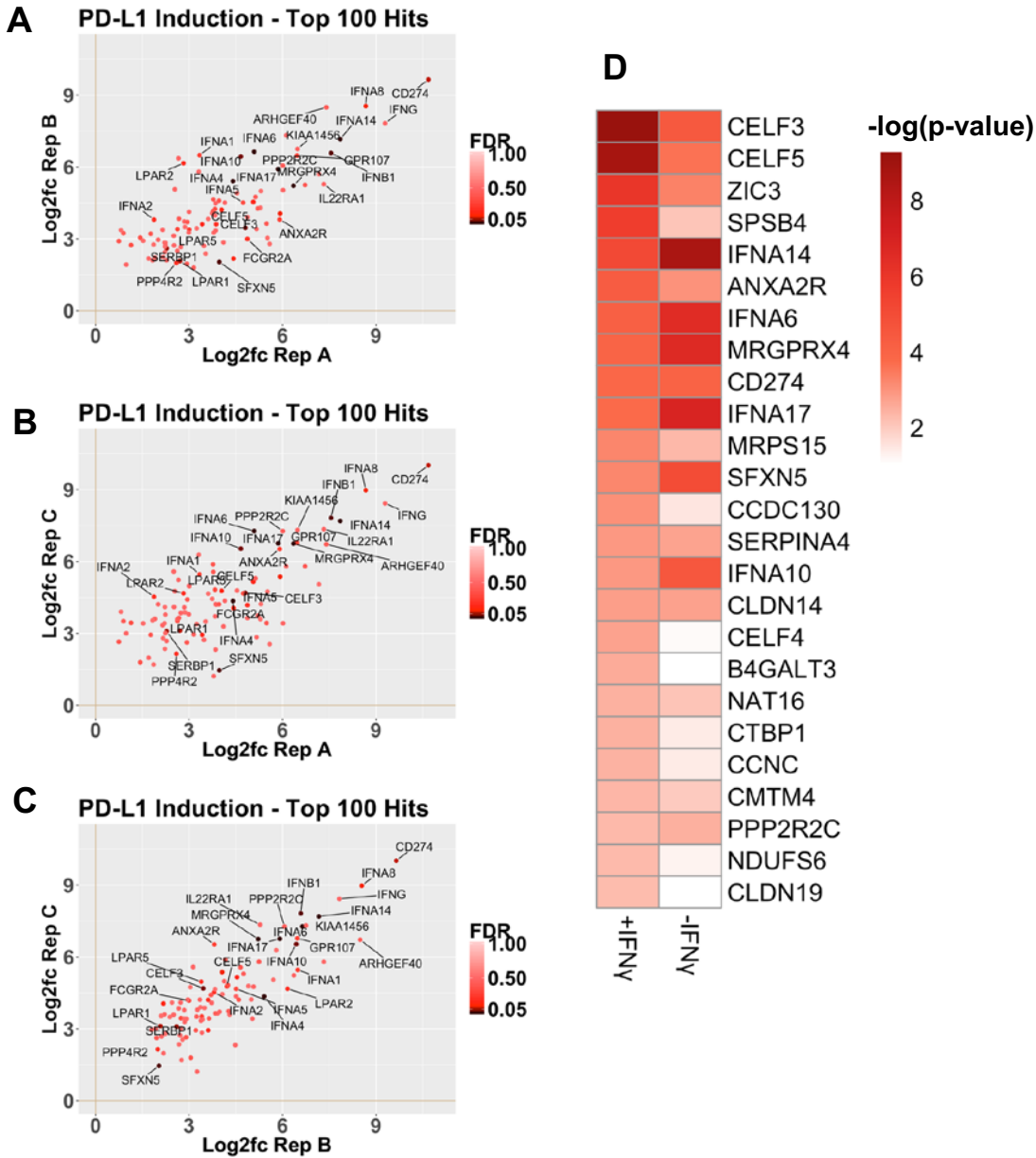
Figure 3-5. Performance of ORFs that induce PD-L1 cell surface expression. (A-C) Average log2-fold-changes between the PD-L1^{high}-sorted versus unstained pre-sorted populations are plotted for the top 100 genes, for each indicated pair of replicates. Genes are represented by a single dot and are color-coded according to their FDR, determined by edgeR camera gene set enrichment analysis. This data was obtained in the absence of IFNγ. **(D)** The strongest PD-L1-inducing ORFs in presence of IFNγ (the IFNγ Response Screen) are displayed as a color-coded heatmap of -log$_{10}$(p-values); all p-values were determined by edgeR camera gene set enrichment analysis. The p-value for each gene is shown in comparison with its p-value for PD-L1 enrichment obtained in the absence of IFNγ (the PD-L1 induction screen).

Table 3-1. List of hits from the PD-L1 Induction Screen. The top 25 hits from the whole-genome PD-L1 induction screen are listed, along with a small selection of additional genes of interest. Gene rank and p-value were determined by edgeR camera gene set enrichment analysis, and they are listed along with the full gene names for each hit. Two gene families are highlighted: CELFs (yellow) and LPARs (green).

Rank	Gene	P-Value	Gene Name
1	IFNA14	2.04E-09	interferon alpha 14
2	IFNA17	1.19E-07	interferon alpha 17
3	MRGPRX4	2.23E-07	MAS related GPR family member X4
4	IFNA6	3.00E-07	interferon alpha 6
5	IFNA4	3.86E-07	interferon alpha 4
6	SFXN5	9.18E-06	sideroflexin 5
7	IFNB1	1.32E-05	interferon beta 1
8	IFNA10	2.83E-05	interferon alpha 10
9	**CELF3**	3.37E-05	CUGBP Elav-like family member 3
10	SERBP1	5.67E-05	SERPINE1 mRNA binding protein 1
11	**LPAR1**	7.64E-05	lysophosphatidic acid receptor 1
12	CD274	8.95E-05	CD274 molecule
13	CWF19L1	1.31E-04	CWF19 like 1, cell cycle control (S. pombe)
14	IFNA8	1.43E-04	interferon alpha 8
15	S1PR2	2.40E-04	sphingosine-1-phosphate receptor 2
16	**CELF5**	2.44E-04	CUGBP Elav-like family member 5
17	BHLHA15	3.17E-04	basic helix-loop-helix family member a15
18	CRTAC1	3.32E-04	cartilage acidic protein 1
19	**LPAR2**	4.00E-04	lysophosphatidic acid receptor 2
20	ZIC3	4.61E-04	Zic family member 3
21	PPP4R2	4.77E-04	protein phosphatase 4 regulatory subunit 2
22	IFNA2	4.88E-04	interferon alpha 2
23	FCGR2A	5.28E-04	Fc fragment of IgG receptor IIa
24	**LPAR5**	5.57E-04	lysophosphatidic acid receptor 5
25	IFNA1	7.83E-04	interferon alpha 1
54	IL22RA1	4.02E-03	interleukin 22 receptor subunit alpha 1
57	IFNW1	4.34E-03	interferon omega 1
58	IFNG	4.38E-03	interferon gamma
99	IFNA13	8.30E-03	interferon alpha 13
110	**TGFB1**	9.39E-03	transforming growth factor beta 1
113	**CMTM4**	9.78E-03	CKLF like MARVEL transmembrane domain containing 4

In addition, we found that overexpression of the α subunit of IL-22 Receptor (*IL22RA1*) increases PD-L1 cell surface expression (Table 3-1, Figures 3-5A-C). IL-22 is a member of the IL-10 family of cytokines that typically binds to a heterodimeric receptor formed from *IL22RA1* and *IL-10RB*. Binding of IL-22 to this receptor complex leads to JAK-STAT signaling, usually mediated by STAT1, which canonically transcriptionally induces PD-L1 through IRF-1 (*258, 270-272*). Overexpression of the α subunit of this receptor may lead to auto-phosphorylation and activation of this well-characterized PD-L1 regulatory pathway.

Of note, we recovered *CMTM4* as a positive PD-L1 regulator with a p-value of 0.00978 in the absence of IFN-γ and a stronger p-value of 0.00476 in the presence of IFN-γ (Tables 3-1 and 3-2). In a reciprocal approach, *CMTM4* and *CMTM6* were identified as PD-L1 positive regulators from loss-of-function screens that isolated the PD-L1[low] FACS-sorted population, using either a CRISPR-based (*262*) or mutagenized haploid cell line library (*263*). Inversely, in our case, we see that overexpression of the *CMTM4* ORF results in its enrichment in the PD-L1[high] sorted population; *CMTM6* was not in our ORFeome library.

Finally, we observe that overexpression of TGF-β (*TGB1*) induces PD-L1 in our NSCLC cell line (Table 3-1). TGF-β is generally an anti-inflammatory cytokine, though it has been shown to exert tissue-specific effects in regulating PD-L1 expression. For example, TGF-β represses PD-L1 expression in monocytes (*273*) and tubular epithelial cells *in vitro (274)*. In contrast, TGF-β induces PD-L1 expression in dendritic cells *in*

vitro (*275, 276*) and in a model of pancreatic islet transplantation *in vivo* (*277*). We see

TGF-β enriching in the PD-L1^{high} sorted cell population, as well as an additional

member of the TGF-β superfamily, BMP4, with a p-value of 0.0045 (Supplemental

Table S3-1).

Table 3-2. List of ORFs that induce PD-L1 in the presence of IFNγ. Genes that produced the highest PD-L1 signal from the IFNγ response screen were determined by normalizing the PD-L1^{high}-sorted population to the PD-L1^{medium}-sorted bulk population. P-values and ranks were calculated by edgeR camera gene set enrichment analysis and the CELF family genes are highlighted (yellow).

Rank	Gene	P-Value	Gene Name
1	CELF3	5.10E-10	CUGBP Elav-like family member 3
2	CELF5	1.60E-09	CUGBP Elav-like family member 5
3	ZIC3	9.10E-07	Zic family member 3
4	SPSB4	2.10E-06	splA/ryanodine receptor domain and SOCS box containing 4
5	IFNA14	6.50E-06	interferon alpha 14
6	ANXA2R	4.30E-05	annexin A2 receptor
7	IFNA6	6.70E-05	interferon alpha 6
8	MRGPRX4	9.80E-05	MAS related GPR family member X4
9	CD274	1.40E-04	CD274 molecule
10	IFNA17	1.70E-04	interferon alpha 17
11	MRPS15	5.50E-04	mitochondrial ribosomal protein S15
12	SFXN5	5.90E-04	sideroflexin 5
13	CCDC130	8.60E-04	coiled-coil domain containing 130
14	SERPINA4	1.20E-03	serpin family A member 4
15	IFNA10	1.20E-03	interferon alpha 10
16	CLDN14	1.50E-03	claudin 14
17	CELF4	1.80E-03	CUGBP Elav-like family member 4
18	B4GALT3	2.70E-03	"beta-1,4-galactosyltransferase 3"
19	NAT16	3.40E-03	N-acetyltransferase 16 (putative)
20	CTBP1	3.40E-03	C-terminal binding protein 1
21	CCNC	3.60E-03	cyclin C
22	CMTM4	4.20E-03	CKLF like MARVEL transmembrane domain containing 4
23	PPP2R2C	4.80E-03	protein phosphatase 2 regulatory subunit B gamma
24	NDUFS6	4.80E-03	NADH:ubiquinone oxidoreductase subunit S6
25	CLDN19	5.20E-03	claudin 19

Although *SMC4* is not present in this ORFeome library, overexpression of wild-type Smad proteins typically increase TGF-β signaling activity (*278, 279*). The fact that SMADs 1, 2, 3, 4, and 9 failed to score significantly in our screen suggests that perhaps TGF-β is inducing PD-L1 signaling through its SMAD-independent signaling pathway in this context, which can engage the MAPK/Erk pathway, p38 MAPK, Rho/Rac, or PI3K/Akt signaling pathways downstream (*280*); BMP4 can also activate these downstream SMAD-independent pathways (*281*). The effect of TGFβ is therapeutically important, as its presence in tumors correlates with intrinsic resistance to anti-PD-L1 blockade (*199*).

GSEA highlights GPCR signaling as positively regulating PD-L1

Because a large number of genes scored in the PD-L1 Induction screen (edgeR camera gene set analysis found 115 genes with a p-value <0.01), we performed two gene set enrichment pathway analyses to identify common pathways and networks among these genes. We removed all type I interferons from the analyses, which otherwise predominate the list of pathways obtained. We used Ingenuity Pathway Analysis (IPA) and DAVID Functional Annotation (*282*) to determine the pathways and networks that result in cell surface expression of PD-L1.

Both analyses recognized the large number of G-protein coupled receptors (GPCRs) that scored in the screen; 12 GPCRs enriched with a p-value <0.01, in addition to a guanine nucleotide exchange factor (GEF) and a monomeric G-protein
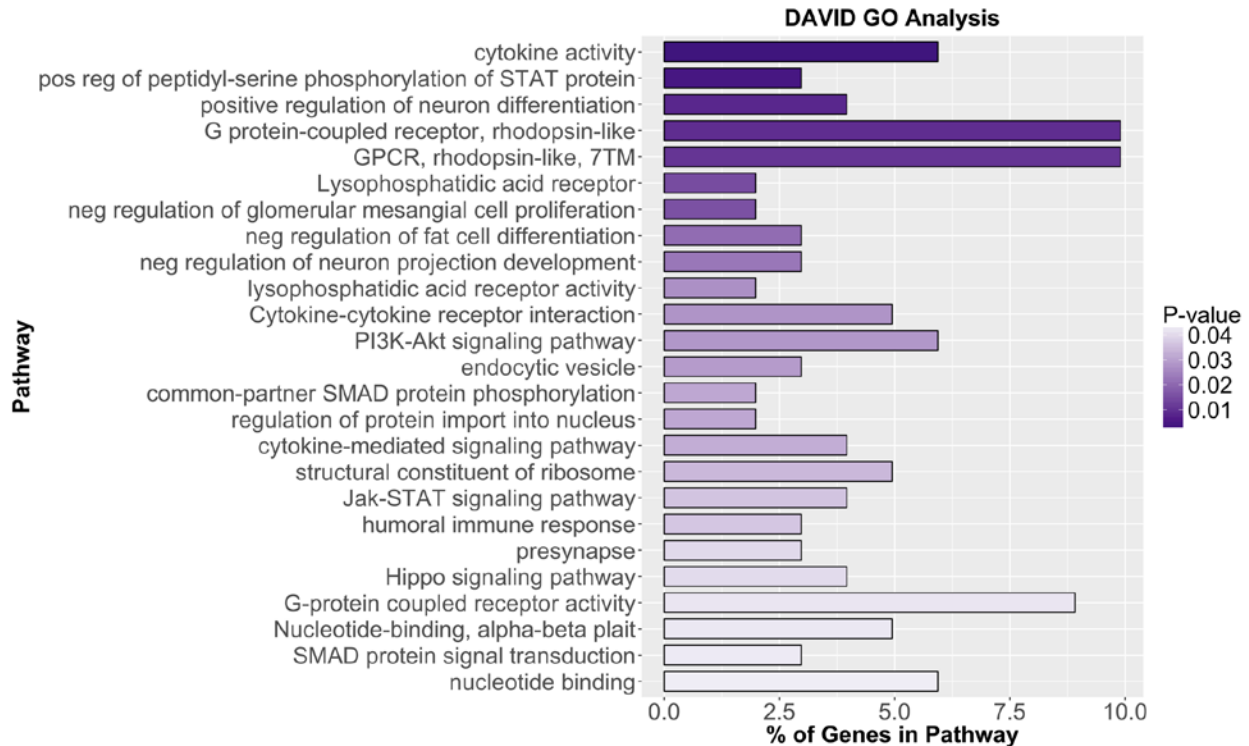
Figure 3-6. DAVID Gene Ontology pathway analysis of PD-L1 induction screen.
Genes that significantly enriched in the PD-L1 Induction Screen (with a p-value of <0.01)
were submitted to DAVID for functional annotation, after filtering out all type I interferons
(which otherwise predominated the analysis). The top 25 identified pathways are shown,
ordered and color-coded by the p-value returned by DAVID for their enrichment. The
percent of genes in the pathway that scored in our screen is plotted on the x-axis.

(Figures 3-6 and 3-7). Three of these GPCRs are members of the same family, the

lysophosphatidic acid receptor family (LPARs). This GPCR family responds to

glycerophospholipid lysophosphatidic acid species (LPA) to activate a variety of

downstream pathways including the Rho/Rac, Ras/MAPK, and PI3K/Akt pathways

(*283*). All of these pathways have been linked to transcriptional or post-transcriptional

regulation of PD-L1 in one or more contexts (*284-286*).
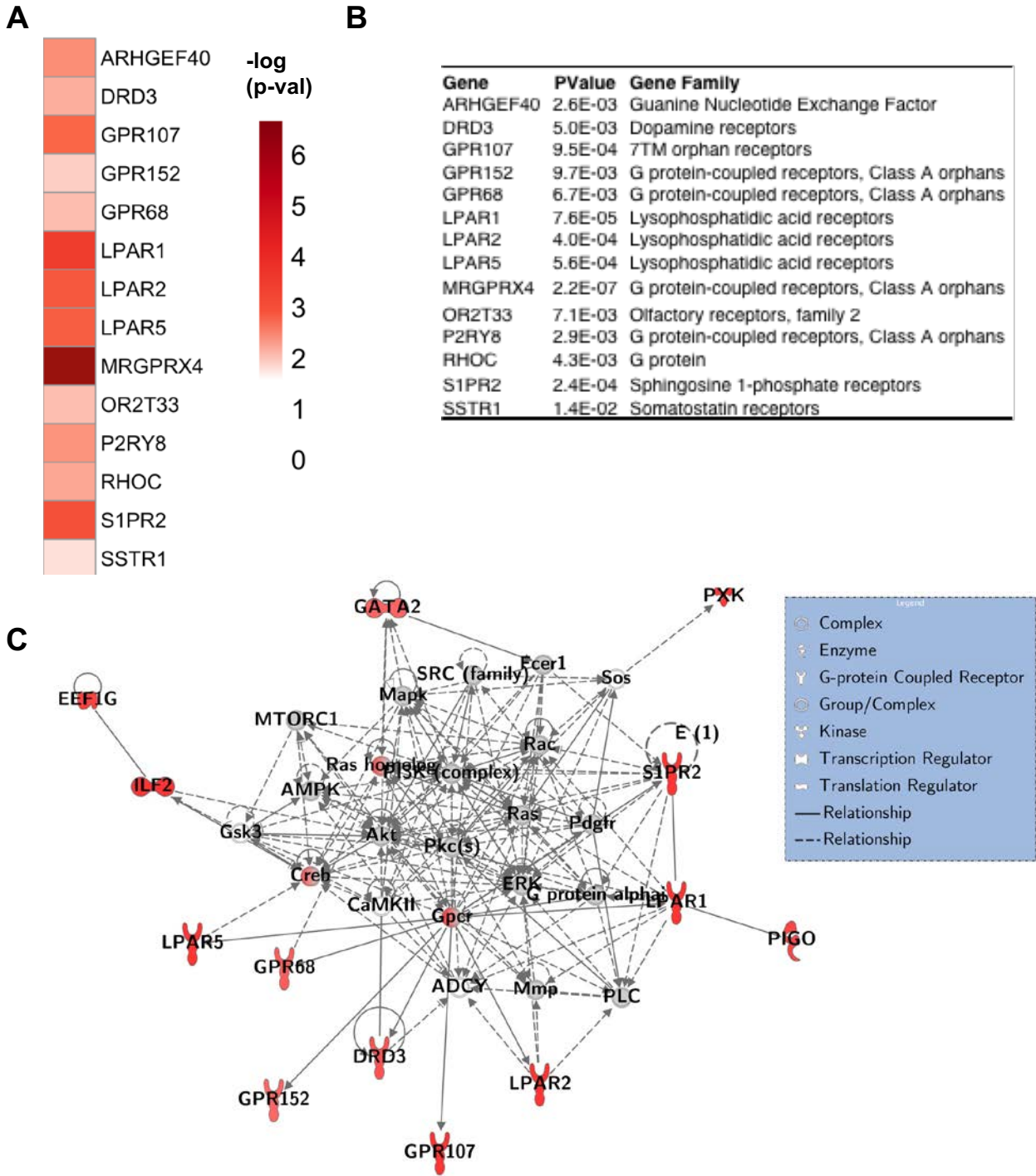
Figure 3-7. GPCRs increase PD-L1 cell surface expression. **(A)** P-values determined by edgeR camera gene set enrichment analysis are plotted as a heatmap for the 12 GPCRs that scored with a p-value <0.01, as well as one GEF and the only monomeric G protein that also met this cutoff. **(B)** List of genes plotted in (A) along with their complete gene names. **(C)** A top network identified by Ingenuity as enriched in the PD-L1 Induction screen. Red shading indicates genes that scored in the screen with a p-value <0.01 determined by edgeR, and the legend identifies the type of protein in the network its shape.

92

Examination of the barcodes corresponding to LPARs 1, 2, and 5 showed

consistent, strong enrichment in the PD-L1^{high} sorted cell population across all three

replicates (Figure 3-8). There are two additional LPAR family members in the library

(LPAR4 and LPAR6) that did not score, and each of these genes has only 2 or 3

barcodes; while *LPAR6* does not appear to enrich, the barcodes for *LPAR4* are

inconsistent (Supplemental Figure S3-1). It is important to note that the H1299 cell

line is *KRAS* wild-type but harbors an activating *NRAS* mutant; thus, it is perhaps the

PI3K-Akt downstream signaling of the LPAR pathway that is predominantly important

in this context.

IPA identified several additional pathways of note (Supplemental Figure S3-2A).

First, it discovered a network of enriched genes centered around activation of Myc,

which is known to be capable of transcriptionally up-regulating PD-L1 (Supplemental

Figure S3-2B) (*287, 288*), though *MYC* itself is not in the ORFeome library. Second,

both IPA and DAVID analysis identified the PI3K-Akt pathway as positively regulating

PD-L1 cell surface expression. Finally, IPA identified the Wnt pathway as enriching in

the PD-L1 induction screen, partially by virtue of Wnt3A (p-value=0.0077) and Wnt16
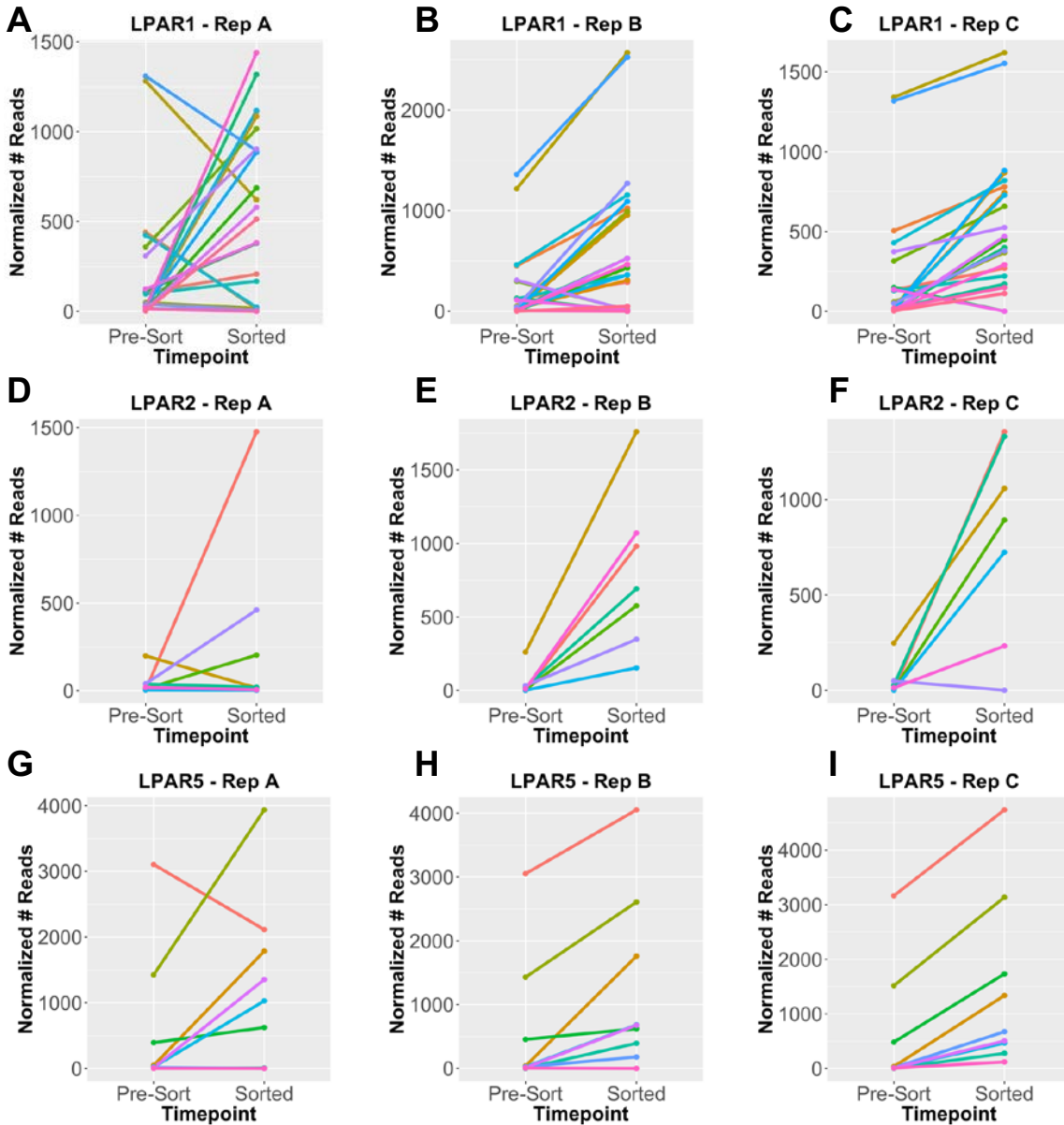
Figure 3-8. Performance of barcodes for LPAR genes that scored in the PD-L1 Induction screen. Normalized read counts are shown for two timepoints: the pre-sorted unstained library of cells, and the PDL1high sorted cell population. Each line represents an individual barcode for the corresponding gene. Performance in all 3 replicates is depicted for *LPAR1* (**A-C**), *LPAR2* (**D-F**), and *LPAR5* (**G-I**).

(p-value=0.026) themselves enriching. The Wnt pathway has been associated with

immunoevasion (*289*) but not directly connected to PD-L1 expression.

CELF3 is a positive regulator of PD-L1 expression that is amplified in lung cancer

Among the genes that enhanced PD-L1 cell surface expression, the most striking previously unreported putative regulators are the CELF genes. *CELF3* (rank=9, p-value=3.37×10^{-5}) and *CELF5* (rank=16, p-value=2.44×10^{-4}) were among the top genes enriched in PD-L1^{high} cells in the absence of IFNγ. Another family member, *CELF4,* also enriched in this sorted cell population, but not as strongly (p-value=0.072). However, in the presence of IFNγ, *CELF3* (rank=1, p-value=5.11×10^{-10}), *CELF4* (rank=17, p-value=1.83×10^{-3}), and *CELF5* (rank=2, p-value=1.63×10^{-9}) all scored as top hits. Examination of the performance of each barcode corresponding to *CELF3* and *CELF5* in the PD-L1 induction screen showed consistent, robust enrichment of barcodes across all three replicates (Figures 3-9A-F).

The CUG-BP, Elav-like Family (CELF) gene family contains 6 members in *homo sapiens*: *CELF1-6* (*290*). *CELF3*, *CELF4*, and *CELF5* are the only members of the CELF family included in the ORFeome library, and as discussed above, they all robustly induced PD-L1 cell surface expression. These genes encode RNA-binding proteins which modulate several steps of RNA processing including alternative splicing, deadenylation or polyadenylation, mRNA stability, and translation efficiency (*290-292*). Though they share a common structure of three RNA-binding domains (RRM) with a divergent domain (DD) between RRMs 2 and 3, CELF proteins fall into two subfamilies based on their expression and activity patterns: CELF1-2 and CELF3-6 (*293*). Whereas

Figure 3-9. CELF3 positively regulates PD-L1 and is amplified in lung tumors. (A-C) Normalized read counts for barcodes corresponding to *CELF3* in the pre-sorted unstained library population and the PD-L1^{high} sorted cell population, in all three replicates. Each line corresponds to an individual barcode associated with CELF3. **(D-F)** Normalized read counts as described above, for *CELF5*. **(G-H)** Analysis of *CELF3* (G) and *CELF5* (H) expression from RNA-seq data (RPKM) across a panel of 1019 cancer cell lines, separated by tissue type as indicated. **(I)** Frequency of amplification of the indicated genes in lung adenocarcinoma. **(J)** Kaplan-Meier survival curve depicting the effect of *CELF3* amplification on survival in lung adenocarcinoma patients.
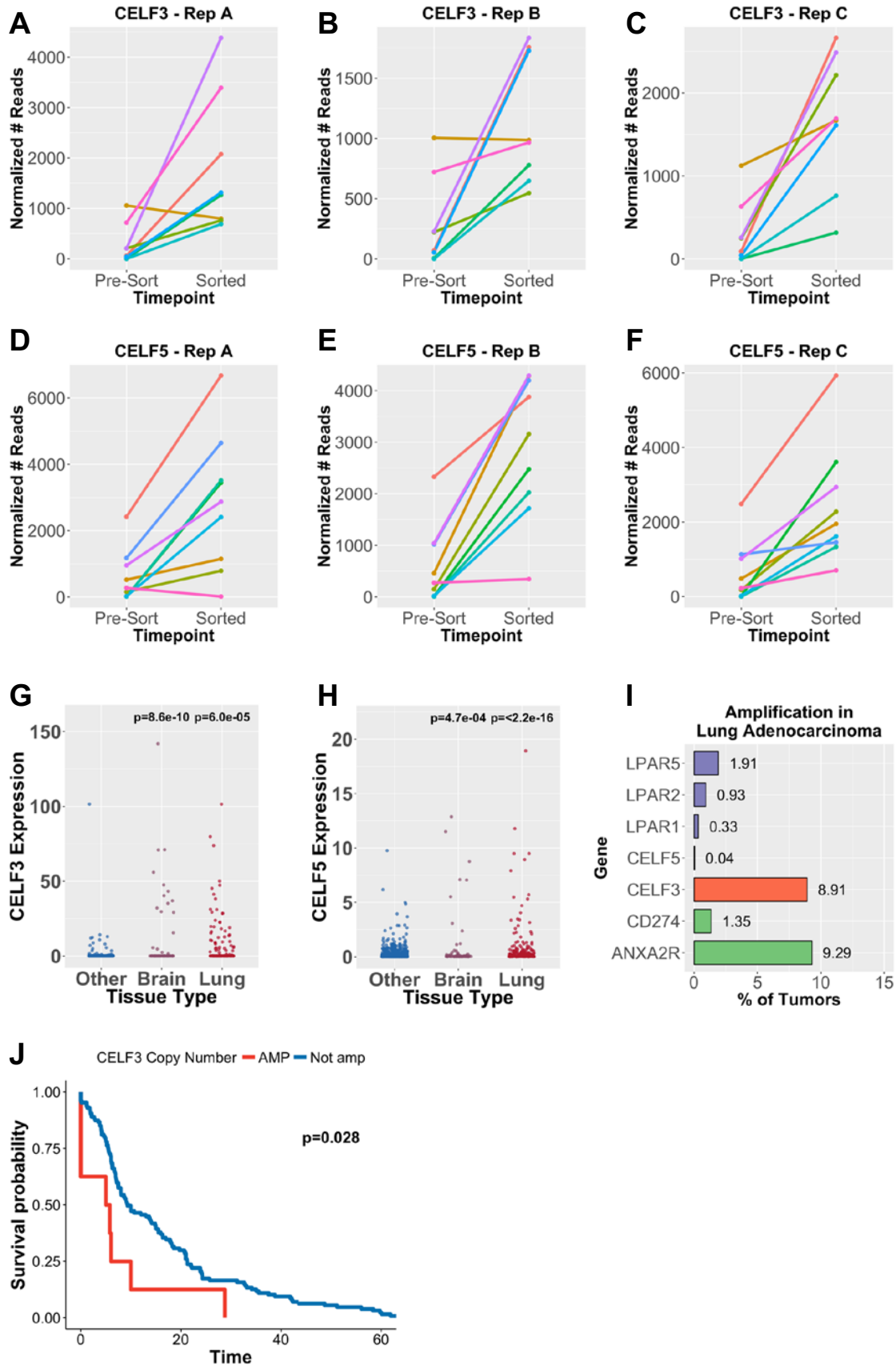
Figure 3-9 (Continued). CELF3 positively regulates PD-L1 and is amplified in lung tumors.

CELF1-2 expression is more ubiquitous, CELF3-6 expression in adult tissue and during embryogenesis is restricted almost entirely to the nervous system (*291, 292, 294*). CELF1-2 have received far more attention than CELF4-6 in the literature, with no RIP-seq or CLIP-seq data available for CELF4-6, and no common RNA-binding consensus sequence being known.

We asked if this family of putative PD-L1 regulators is expressed in lung cancer. To address this question, we surveyed a compilation of RNA-seq data from 1019 cancer cell lines. We found that both CELF3 and CELF5 are expressed specifically in lung cancer cell lines, and also to some degree in cancer cell lines derived from brain tissues (Figures 3-9G and 3-9H). We also examined whether CELF family members undergo mutation or copy number alterations in tumors. We find that CELF3 is amplified in 8.91% of lung adenocarcinoma but is not subject to similar levels of copy number alteration in other tumors (Figure 3-9I).

Unlike *CELF4* and *CELF5*, *CELF3* falls into a region of frequent focal and large-scale amplification in lung adenocarcinoma, 1q21.3 (*295*). This focal region contains a proto-oncogene driver, aryl hydrocarbon receptor nuclear translocator (*ARNT*), which may explain the frequency of its amplification in lung cancers. We performed a survival analysis to ask if *CELF3* amplification has a statistically significant correlation with prognosis using a dataset from lung adenocarcinoma patients (*295*); we found that increased copy number of CELF3 is associated with worse prognosis (Figure 3-9K). Although this finding is consistent with its putative effect on PD-L1 levels, there are
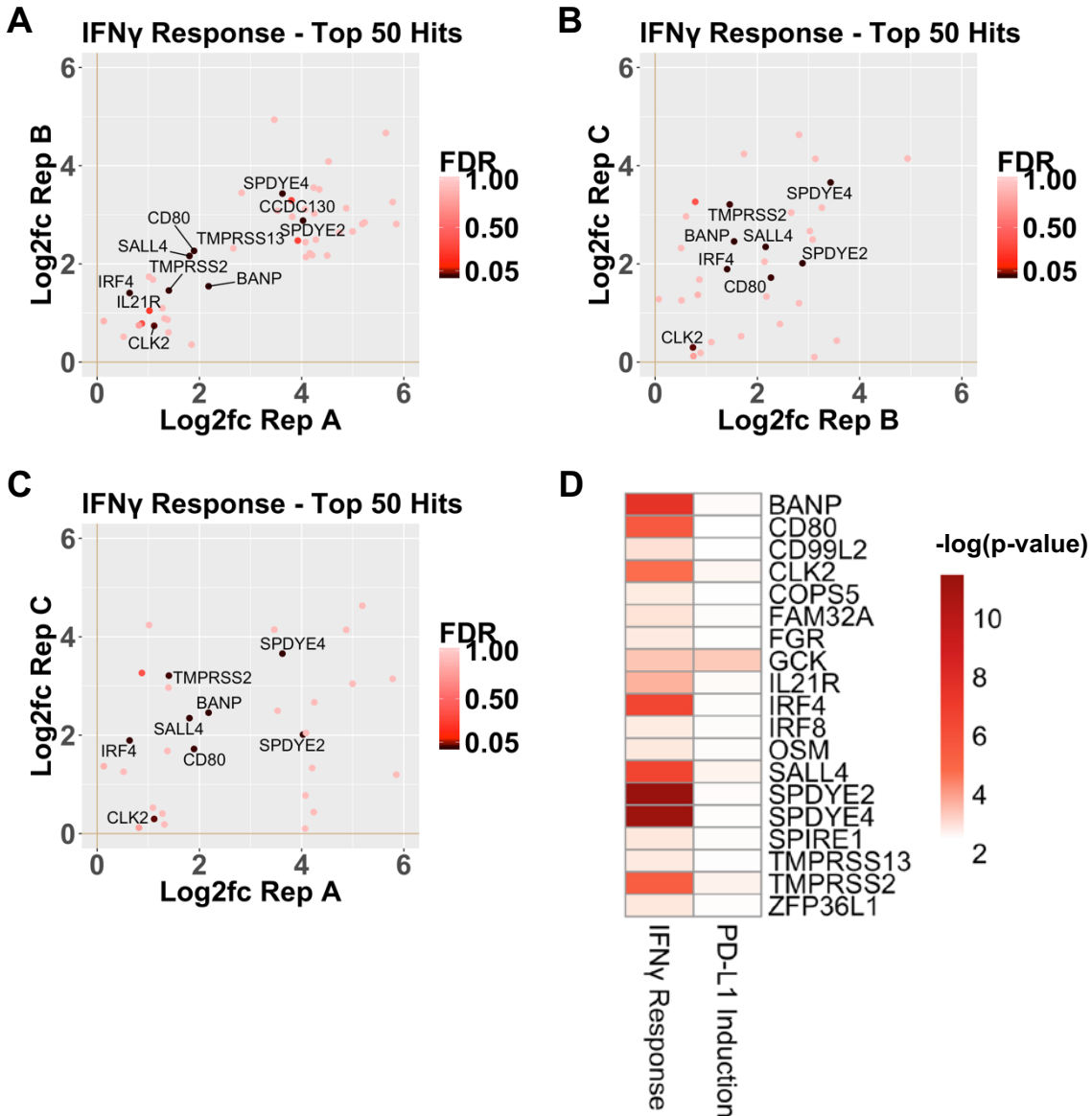
Figure 3-10. Results of the IFNγ Response Screen. (A-C) Average log2-fold-changes between the PD-L1^{low} sorted versus unstained pre-sorted populations are plotted for the top 50 genes, for each pair of replicates. Genes are represented by a single dot and are color-coded according to their FDR, determined by edgeR camera gene set enrichment analysis. This data was obtained in the presence of IFNγ. **(D)** The ORFs which enrich most strongly in the IFNγ Response screen are displayed as a heatmap of -log$_{10}$(p-values); all p-values were determined by edgeR camera gene set enrichment analysis. The p-value for each gene is shown in comparison with its p-value in the PD-L1 induction screen and plotted by color-coding according to the adjacent color scale.

many co-variates with CELF3 amplification, including amplification of the driver gene *ARNT*.

SPDYE2/4 or CLK2 Overexpression Interferes with IFNγ Signaling

In addition to screening for ORFs that induce PD-L1 expression, we also performed a parallel FACS-based ORFeome screen in the presence of IFNγ, to screen for ORFs that block induction of PD-L1. The object of this screen was to identify ORFs that interfere with the IFNγ signaling response, using PD-L1 as a readout of functional IFNγ signaling. Because defects in IFNγ signaling appear to be the most significant cause of acquired resistance to immunotherapy (*201, 265*), it is important to identify genetic modifiers of this pathway.

Unlike the PD-L1 Induction screen, in which a large set of genes enriched in the PD-L1^{high} population with a significant p-value, the IFNγ Response screen revealed two genes with a particularly high p-value for enrichment in the PD-L1^{low} population: *SPDYE2* and *SPDYE4* (Figures 3-10 and 3-11, Table 3-3). These genes encode two members of the Speedy family of non-cyclin CDK activators. Speedy genes are mammalian orthologs of the *Xenopus* gene *XRINGO* (*Xenopus* Rapid Inducer of G2-M in Oocytes), and they include 12 related family members, 4 of which are in our ORFeome library (*SPDYC, SPDYE2, SPDYE4,* and *SPDYE9P*) (*296*).

Although most of these family members promote cell cycle progression through activation of CDK1, CDK2, and CDK5 (*297*), Speedy E (*SPDYE1*) has been shown to

impair cell cycle progression of mammalian U2-OS osteosarcoma cells by sequestration of CDK1 from its substrates (*298*). Thus, it remains unclear whether *SPDYE2* and *SPDYE4* impair or enhance cell cycle progression in mammalian cells. Nonetheless, it may be relevant that PD-L1 expression is regulated by the cell cycle, through CDK4/6 phosphorylation of SPOP, which mediates ubiquitination and degradation of

Table 3-3. List of Genes that Scored in the IFNγ Response Screen. Genes that enriched in the PD-L1^{low} sorted cell population after IFNγ treatment were ranked according to their p-value determined by edgeR camera gene set enrichment analysis. The top two families of genes are highlighted: *SPDYE* genes (yellow) and *TMPRSS* genes (green).

Rank	Gene	P-Value	Gene Name
1	SPDYE2	3.33E-12	speedy/RINGO cell cycle regulator family member E2
2	SPDYE4	5.39E-12	speedy/RINGO cell cycle regulator family member E4
3	BANP	2.70E-08	BTG3 associated nuclear protein
4	SALL4	2.73E-07	spalt like transcription factor 4
5	IRF4	2.94E-07	interferon regulatory factor 4
6	CD80	2.50E-06	CD80 molecule
7	TMPRSS2	4.18E-06	transmembrane serine protease 2
8	CLK2	1.84E-05	CDC like kinase 2
9	IL21R	4.14E-04	interleukin 21 receptor
10	GCK	9.82E-04	glucokinase
11	CD99L2	4.08E-03	CD99 molecule like 2
12	FAM32A	6.21E-03	family with sequence similarity 32 member A
13	SPIRE1	1.19E-02	spire type actin nucleation factor 1
14	ZFP36L1	1.66E-02	ZFP36 ring finger protein like 1
15	OSM	1.75E-02	oncostatin M
16	TMPRSS13	2.07E-02	transmembrane serine protease 13
17	FGR	2.31E-02	FGR proto-oncogene, Src family tyrosine kinase
18	IRF8	2.84E-02	interferon regulatory factor 8
19	COPS5	3.00E-02	COP9 signalosome subunit 5
20	NAE1	3.94E-02	NEDD8 activating enzyme E1 subunit 1
21	HSBP1	5.73E-02	heat shock factor binding protein 1
22	GPA33	6.43E-02	glycoprotein A33
23	TBCA	7.60E-02	tubulin folding cofactor A
24	NAPA	7.61E-02	NSF attachment protein alpha
25	GRHL2	8.38E-02	grainyhead like transcription factor 2

PD-L1 (*264*). Examination of the barcodes corresponding to *SPDYE2* and *SPDYE4* showed strong enrichment in the PD-L1^{low} sorted cell population of all barcodes corresponding to the *SPDYE*2 and *SPDYE4* ORFs across all three replicates (Figure 3-11).

Although *SPDYE2* and *SPDYE4* outperform the rest of the library, it is noteworthy that CD80 is among the top ORFs that block PD-L1 induction in the presence of IFN-γ. CD80 is a direct ligand of PD-L1, and it binds to the same region of PD-L1 as the monoclonal antibody we used for screening (*299*). CD80 overexpression on the cell surface may physically interfere with antibody binding, due to direct binding and sequestration of PD-L1. It is also significant that, from a whole-genome library, two related transmembrane serine proteases (*TMPRSS2* and *TMPRSS13*) appear to block PD-L1 induction; overexpression of these serine proteases may interfere with cell surface display of PD-L1 in some way.

The most clinically relevant gene that impaired the ability of IFNγ to induce PD-L1 is CDC-like kinase 2 (*CLK2*). *CLK2* is very frequently amplified and overexpressed in breast (*300*) (21.36% of METABRIC), prostate (29.9%), and pancreatic (22.2%) cancer (*301*). These three tumor types notoriously respond poorly to immunotherapy, particularly breast and prostate cancer (*189*). Based on our results, we would hypothesize that tumors with *CLK2* amplification may have intrinsic resistance to immunotherapy. Segregating patients based on *CLK2* status may improve outcomes in these immunotherapy-intractable tumor types.
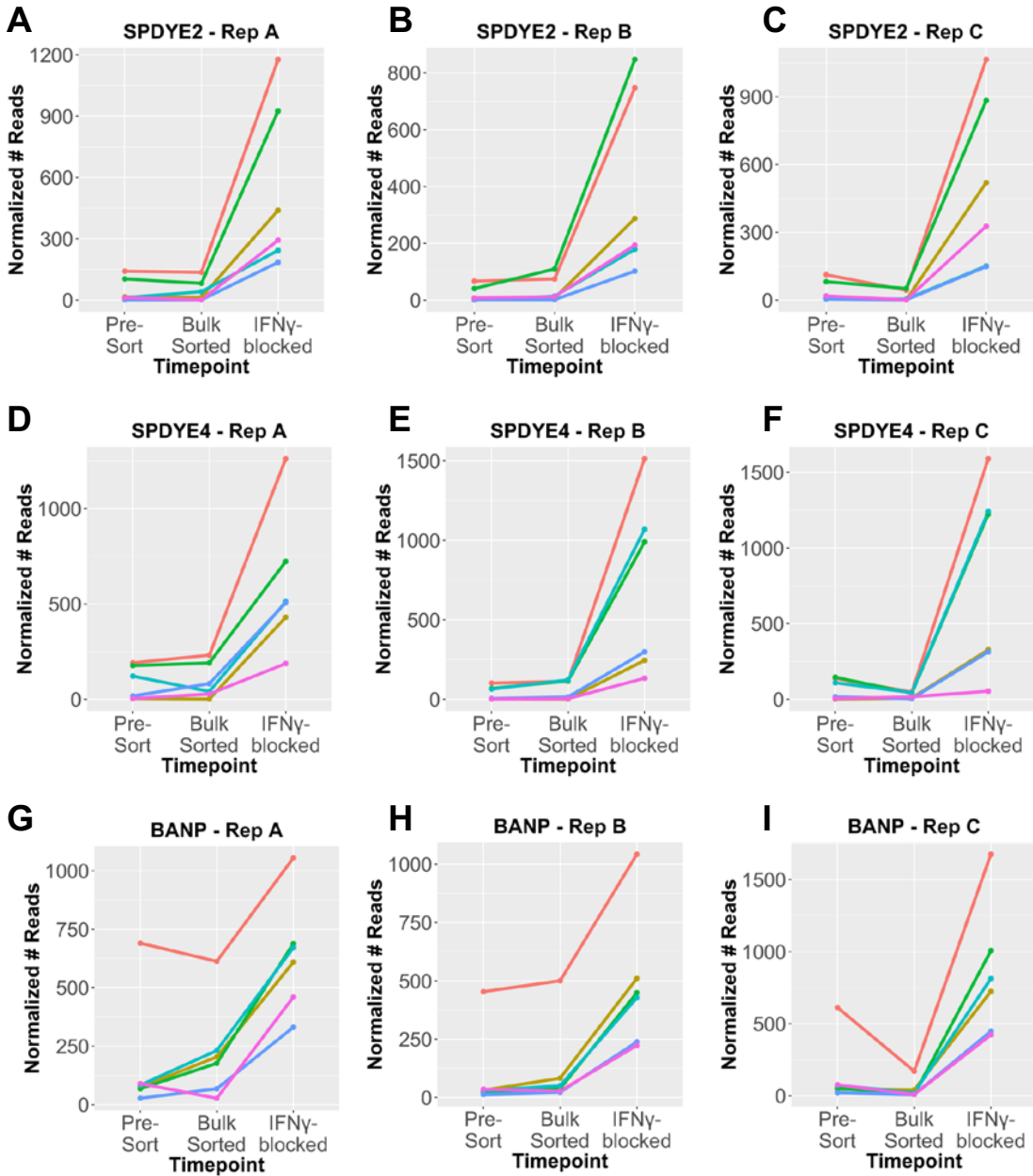
Figure 3-11. Performance of barcodes for top hits in the IFNγ Response Screen. Normalized read counts of barcodes mapped to *SPDYE2* (**A**), *SPDYE4* (**B**) and *BANP* (**C**) are shown for three timepoints of the IFNγ response screen: the pre-sorted unstained library of cells, the bulk PD-L1$^{\text{medium}}$ sorted cell population, and the PDL1$^{\text{low}}$ sorted cell population. Each line represents an individual barcode corresponding to the indicated gene. Performance in each replicate is depicted in a separate panel.

Furthermore, a straightforward mechanism for the role of *CLK2* in antagonizing IFNγ signaling immediately suggests itself from the literature. *CLK2* is a dual-specificity kinase that phosphorylates a variety of serine/threonine and tyrosine-containing substrates, one of which is *PTPN1* (*302*). *PTPN2* has been shown to antagonize JAK-Stat signaling by directly dephosphorylating JAK1 and STAT1 (*266, 303, 304*). Thus, *CLK2* phosphorylation of *PTPN1* (and/or *PTPN2*), which has been shown to promote its function (*302*), likely directly antagonizes JAK-Stat signaling. This may result in defective antigen presentation machinery and cause tumors with *CLK2* amplification to be intrinsically resistant to immunotherapy.

III.   Discussion

We performed parallel whole-genome screens for regulators of PD-L1 expression and modifiers of IFNγ signaling using a barcoded ORFeome library. We recovered a distinct set of positive PD-L1 regulators from our gain-of-function study, compared with the hits recovered from loss-of-function screening. Burr *et al.* *(262)* and Mezzadra *et al.* *(263)* found that interfering with components of the JAK-STAT signaling pathway, CMTM6, or PD-L1 itself led to the strongest reduction in PD-L1 signal by FACS. While our results included *CMTM4*, *JAK1* and *JAK2* were not in our ORFeome library, and we found that overexpression of interferons themselves, rather than interferon receptors, was more potent in upregulating PD-L1, in the presence or absence of IFNγ. Overall the signal to noise from FACS-sorting for PD-L1[high] cells was favorable, leading

104

to a large list of recovered genes with significant p-values. The overlap and distinctions in these gene lists reflect the pros and cons of loss-of-function versus gain-of-function screening approaches to this type of biological question.

Among the positive PD-L1 regulators discovered by our screen, we found that the RNA-binding proteins *CELF3*, *CELF4*, and *CELF5* are capable of robustly inducing PD-L1 expression when overexpressed. We found that *CELF3* and *CELF5* are expressed in lung cancer cell lines, and that *CELF3* is amplified in 8.9% of lung adenocarcinoma. Its amplification may be driven by its chromosomal linkage to the lung cancer oncogene *ARNT*.

Although the complete recovery of these three family members in both screens and the consistency of barcode performance corresponding to these ORFs provides a fairly large degree of internal validation, we have cloned the ORFs encoding *CELF3*, *CELF4*, and *CELF5* into an inducible lentiviral vector for individual validation. We plan to test the straightforward hypothesis that these proteins are regulating PD-L1 post-transcriptionally, as their primary functional domains are RNA-binding motifs. We plan to test whether they are affecting stability of the PD-L1 transcript, or perhaps promoting alternative splicing. There are two main transcript variants of PD-L1, one of which lacks the exon required for interaction with PD-1, as well as recognition by the antibody used in our screen (*305*).

We found 12 GPCRs that positively regulate PD-L1 cell surface expression in our NSCLC line, including three members of the LPAR family. Several connections

105

between GPCR signaling and PD-L1 imply possible mechanisms for this regulatory influence. For example, KRAS signaling is known to regulate PD-L1 post-transcriptionally, through phosphorylation of the 3'UTR-binding protein TTP (*285*). In addition, cross-talk with PI3K, MAPK, and NFκB signaling can upregulate PD-L1 transcript levels (*101*). Importantly, the H1299 NSCLC cell line possesses an activating mutation in *NRAS*. Because PI3K signaling downstream of LPARs is particularly strong (*283*), we hypothesize that the PI3K signaling pathway may be mediating this GPCR effect, consistent with the observation that interfering with the PI3K pathway leads to loss of PD-L1 expression (*258*). Because PI3K signaling has complex roles in potentiating and responding to interferon signaling (*306*), we plan to test if these GPCRs are inducing interferon expression.

Finally, we recovered a group of genes that block IFNγ-induced expression of PD-L1. These genes, if overexpressed in tumors, would represent potential resistance mechanisms to immunotherapy. For this reason, we are exploring datasets such as the Van Allen *et al.* RNA-seq and survival analysis (*307*), to ask if overexpression of these genes correlate with poor response to immunotherapy. Two genes belonging to the same family achieved a notably high p-value in this screen: *SPDYE2* and *SPYDE4*. We have cloned these ORFs into a lentiviral vector to test the effect of their overexpression on the cell cycle, and to ask if their effect on the cell cycle is directly responsible for abrogating PD-L1 expression.

Potentially the most important finding from our IFNγ Response screen is the observation that *CLK2* overexpression interferes with functional IFNγ signaling. *CLK2* is frequently amplified in breast, prostate, and pancreatic cancer, three tumor types which have responded poorly to immunotherapy in clinical trials, for reasons that are not well understood (*189*). Based on our results, we suggest that the fraction of patients with breast, prostate, or pancreatic cancer that possess *CLK2* amplification (~20-30%) may be intrinsically resistant to immunotherapy. Segregating patients based on *CLK2* status may improve response rates in these immunotherapy-intractable tumor types, and supplementation with *CLK2* inhibitors in tumors with *CLK2* amplification may improve outcomes.

IV. Methods

Cell culture

Human H1299 cells and all clones derived from this parental cell line were cultured in RPMI-1640 Medium (ATCC modification) supplemented with 10% (v/v) fetal bovine serum (FBS), 100 units/mL penicillin, and 100 µg/mL streptomycin. HEK293T cells, utilized for lentivirus production, were maintained in DMEM supplemented with 10% FBS, 100 units/mL penicillin, and 100 µg/mL streptomycin. U2-OS osteosarcoma cells from the American Tissue Type Collection (ATCC), utilized

for titering virus, were cultured in DMEM supplemented with 10% FBS, 100 units/mL

penicillin, and 100 μg/mL streptomycin.

Lentivirus production and titering

To produce lentivirus, HEK293T cells were seeded in tissue culture dishes at a

density equivalent to 6×10^5 cells per 0.9 cm2 surface area. Plasmid DNA was diluted

into serum-free medium with a lentiviral packaging plasmid mixture of SV40 VSVg,

Gag/Pol, Tat, and Rev, and transfected with PolyJet. Cell culture media was changed

24 hours later. After 48 h, the supernatant was harvested, filtered through a low-

protein binding HT Tuffryn® membrane with 0.45 μm pores (Pall cat. #4184), and

stored at -80°C. Lentiviral titer was determined by transducing the cell line of interest

or U2-OS cells plated at clonogenic density with serial dilutions of virus in the presence

of 4-8 μg/mL polybrene. After selecting with puromycin, colonies were counted to

determine viral titer.

H1299 clone design and selection

First, the non-small cell lung carcinoma cell line H1299 was assessed for its

suitability to this screening application. H1299 cells were stained with an APC-

conjugated CD274 monoclonal antibody (eBioscience, clone MIH1) or its APC-

conjugated mouse IgG1 κ isotype control, in the presence or absence of varying doses of

IFNγ, ranging from 50ng/μL-1μg/μL; these stained cells were analyzed by fluorescence-activated cell sorting (FACS). 100ng/μL IFNγ or higher dosing induced maximal expression of PD-L1, so 100ng/μL was selected as the optimal dose for further experimentation. Moreover, H1299 background expression of PD-L1 in the absence of IFN-γ was found to relatively low, even when stained with 5x or 10x the recommended antibody dose. Thus, because of its low background PD-L1 cell surface expression and its highly inducible expression of PD-L1 in response to IFNγ, as well as its therapeutically relevant lung cancer origin, the H1299 cell line was chosen as the basis for screening.

To produce an optimal clonal screening context, parental H1299 cells were transduced in the presence of 8μg/mL polybrene with lentivirus containing an rtTA-neo construct. This construct was previously derived from pInducer20 (*248*) by BstXI digestion, gel purification, self-ligation, and sequence verification. After selection with 500μg/mL neomycin, cells were plated at clonogenic density and 96 clones were individually isolated for further study.

These 96 individual clones were tested for their background cell surface expression of PD-L1, as well as their cell surface expression of PD-L1 induced by treatment with IFN-γ. Among these 96 clones there was some variability in background PD-L1 expression in the absence of IFN-γ, as well as the level of PD-L1 induced in the presence of IFN-γ. 12 clones with the lowest background PD-L1 cell surface expression and highest PD-L1 inducibility in response to IFN-γ were chosen for further selection.

These 12 selected clones were divided into two populations: one for testing, and one for maintenance. The test population was transduced in the presence of 8µg/mL polybrene with lentivirus containing a pBI-TRE construct, which was cloned and generously shared by Qikai Xu. This construct possesses a tet-On bidirectional TRE promoter, cloned into the pHAGE backbone from pBI-TRE (Clontech) and engineered to drive expression of GFP and DsRed. After selection of this construct with puromycin, GFP and DsRed were induced by treatment with 1µg/µL doxycycline, and fluorophore expression was measured in these 12 clones by FACS. The 5 clones with the most highly inducible GFP and DsRed expression upon doxycycline treatment were chosen for the final phase of characterization.

Finally, leakiness from integrated rtTA in the 5 most inducible clones was assessed using the same pBI-TRE-infected test cells. Levels of DsRed and GFP induced with 1µg/µL doxycycline were compared to levels in the absence of doxycycline and normalized to uninfected maintenance populations. Clone 2 exhibited the best induction and lowest background of DsRed and GFP in the test population. Test populations were discarded and the maintenance population of clone 2 was expanded to serve as the basis for screening.

Whole-Genome ORFeome Screen

A whole-genome barcoded open reading frame (ORF) library was generated as previously described (*255*). In brief, a pool of 24-mer barcodes was subcloned into the

pHAGE-CMV-Dest-PGK-puro vector to generate a pHAGE-CMV-Dest-PGK-puro-3'

barcoded library. LR reactions were performed between this library and pDONR221

Orfeome v8.1 pools, which contain 10,298 unique genes. Barcode-ORF pairs were

mapped by Illumina sequence, revealing an average complexity of 5 barcodes per gene.

Although it is unlikely to affect the function of most ORFs, this LR reaction produces a

small 9 amino acid sequence (NPAFLYKVV) at the C-terminus of the protein, before

each stop codon.

This barcoded ORFeome library, containing a total of ~12.8K ORFs, was

packaged into lentivirus in a pooled format, and titer was measured by the colony

formation assay in U2-OS cells described above. Functional titer was confirmed in a

pilot test infection of H1299 cells before proceeding with screen infection; after infection

and puromycin selection at estimated multiplicity of infection (MOI) ranging from 0.1-1,

visual inspection of the percent of cell death was utilized to confirm the correct titer.

A pooled ORFeome screen was then performed by transducing H1299 cells in the

presence of 8μg/mL polybrene with the barcoded ORFeome library, at an MOI of 0.1

and a representation of 1000 cellular integrations per ORF. After selection in 3μg/mL

puromycin for 3 days and expansion for 2 days, the screen was split into two equivalent

populations, each of which maintained the original representation of 1000 cellular

integrations per ORF. One screen population was treated with 1μg/μL doxycycline for

48 hours, while the second population was treated with 1μg/μL doxycycline for 48 hours

and 100ng/μL IFN-γ for the latter 24 hours. After treatment, cells were harvested and

a pre-sort starting pellet was collected from both IFNγ+ and IFNγ- populations before

proceeding with cell staining. The remaining cells were stained with an APC-conjugated

CD274 monoclonal antibody (eBioscience, clone MIH1), subjected to FACS sorting, and

pelleted for analysis. At least 12×10^6 cells were sorted for each replicate, ensuring

maintenance of the entire representation of 1000 cellular integrations per ORF. For

both screens, three populations were collected from sorting: PD-L1^{high}, PD-L1^{med}, and

PD-L1^{low}. In each case, gates were set so that the PD-L1^{high} and PD-L1^{low} populations

received ~5% of stained cells.

Screen cell pellets were lysed overnight at 55°C in 10mM Tris pH8.0, 10mM

EDTA, 0.5% SDS, and 0.5mg/mL proteinase K. Genomic DNA was isolated from cell

pellets by phenol:chloroform extraction using Phaselock tubes (5 PRIME), followed by

chloroform extraction. RNase A was added at a final concentration of 25μg/mL and

incubated for 4 hours at 37°C before a second round of phenol:chloroform and

chloroform extractions. DNA was ethanol precipitated, recovered by centrifugation,

washed three times with 70% ethanol, and resuspended in 10mM Tris-Cl pH 8.5.

ORF barcodes were PCR-amplified from resuspended genomic DNA and adapted

for Illumina sequencing through the addition of a P5 adapter, standard Illumina primer

binding site, and a stagger sequence of variable length (0-8bp) 5' of the ORF barcode,

followed by a 3' Illumina index sequence primer binding site, a 7 base pair index

sequence, and a P7 adaptor. The relative representation of library reagents in each

sample was determined by Illumina sequencing of PCR-amplified ORF barcodes on a NextSeq500 System.

Analysis of Screen Results

Sequencing reads were trimmed to isolate ORF barcodes and aligned to the reference library using Bowtie (*308*). Read counts were analyzed by MAGeCK (*223*) and edgeR (*224*), utilizing camera gene set analysis to calculate gene rank lists, p-values, and FDRs for each gene. The log2-fold-change for each barcode was manually determined by normalizing samples for total read counts and then calculating the log2-fold-change for each barcode between the relevant sorted bin and the pre-sorted starting cells. The average log2-fold-change for each gene was calculated by averaging the fold-changes for all barcodes corresponding to the given ORF.

Pathway analysis was performed on analyzed read counts from both screens, using two methods. First, all genes meeting a cutoff p-value of 0.01 by edgeR were submitted to the Database for Annotation, Visualization, and Integrated Discovery (DAVID) version 6.8 (*282*), after removing all type-I interferons from the submitted gene list. Similarly, all genes meeting a cutoff p-value of 0.01 by edgeR, except the type I interferons, were analyzed through the use of Ingenuity Pathway Analysis (IPA) (Qiagen, Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis).

The frequency of CELF3 amplification in lung adenocarcinoma was calculated by downloading CNA files for all lung adenocarcinoma studies with available CNA data through cBioPortal (*301, 309*). CELF3 expression in CCLE cell lines was determined from CCLE cell line RNA-Seq data, which was downloaded and compiled by Kamila Naxerova. The CELF3 survival analysis was performed with a publicly available dataset including both CNA data and survival information from 183 lung adenocarcinoma patients (*310*).

Chapter 4: A systems approach to mapping the functionality of cancer driver genes

Attributions

The work presented herein was conceived by Steve Elledge and myself. I selected, designed, and optimized conditions for the 30 HMEC screens described, and performed 23/30 of the HMEC screens in their entirety; the results of many of these screens (outlined in Figure 4-13) are pending sequencing and analysis. Eric Wooten performed the tissue culture for the 7 TSG and OG drug screens outlined in Figure 4-13. I generated and titered library lentivirus used for all screens, performed all genomic DNA extractions, PCRs, performed all analysis contained herein, and wrote the following chapter.

Teresa Davoli and Yumei Leng designed the oncogene library, and Yumei Leng performed the site-directed mutagenesis and cloning steps involved in generation of OG library pools. Mei Yuk Choi performed infections, puromycin selection, and cell passaging for OG and TSG screens in p53+/- HCECs and HPNEs, as well as constitutive promoter-driven OG library screens in HMECs and hTERT-IMR90s.

Mamie Li performed cloning steps to generate the shRNA and CRISPR TSG libraries and ORF OG libraries. Qikai Xu and Teresa Davoli designed the shRNA and sgRNA sequences used to generate the TSG libraries. Rupesh Patel performed the hTERT-IMR90 CRISPR TSG screen as the proliferation control arm for his senescence

bypass screen. Timothy Martin generated the clonal p53-/- HCEC and p53-/- HPNE cell lines.

In collaboration with Kornelia Polyak's lab, Andriy Marusyk performed the orthotopic injections for the *in vivo* pilot screen and Doris Tabassum performed the orthoptic injections, tumor measurements, and luciferase imaging for the *in vivo* CRISPR screens.

I.   Introduction

Numerous genes undergo point mutations in the course of tumorigenesis, yet only a fraction of these genes functionally contribute to the tumorigenic process. The most potent drivers of oncogenesis have been discovered and recurrently recovered through tumor sequencing efforts (*311*); these genes have been termed "mountains," reflecting their high frequency of alteration in tumors (*12*). However, driver genes that are less frequently mutated, termed "hills," nonetheless can play an important role in tumor formation, and can be challenging to identify from tumor sequencing data. Several analyses have interpreted point mutations in tumors with the goal of identifying driver genes (*12, 312*), and the Pan-Cancer Atlas recently released a final consensus list of 299 driver genes (*26*).

We developed an approach to this computational challenge called Tumor Suppressor and Oncogene (TUSON) Explorer (*15*). The TUSON Explorer algorithm computes the likelihood of every gene being either a tumor suppressor gene (TSG) or
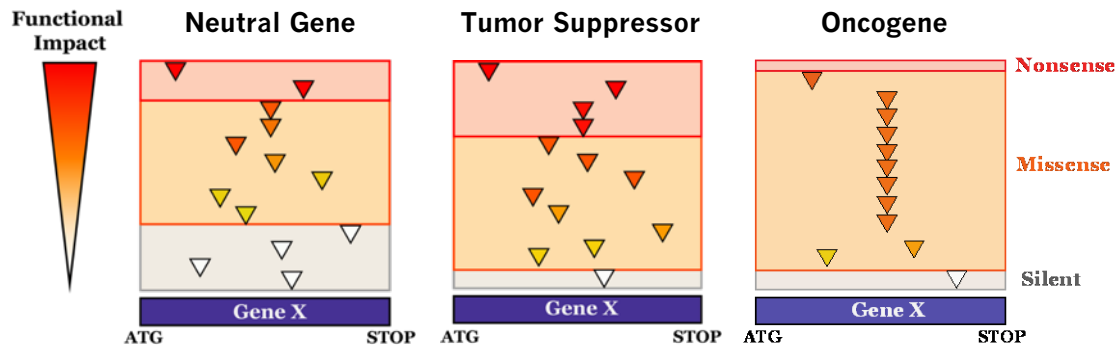
Figure 4-1. Pattern of Mutations Expected by TUSON Explorer for a neutral gene, a TSG, or an OG. Schematic depiction of the mutation patterns interpreted by TUSON Explorer in predicting the likelihood that any given gene is a neutral gene, a TSG, or an OG. Compared to neutral genes, TSGs are expected to display a higher proportion of nonsense and functionally impactful missense point mutations. Furthermore, functionally impactful point mutations in OGs should be clustered along the length of the gene body.

oncogene (OG) based on the pattern of mutations that occur in that gene in tumor sequencing data. The analysis relies upon simple principles: it assumes that the ratio of functionally impactful to functionally benign mutations should be higher for a driver versus a passenger gene (Figure 4-1). Furthermore, it expects that mutations are more likely to be clustered in particular positions along the length of the gene body for OGs rather than TSGs, because there are fewer potential gain-of-function point mutations than loss-of-function mutations in general (Figure 4-1).

TUSON Explorer predicted a larger list of putative driver genes with a statistically significant p-value than most similar studies. We utilized the predictions generated by TUSON Explorer to create small genetic sublibraries targeting the top 500 TSGs and top 350 OGs, to utilize for systematic genetic screens. We created shRNA

and CRISPR-based libraries to the top 500 TSGs, and barcoded ORF libraries to the top OGs. The compact nature of these libraries enabled high-throughput screening.

We sought to systematically explore if TUSON-predicted TSGs or OGs could functionally drive cancer, either by promoting proliferation or other hallmark cancer phenotypes. To that end, we selected and optimized conditions that recapitulate both classic and stress hallmark phenotypes of cancer; we screened both the TSG and OG library in these conditions. In addition, recent work has shown that driver genes can vary significantly by tissue type (*255, 313*). Thus, we screened both the TSG and OG libraries for proliferation in immortalized cell lines of four different tissue types, to uncover driver genes and compare differences in genes that drive proliferation in cell lines derived from four different tissue sources.

II.   Results

Screens of TSG and OG Libraries Identify Proliferation Drivers and Reveal Tissue-Specific Proliferation Pathways

Based upon TUSON Explorer's predictions for the likelihood of genes being either a TSG or OG, we designed two sets of genetic sublibraries: TSG libraries and an OG library. For TSGs, we designed both CRISPR and shRNA-based libraries containing 10 gRNAs or shRNAs targeting the top 500 genes predicted to be TSGs by TUSON Explorer. For OGs, we generated a barcoded ORF library combining 150 genes whose

118

mutational signatures implicate them as potential OGs by TUSON Explorer with 172

genes that are likely to be OGs based on prior screen data (*255, 314*), frequent

amplification in tumors (*21*), or previous annotation as cancer drivers (*22, 315*). To this

pool of genes, we added 150 OGs altered by site-directed mutagenesis to include the

most common point mutation found in tumors. Thus, the OG library contains a

mixture of 300 wild-type and 150 mutant ORFs corresponding to known and predicted

OGs.

We performed a series of systematic screens with these libraries, beginning with

screens for proliferation in four independent immortalized non-cancerous cell lines.

These lines included telomerase (hTERT)-immortalized human mammary epithelial cells

(HMECs), hTERT-immortalized human nestin-expressing pancreatic epithelial cells

(HPNEs), human colonic epithelial cells (HCECs), and human fetal lung fibroblasts

(IMR90s). Importantly, hTERT-HMECs lose their luminal lineage markers in culture

and become more similar to basal-like cells (*316, 317*). In this group of screens, we

intended to ask two related questions: first, did TUSON predict novel TSGs or OGs

that functionally drive proliferation? And second, are there differences in genes that

drive proliferation in one tissue context versus others?

For each of these screens, we transduced the corresponding cell line with either

the OG or TSG library, at a multiplicity of infection of 0.2 and a representation of 1000
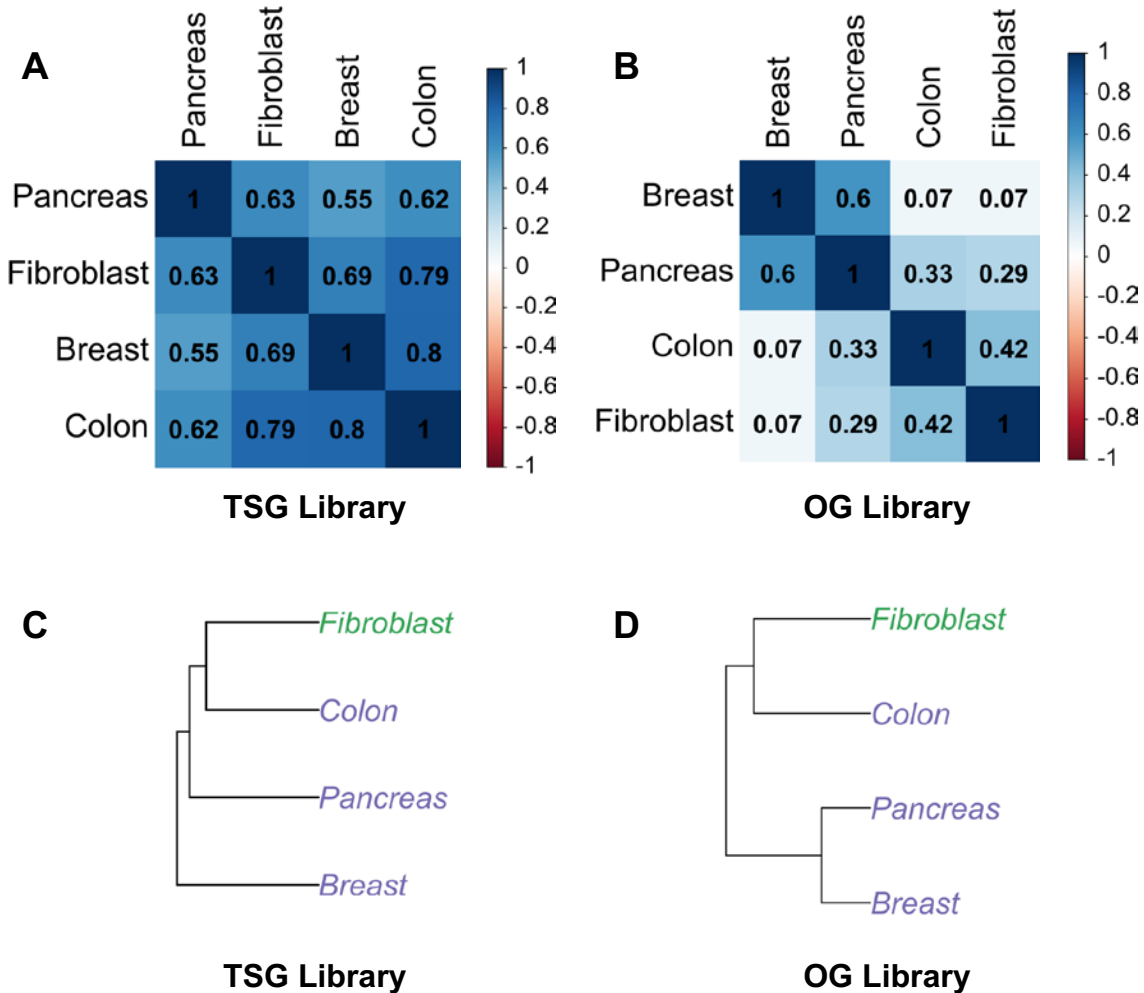
Figure 4-2. Correlation of TSG and OG library performance in proliferation screens of immortalized cell lines. (**A-B**) Correlation between proliferation screens performed in the indicated immortalized cell lines with either the CRISPR TSG library (**A**) or the ORF OG library (**B**) is depicted as a Pearson correlation matrix. The log2-fold-change of each gene was used to calculate Pearson correlation coefficients, which are displayed and color-coded according to the adjacent color scale. (**C-D**) Hierarchical clustering of all statistically significant log2-fold-changes from proliferation screens of the TSG (**C**) or OG (**D**) library in the four indicated cell lines.

cellular integrations per sgRNA or 500 cellular integrations per ORF barcode. After

puromycin selection, we collected initial cell pellets and passaged the cells for 12

population doublings. We isolated genomic DNA from initial and final cell pellets and
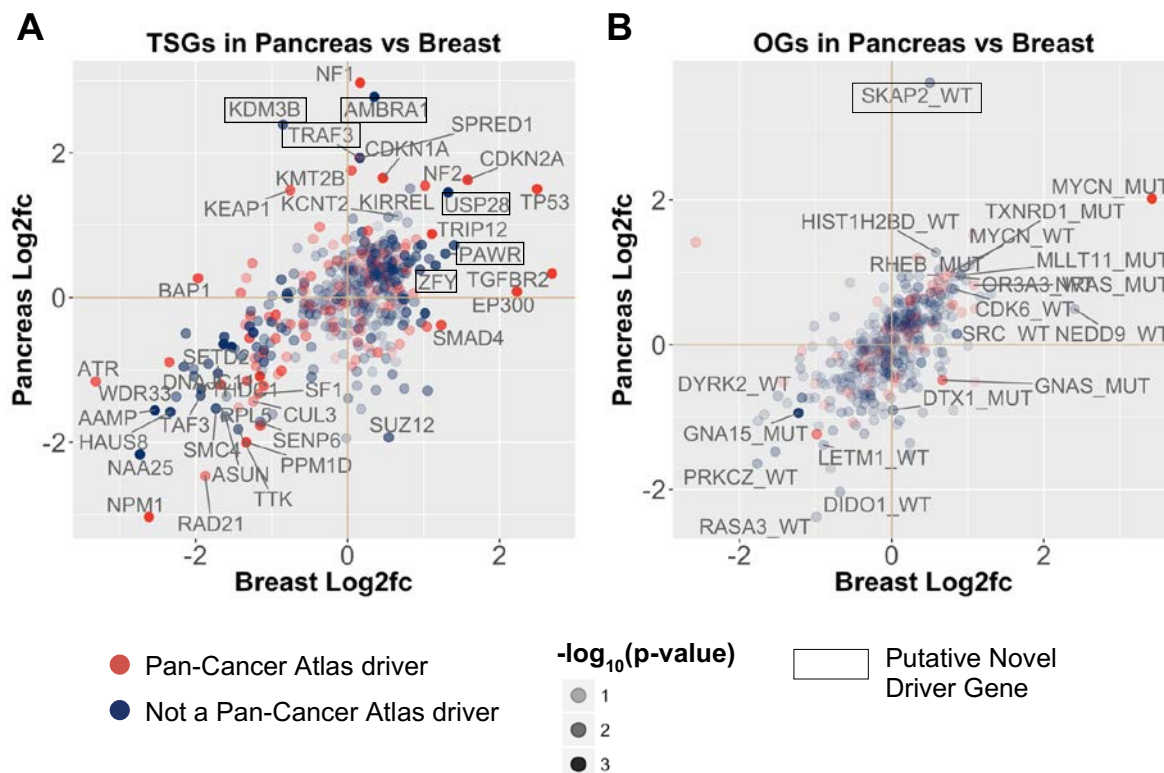
Figure 4-3. Proliferation of TSGs and OGs in Pancreas versus Breast Cells.
Scatterplots depicting the log2-fold-change of each gene in the CRISPR TSG **(A)** or ORF
OG **(B)** library in telomerase-immortalized pancreas versus breast cells. Genes are colored
according to whether they are (red) or are not (blue) annotated as a cancer driver by the
Pan-Cancer Atlas. Each dot represents a single gene, and its opacity is graded according to
the significance of its performance according to MAGeCK (A) or edgeR (B). Putative novel
driver genes are outlined, and genes of interest are labeled.

PCR-amplified gRNAs or ORF barcodes to determine the relative distribution of

reagents at the beginning versus the end of the screen. We calculated this difference as

an average log2-fold-change for each gene and determined the p-value for each gene's

performance by MAGeCK (*223*) for CRISPR and edgeR (*224*) for ORFs.

We began by calculating the correlation between screens performed in different

cell lines, comparing the log2-fold-changes of all genes (Figure 4-2). We found all
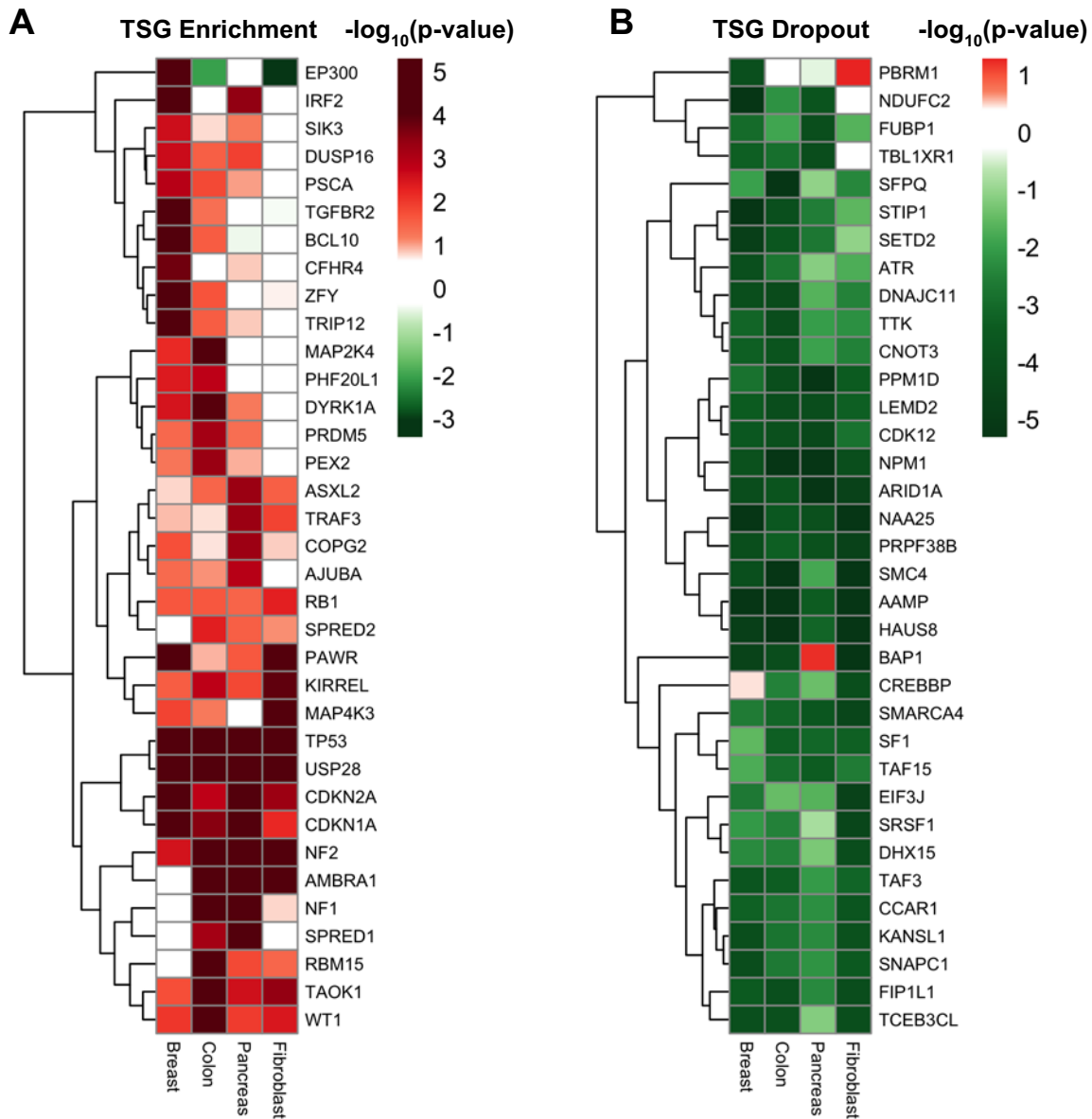
Figure 4-4. TSG Proliferation in Cell Lines from Four Tissue Types. Genes with the most significant p-values by MAGeCK for enrichment **(A)** or dropout **(B)** in proliferation screens across four immortalized cell lines are depicted as heatmaps. The four columns of each heatmap correspond to the indicated cell lines. The p-value of each gene is plotted by color-coding according to the adjacent color legend, on a -$\log_{10}$scale. To clarify directionality, this value is plotted as positive for genes that enriched in the screen (with a positive log2-fold change) and negative for genes that dropped out in the screen (with a negative log2-fold-change. Genes were hierarchically clustered using the complete linkage method.

correlations to be positive, and correlations were notably stronger for the TSG library, possibly because this library contains multiple essential genes, so strong dropout is concordant across all four lines (Figure 4-2A). Correlations were less uniform for the OG library; log2-fold-changes in breast and pancreas cells overall correlated with a Pearson correlation coefficient of 0.6, bearing less resemblance to log2-fold-changes in colon epithelial cells and even less resemblance to log2-fold-changes in fibroblasts (Figure 4-2B). When we hierarchically clustered screen data using all statistically significant log-2-fold-changes, these relationships were apparent from the OG library screens, but they were generally masked in the TSG library screens by the tight correlation of performance of this library across cell lines (Figures 4-2C and 4-2D).

We examined individual gene performance in breast versus pancreas cells, for both the TSG and OG library, in more detail (Figure 4-3). In the TSG library, a number of genes were essential in both immortalized cell lines, and we observed particularly robust, consistent dropout of genes involved in cell division including sister chromatid cohesion (*SMC4 NIPBL*), mitotic spindle assembly (*TTK*, *HAUS8*, *NPM1*), and chromosome segregation (*INO80*, *RAD21*); we noted equally potent dropout of the master regulator of the DNA damage response, *ATR* (Figure 4-3A). In terms of enrichment, CRISPR-based knockout of *TP53* potently drove proliferation in both cell lines, as did knockout of *USP28*, which deubiquitinates p53 to promote its stability (*318, 319*). *TRIP12*, another shared positive regulator of proliferation, is an E3 ubiquitin-protein ligase involved in degradation of p19(ARF), a p53-stabilizing product
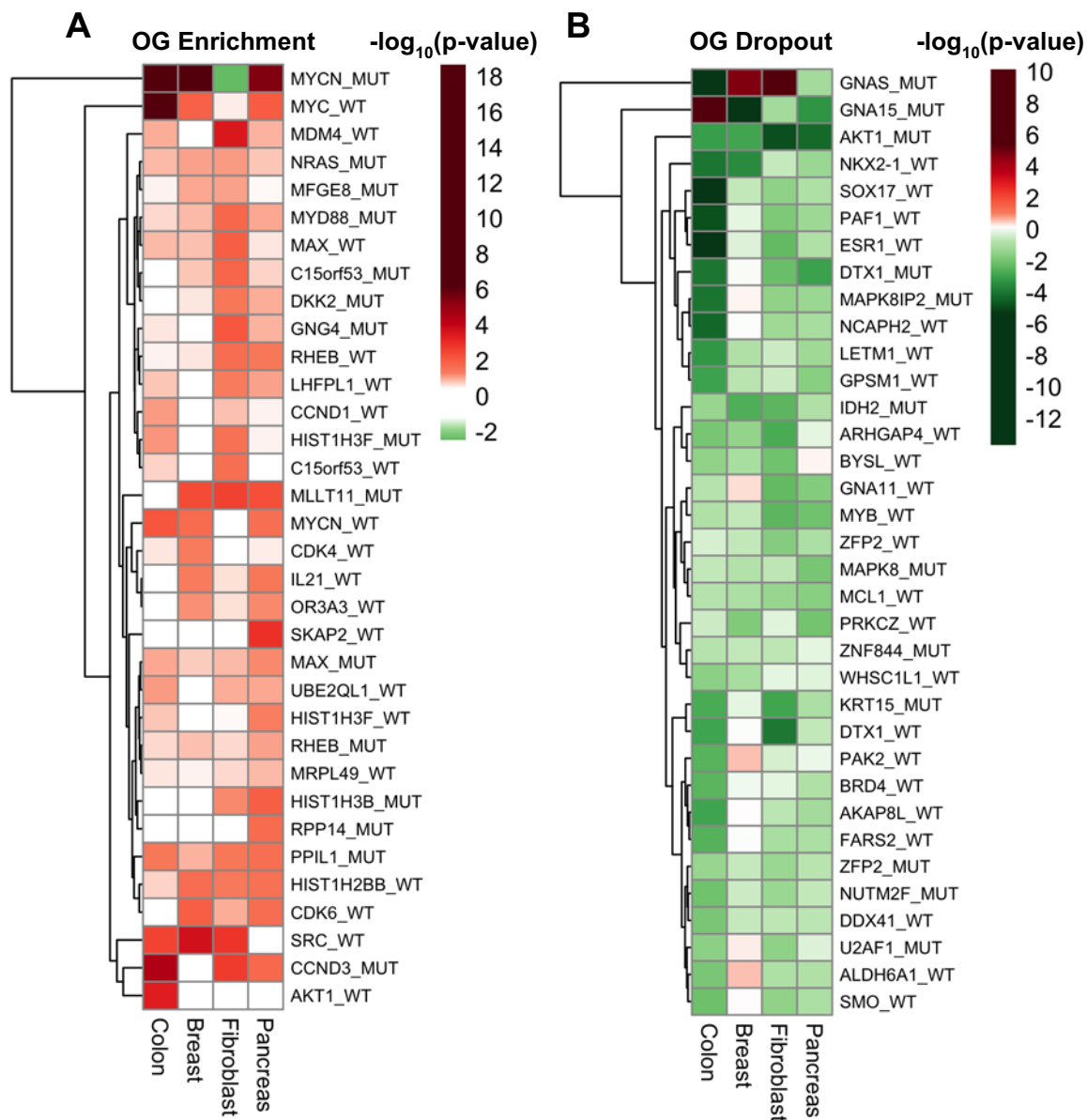
Figure 4-5. OG Proliferation in Cell Lines from Four Tissue Types. Genes with the most significant p-values by edgeR for enrichment **(A)** or dropout **(B)** in proliferation screens across four immortalized cell lines are depicted as heatmaps. The four columns of each heatmap correspond to the indicated cell lines. The p-value of each gene is plotted by color-coding according to the adjacent color legend, on a -$\log_{10}$ scale. To clarify directionality, this value is plotted as positive for genes that enriched in the screen (with a positive log2-fold change) and negative for genes that dropped out in the screen (with a negative log2-fold-change. Genes were hierarchically clustered using the complete linkage method.

of the *CDKN2A* locus (*320, 321*). Thus, the *TP53* pathway is a potent regulator of both breast and pancreas lines.

In contrast, loss of certain genes enhanced proliferation of breast or pancreas cells, but not both. For breast cells, these genes included *TGFBR2* and *SMAD4*, both members of the TGFβ signaling pathway. Although widely distributed *in vivo*, TGFβ exerts diverse tissue-specific growth effects, and has been shown to specifically inhibit mammary growth and morphogenesis (*322*). Breast-specific suppressors of proliferative growth also included *EP300*, *PAWR*, and *ZFY*. Interestingly, although *ZFY* is genomically located on the Y chromosome and these are breast cells derived from reduction mammoplasty tissue of a female patient, 10/10 of the gRNAs targeting this gene also target the highly related paralog *ZFX*, located on the X chromosome. In the opposing cell line, pancreas-specific drivers of proliferation included *AMBRA1*, *NF1*, *SPRED1*, *TRAF3*, and *KDM3B*.

From the OG library, by far the strongest driver of proliferation in both breast and pancreas cells was a mutated version of the *MYCN* gene (Figure 4-3B). The wild-type *MYCN* ORF was modified by site-directed mutagenesis to include the P44L point mutation, a somatic mutation found in neuroblastoma, Wilms tumor (*323*), glioma (*324*), medulloblastoma (*325*), neoplastic cysts of the pancreas (*326*), and other tumor types (*327*). This point mutation is found in 1.7% of neuroblastoma cases without *MYCN* amplification (*328*). Based on its frequent recurrence, the P44L mutation has been presumed to be activating, but it has not been functionally or biochemically

characterized (*329*). Our data shows that this point mutation potently increases the activity of *MYCN* in promoting proliferation, and indeed is a functionally strong activating mutation. The P44L *MYCN* mutant exhibited a 2-fold higher log2-fold-change than wild-type *MYCN* in pancreas cells, and a 4.5-fold higher log2-fold-change in breast cells. Wild-type *MYC* also promoted proliferation of both cell lines, but there is no mutant version of the *MYC* ORF in the OG library.

As with TSGs, several OGs enhanced proliferation of either breast or pancreas cells specifically, but not both. For example, *SKAP2*, an adaptor protein essential in Src pathway signaling, enhanced proliferation of pancreas cells specifically. In general, Src pathway activation accelerates formation of pancreatic tumors in mouse models and correlates with poor prognosis patients (*330*), so there is some contextual basis in concordance with this tissue-specific OG activity. On the other hand, *NEDD9* promotes breast cell proliferation more strongly than pancreas cell proliferation; *NEDD9* is a multi-domain scaffold protein that assembles signaling complexes downstream of a variety of pathways, including integrins and receptor tyrosine kinases.

The recently released Pan-Cancer Atlas published a final consensus list of Cancer Driver Genes, consisting of 299 unique genes (*26*). 25% of our TSG library genes are annotated as Cancer Driver Genes by the Pan-Cancer Atlas, and 26% of our OG library genes are annotated as Cancer Driver Genes. We asked if the genes that functionally drove proliferation in our cell lines are annotated as drivers (Figure 4-3). While many proliferation drivers are annotated as Cancer Driver Genes (red), several are not

annotated on this consensus driver gene list (blue), and we note putative novel driver genes that could be added to this list as including *AMBRA1, TRAF3, KDM3B, PAWR, ZFX,* and *USP28.*

This analysis can be extended to include all four cell lines screened, and most effectively visualized as a heatmap (Figures 4-4 and 4-5). As seen in comparing breast versus pancreas proliferation drivers, loss of some genes drove proliferation across all cell types, including *TP53* and *CDKN2A* (p16) (Figure 4-4, Supplemental Figure S4-1). Among our list of putative novel drivers, *AMBRA1* and *PAWR* impressively drove proliferation in three of the four cell lines tested. *USP28* knockout drove proliferation in all four cell lines, likely through its role in promoting p53 stability. In contrast, certain genes enhanced proliferation in only one particular cell line, behaving as tissue-specific drivers.

By far, the most prominent driver in the OG library is the P44L mutant form of *MYCN*, discussed above (Figure 4-5, Supplemental Figure S4-2). It appears to be an epithelial-specific driver, as it enhances proliferation of breast, pancreas and colon epithelial cell lines, but not fibroblasts. The second strongest driver across all four cell lines is a mutated form of cyclin D, *CCND3*, containing an I290R mutation. This point mutation occurs very near the C-terminus of the cyclin D protein, at a highly conserved isoleucine residue; this isoleucine is not only highly conserved among cyclin D3 homologs in different species, but similar amino acids occur at the corresponding position in cyclin D1 and cyclin D2. A nearby threonine (T283) is an important a site of phosphorylation

Figure 4-6. Effect of p53 Deletion on TSG Proliferation. (A-B) Scatterplots depicting the log2-fold-change of each gene in the CRISPR TSG library in p53 knockout versus p53 wild-type colon **(A)** or pancreas **(B)** cells. Genes are colored according to whether they are (red) or are not (blue) annotated as a cancer driver by the Pan-Cancer Atlas. Each dot represents a single gene, and its opacity is graded according to the significance of its performance according to MAGeCK. **(C)** Genes with the most significant p-values by MAGeCK are depicted as a heatmap. The p-value of each gene is plotted by color-coding according to the adjacent color legend, on a -$\log_{10}$ scale. To clarify directionality, this value is plotted as positive for genes that enriched in the screen (with a positive log2-fold change) and negative for genes that depleted in the screen (with a negative log2-fold-change. Genes were hierarchically clustered using the complete linkage method.

128

that triggers degradation of cyclin D3, and the I293R mutation has been shown to promote stability of the cyclin D3 protein (*331*). The mutant form of *CCND3* far outperforms the only wild-type cyclin included in the OG library (*CCND1*), for all cell lines, underscoring the general theme that in this competitive setting, point mutations can dramatically enhance the ability of OGs to promote proliferation.

We also screened our TSG and OG libraries in two p53 knockout cell lines (p53^{KO}), to investigate the effect of p53 loss on the ability of drivers to promote proliferation. Both colon and pancreas p53 knockout lines were generated by transient transfection of the corresponding immortalized parental cell line (p53^{WT}) with Cas9 and a gRNA targeting p53. We compared TSG and OG library screens in these p53 knockout lines with screens performed in the parental wild-type lines from which they were derived (Figure 4-6, Figure 4-7, Supplemental Figure S4-3). Most TSGs behaved quite similarly in p53^{WT} and p53^{KO} cells, with some notable exceptions. For example, *TP53* drove proliferation in the p53^{WT} cell line but not in the p53^{KO} cell line, as expected. *USP28* and *RBM15* exhibited the same pattern, and this result was true for both the colon and pancreas p53^{WT} and p53^{KO} cell line pairs. As discussed above, *USP28* deubiquitinates and stabilizes p53 (*318, 319*). Although not a prominent member of the p53 pathway, *RBM15* knockout has been shown to reduce Ras-induced p53 protein levels in murine embryonic fibroblasts (MEFs), through a proteasome-dependent mechanism (*332*).
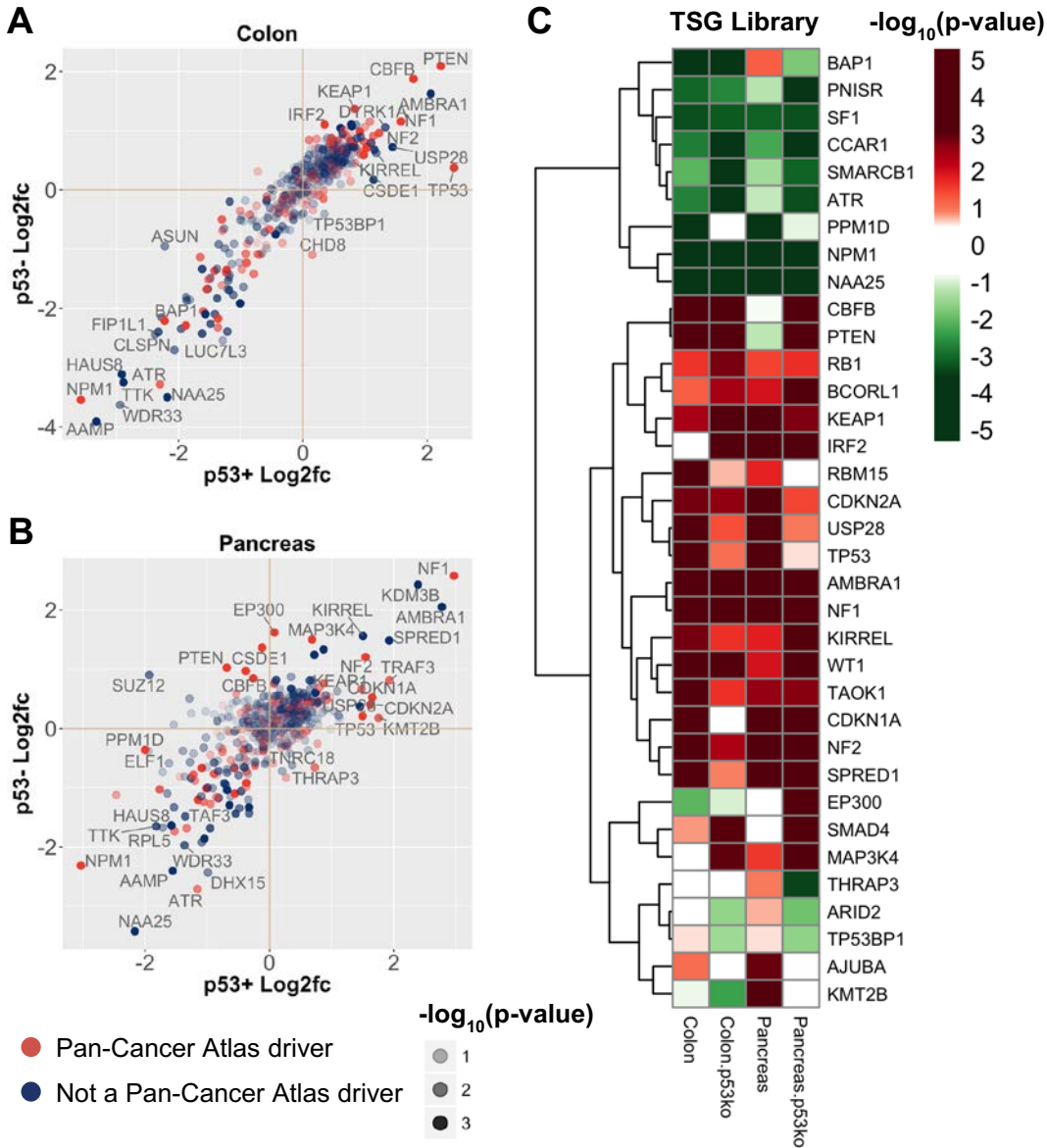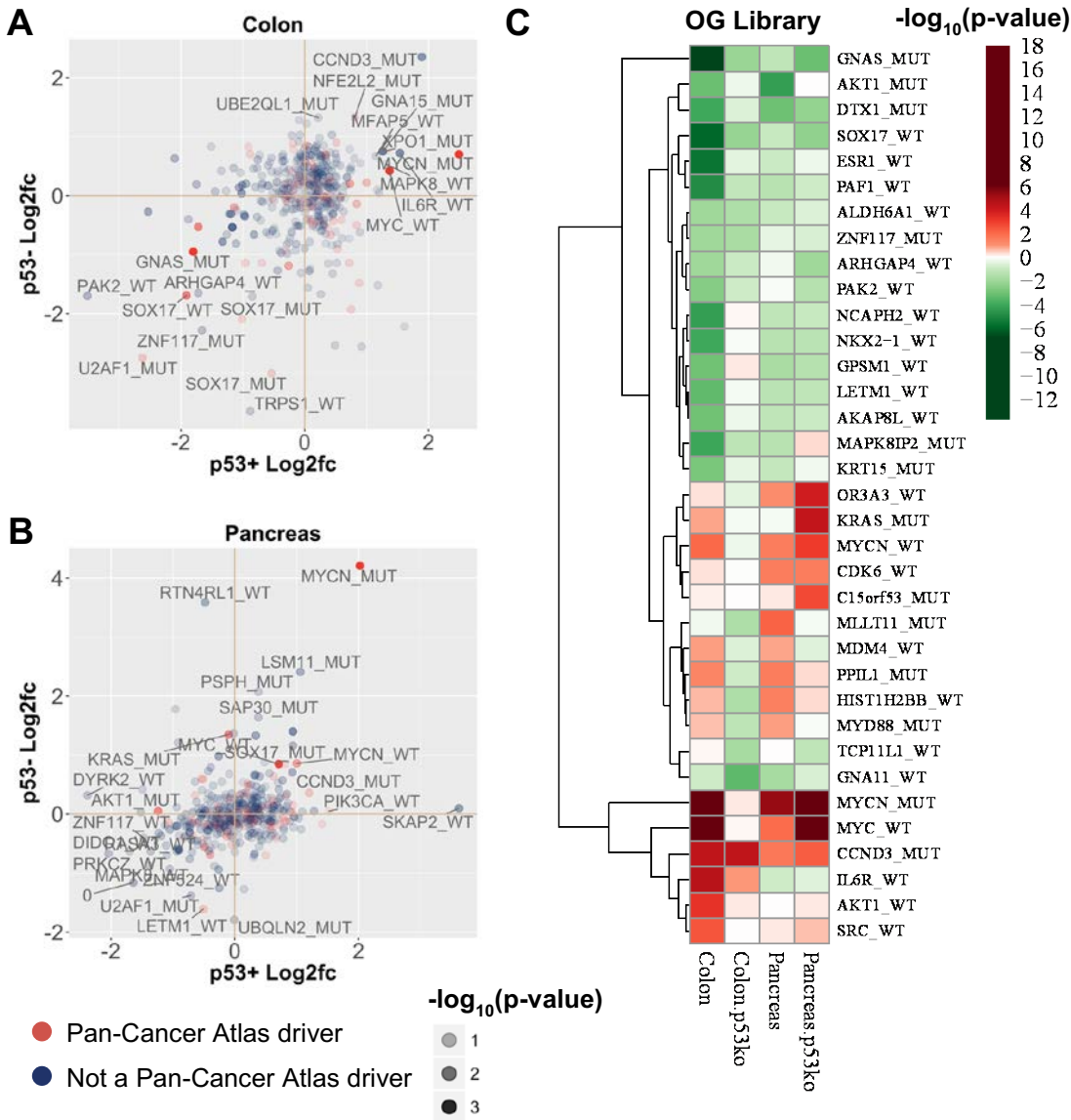
Figure 4-7. Effect of p53 Deletion on OG Proliferation. (A-B) Scatterplots depicting the log2-fold-change of each gene in the ORF OG library in p53 knockout versus p53 wild-type colon **(A)** or pancreas **(B)** cells. Genes are colored according to whether they are (red) or are not (blue) annotated as a cancer driver by the Pan-Cancer Atlas. Each dot represents a single gene, and its opacity is graded according to the significance of its performance according to edgeR. **(C)** Genes with the most significant p-values by edgeR are represented as a heatmap. The p-value of each gene is plotted by color-coding according to the adjacent color legend, on a -log10 scale. To clarify directionality, this value is plotted as positive for genes that enriched in the screen (with a positive log2-fold change) and negative for genes that depleted in the screen (with a negative log2-fold-change. Genes were hierarchically clustered using the complete linkage method

Overall, we found that some genes were tissue-specific proliferation drivers, while other genes drove proliferation in a manner that was context-independent. We noticed certain pathway themes in tissue-specific behavior of drivers; for example, the TGFβ pathway appeared to be breast and colon-specific (Figure 4-4A). Thus, we asked, in an unbiased manner, which pathways were overall most important in driving proliferation of each cell line. To do this, we utilized the recently published curated set of 10 canonical oncogenic signaling pathways released by the Pan-Cancer Atlas (*1*). We classified the genes in our TSG and OG libraries according to these curated pathways and calculated the significance of each pathway in driving proliferation of the four cell lines we screened (Figure 4-8).

This pathway analysis confirmed our incidental observation that the TGFβ pathway drove proliferation of breast and colon cells, but not pancreas or fibroblast cells. The *TP53* pathway and the Cell Cycle pathway (containing *CDKN1A, CCND1, CDKN2A, RB1*, and others) promoted proliferation of all four cell lines. The Myc pathway potently drove proliferation of all three epithelial cell lines and was a less potent driver of fibroblasts. Among the other canonical oncogenic signaling pathways, the Hippo and PI3K pathways were the most potent drivers of proliferation, with the PI3K pathway performing strongest in breast cells. Though altered in many tumor types, the PI3K-Akt pathway is altered in breast tumors with particularly high frequency (*1*), and a recent whole-genome gain-of-function study from our laboratory also found the PI3K pathway to be specifically important in driving proliferation of
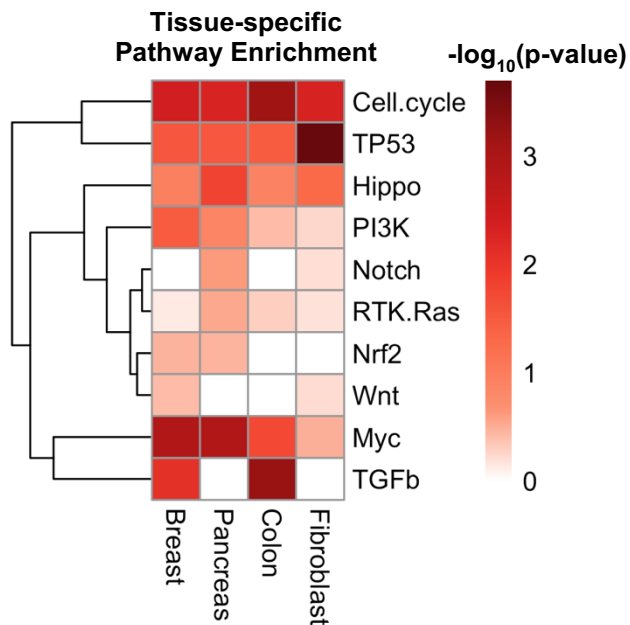
Figure 4-8. Tissue Specificity of Pan-Cancer Atlas Canonical Oncogenic Signaling Pathways. P-values for enrichment of each of the 10 canonical oncogenic signaling pathways curated by the Pan-Cancer Atlas (*1*) are plotted as a heatmap on a color-coded -log$_{10}$ scale. Each column of the heatmap corresponds to a cell line of a different tissue type in which proliferation of TSG and OG library screens were performed. Pathway significance was calculated using a two-tailed Fisher's exact test for the proportion of genes in that pathway that scored in the indicated cell line (requiring a p-value<0.01 to consider a gene as having scored).

breast cells (*255*). In contrast, the fact that the RTK-Ras pathway enhanced

proliferation more modestly may be a matter of relative competition; the log-2-fold-

changes for many of the genes in the RTK Ras pathway were positive, but not

statistically significant in the presence of these other potent drivers of proliferation.

To ask if we could observe these pathway relationships on a broad scale, we

combined all statistically significant OG and TSG proliferation drivers across our four

cell lines and hierarchically clustered them (Supplemental Figure S4-4); this approach

accurately clustered some genes by pathway, and perhaps suggests some unknown

functional relationships between genes, but is not a strictly literal map of pathway relationships.

Mapping Drivers of Hallmark Cancer Phenotypes

Increased proliferation is one of 12 proposed hallmark phenotypes shared by tumors (*55, 56, 333*). We hypothesized that TSGs and OGs recurrently mutated in tumors may also be driving one or more of other the hallmark phenotypes of cancer, in addition to enhanced proliferation. We designed and conducted a systematic series of screens to probe the hallmark phenotypes of cancer with our TSG and OG libraries in HMECs.

We began by screening HMECs for several growth-related phenotypes, with both our CRISPR and shRNA-based TSG libraries. As a baseline, we first compared CRISPR versus shRNA libraries in proliferation screens performed in HMECs, conducted as described above. We passaged cells in culture for 12 population doublings and sequencing PCR-amplified gRNAs or half-hairpins (*219*) from genomic DNA of cell pellets collected at the beginning and the end of the screen. We saw broad overall similarity in gene performance, with some notable discrepancies in gene behavior (Figure 4-9). For example, CRISPR-based knockout of *ZFY* nominally enhanced proliferation, while shRNAs targeting *ZFY* had no effect. As discussed above, *ZFY* is located on the Y chromosome, while its closely related paralog *ZFX,* located on the X chromosome, is also targeted by 10/10 of these gRNAs. In HMECs derived from female patients, it is possible that these gRNAs may have a phenotype through their effect on *ZFX*, but

133
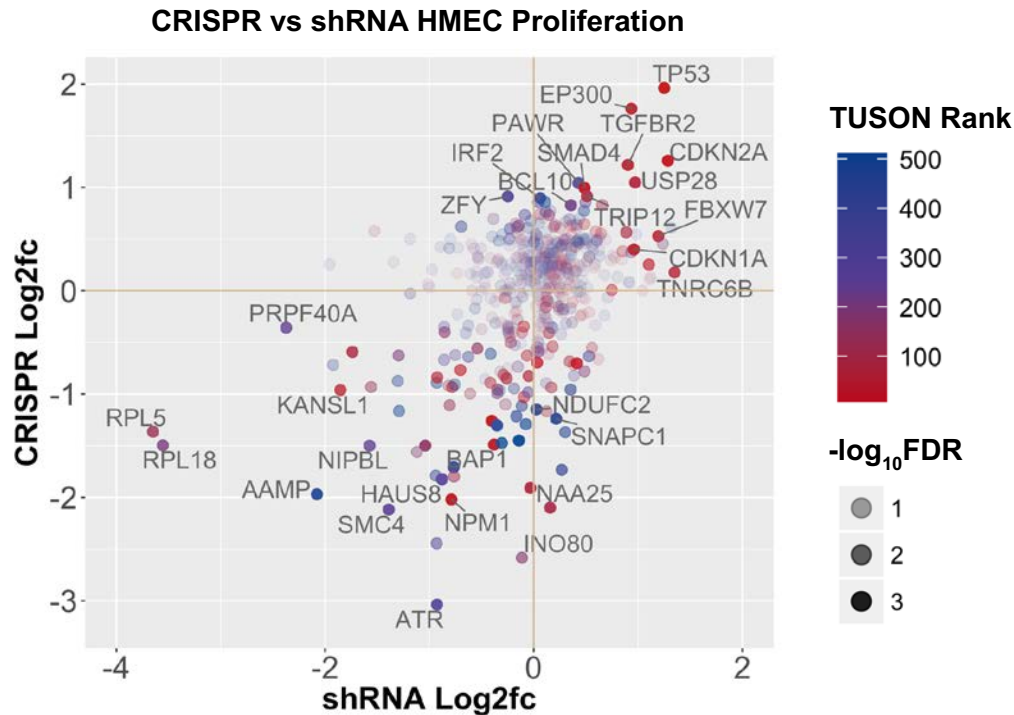
Figure 4-9. HMEC CRISPR versus shRNA TSG Proliferation. Scatterplot comparing the log2-fold-change of each TSG for the CRISPR TSG HMEC proliferation screen versus shRNA TSG HMEC proliferation screen. Each dot represents the average of 10 gRNAs targeting a single gene and is color-coded according to the original rank assigned to that gene by TUSON Explorer. The opacity of each gene corresponds to the significance of that gene's performance as determined by MAGeCK.

shRNA targeting *ZFY* have no activity toward *ZFX*. *ZFX* is a transcriptional activator expressed in multiple tumor types whose knockdown downregulates many genes, including *CDKN2A* and *TGFBR1 (334)*.

In the contrary direction, *TNRC6B* knockdown by shRNA drove proliferation, while CRISPR knockout of *TNRC6B* did not produce the same effect. *TNRC6B* is an essential component of Argonaute-associating microRNA machinery (*335*), and thus its knockout by CRISPR is likely somewhat toxic. In terms of proliferative enhancement, positive drivers of proliferation recognized by both libraries include TGFβ pathway

members *TGFBR2* and *SMAD4* and the transcriptional repressor *PAWR*, a notably

strong enhancer of proliferation (rank=6) relative to its original TUSON Rank

(rank=310). In general, even after mean-normalizing data to facilitate comparison,

CRISPR produced more potent log2-fold-changes than shRNA-based reagents (Figure 4-

9).

In parallel with proliferation, we screened our libraries in HMECs under several

additional growth conditions designed to emulate the stress phenotypes experienced by

cancer cells during growth and invasion: at low clonogenic density, in media with

limiting growth factors, and on ultra-low attachment tissue culture plates (anoikis-

inducing conditions). Particularly because HMECs grow as epithelial islands, growth at

clonogenic density deprives cells of paracrine support signals and increases the

stringency of intrinsic survival signals required for clonal growth (*336*). Growth in the

presence of limiting growth factors increases the demand for self-sufficient proliferative

signaling. Growth on ultra-low attachment plates deprives cells of the simulated

basement membrane provided by serum proteins that attach to plasma-treated tissue

culture plates, inducing a particular form of apoptosis known as anoikis (*337*). These

three culture conditions impose phenotypic demands on the cells that correspond to the

hallmark cancer phenotypes of "evading apoptosis" and "self-sufficiency in growth

signals."

We performed these screens with both CRISPR (Figure 4-10 and Supplemental

Figure S4-7) and shRNA (Supplemental Figures S4-5 and S4-6) TSG libraries and
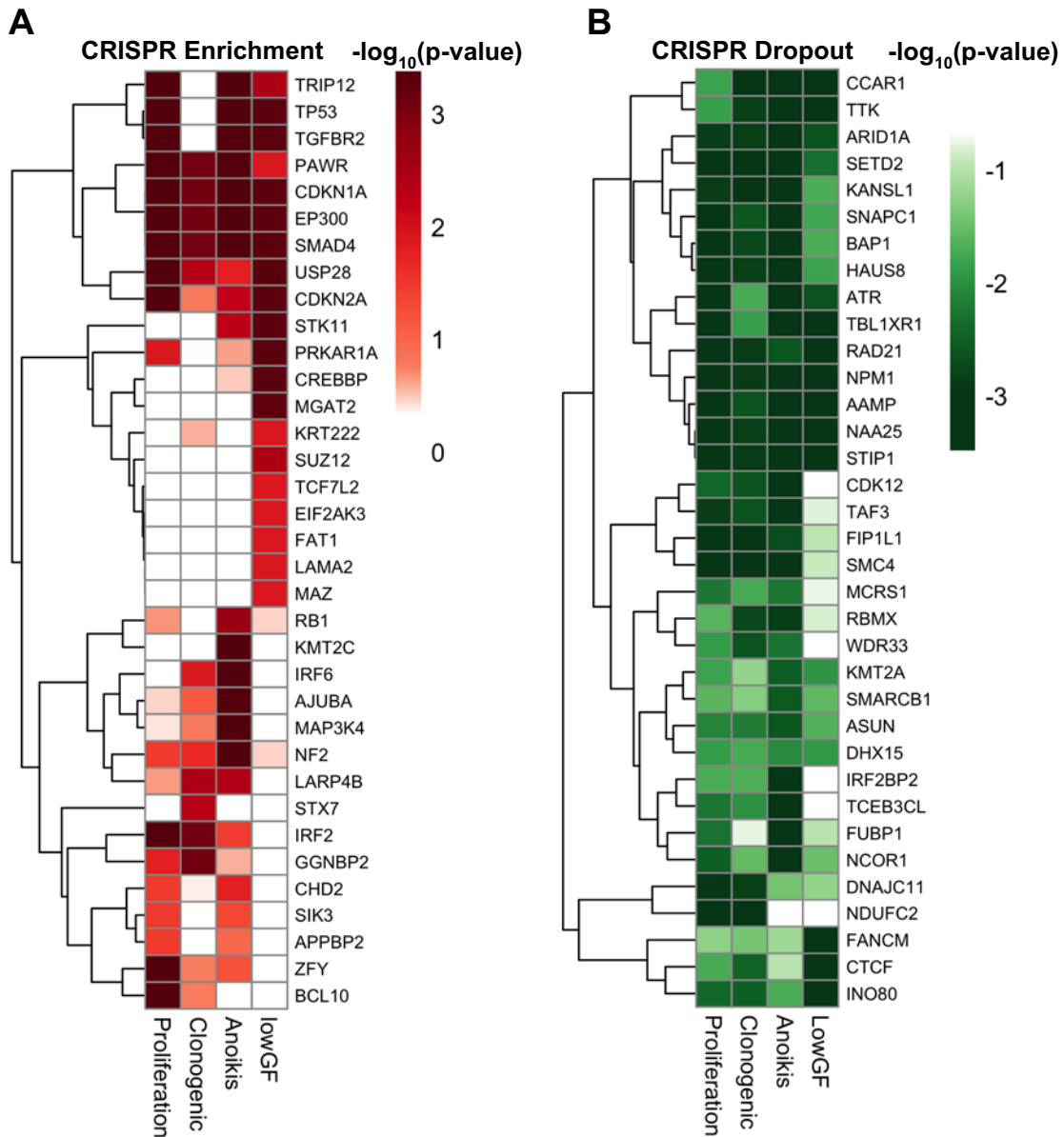
Figure 4-10. CRISPR TSG Library Performance Across Four Different Growth Phenotypes in HMECs. Heatmaps depicting genes with the most significant p-values by MAGeCK for enrichment **(A)** or dropout **(B)** in HMEC CRISPR TSG screens for proliferation, clonogenic growth, anoikis bypass, and growth factor deprivation. The p-value of each gene is plotted by color-coding according to the adjacent color legend, on a -$\log_{10}$scale. To clarify directionality, this value is plotted as positive for genes that enriched in the screen (with a positive log2-fold change) and negative for genes that dropped out in the screen (with a negative log2-fold-change. Genes were hierarchically clustered using the complete linkage method.

compared the results to their corresponding proliferation screens. Some genes promoted growth in all contexts, while others became important under particular stress conditions. For example, *EP300, PAWR, SMAD4, CDKN2A,* and *USP28* promoted survival or growth in every screen. In contrast, knockout of *NF2, MAP3K4*, and *IRF6* was particularly beneficial in allowing cells to bypass exposure to anoikis-inducing culture conditions.
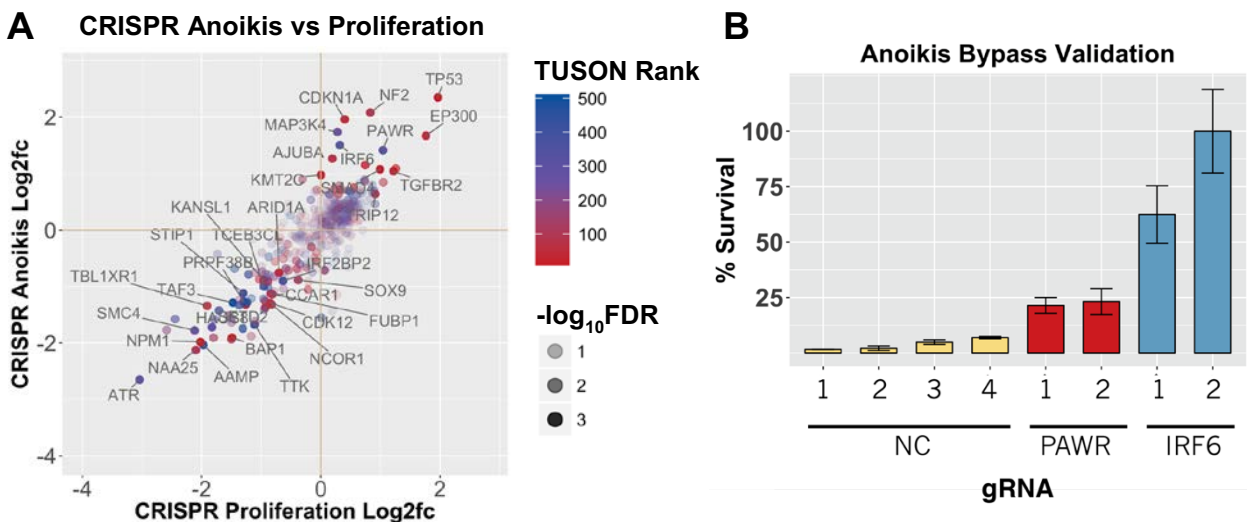
Figure 4-11. *IRF6* depletion allows HMECs to bypass anoikis. (A) Scatterplot depicting the log2-fold-change of each gene in the CRISPR TSG library during the HMEC CRISPR TSG anoikis bypass screen versus the HMEC CRISPR TSG proliferation screen. Each dot represents the average of 10 gRNAs targeting a single gene and is color-coded according to the original rank assigned to that gene by TUSON Explorer. The opacity of each gene corresponds to the significance of that gene's performance according to MAGeCK. **(B)** gRNAs to *IRF6* and *PAWR* were tested individually for their ability to confer survival under anoikis stress conditions, compared with negative control (intergenic-cutting) gRNAs. HMECs were transduced with lentivirus expressing the indicated gRNA under U6 promoter control and Cas9 under EFS promoter control. After selection, cells were subjected to low attachment plating for 24-hours, recovered for 12-hours, and counted with a Coulter counter to measure survival.

We explored these putative anoikis bypass drivers in more detail (Figure 4-11A) and saw that *MAP3K4, IRF6,* and *PAWR* were low-ranking TUSON Explorer genes that had particularly strong effects in bypassing anoikis. While *PAWR* also promoted proliferation, *IRF6* appeared to be an anoikis-specific driver. We cloned individual gRNAs targeting *IRF6* and *PAWR* in parallel with control gRNAs into a lentiviral vector. We packaged this vector, which also contains Cas9 under EFS promoter control, into lentivirus and transduced it into HMECs. We subjected HMECs to anoikis stress conditions and recovered overnight to quantify their survival (Figure 4-11B). We tested two gRNAs targeting *IRF6*, both of which dramatically enhanced
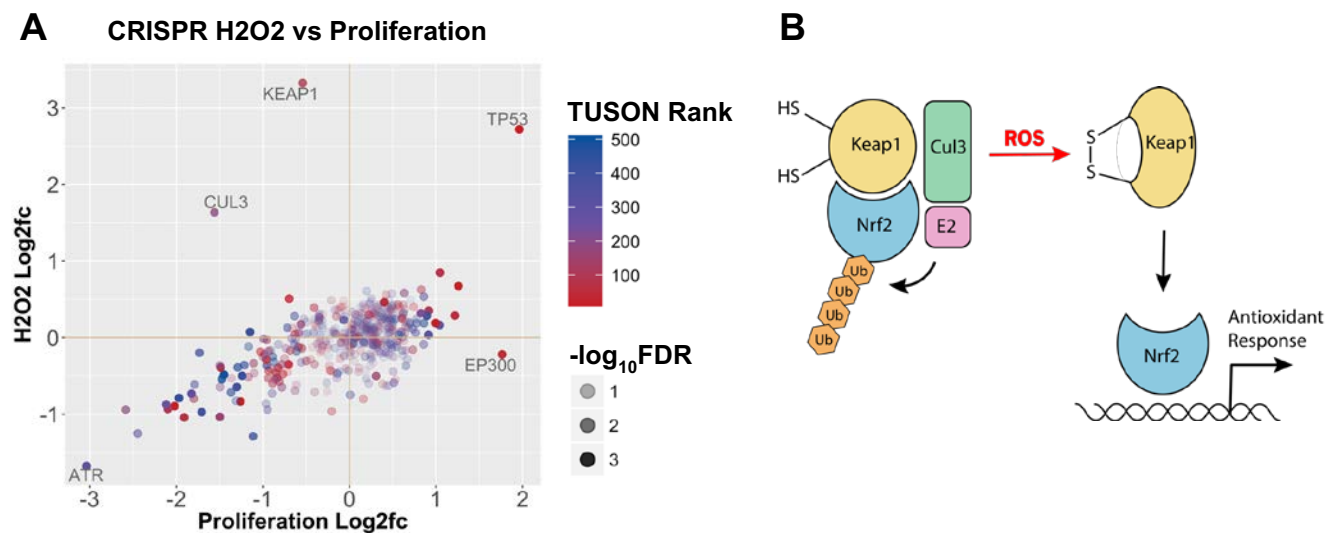
Figure 4-12. H2O2 Bypass Screen Identifies Nrf2-Cul3-Keap1 Pathway. (A)
Scatterplot depicting the log2-fold-change of each gene in the CRISPR TSG library during the HMEC H2O2 bypass screen versus the HMEC proliferation screen. Each dot represents the average of 10 gRNAs targeting a single gene and is color-coded according to the original rank assigned to that gene by TUSON Explorer. The opacity of each gene corresponds to the significance of that gene's performance according to MAGeCK. **(B)** Graphical representation of the Nrf2-Cul3-Keap1 pathway.

the ability of HMECs to bypass this *in vitro* culture stress relative to negative control gRNAs.

In addition to these hallmark growth-related phenotypes, cancers also exhibit several forms of characteristic cellular stress, termed the "stress phenotypes of cancers" (*56, 127*). Though these phenotypes do not initiate the tumorigenic process, they are common features of the tumorigenic state, and they include replication stress, proteotoxic stress, metabolic stress, oxidative stress, and mitotic stress. We hypothesized that genes frequently mutated in tumors may promote tolerance to these forms of stress.

To test this hypothesis in a pilot screen, we exposed cells to reactive oxygen species (ROS) stress, by simple hydrogen peroxide supplementation during screening. We found 3 genes that potently allowed cells to survive in the presence of hydrogen peroxide (Figure 4-12A): *TP53, KEAP1*, and *CUL3*. While *TP53* was also a proliferation driver, neither *KEAP1* nor *CUL3* drove proliferation, but both provided a robust resistance to ROS stress. *KEAP1* and *CUL3* are negative regulators of Nrf2, the master transcription factor of the anti-oxidant stress response. In its reduced state, Keap1 tethers Nrf2 to the E3 ligase Cul3, promoting its ubiquitin-mediated degradation (Figure 4-12B). However, Keap1 possesses cysteine sensor residues that, when oxidized, alter the conformation of Keap1 to release Nrf2 and allow it to translocate to the nucleus, where it upregulates antioxidant response target genes.

Immunoevasion

Screen	Type
PD-L1 (+/- IFNγ)	FACS
MHC-1 (+/- IFNγ)	FACS
pStat (+/- IFNγ)	FACS
CD47	FACS
B7-H3	FACS
B7-H4	FACS
HVEM	FACS
Galectin-9	FACS
CD122	FACS
CD155	FACS

Self-sufficiency In growth signals

Screen	Type
Proliferation	In vitro
Clonogenicity	In vitro

Evading Apoptosis

Screen	Type
Anoikis	In vitro
Low GFs	In vitro

Invasion and Metastasis

Screen	Type
Metastasis	In vivo
**With Polyak Lab	

Stemness

Screen	Type
CD44/CD24	FACS
N-cad/E-cad	FACS

Oxidative and Hypoxic Stress

Screen	Type
H2O2	drug
DMOG	drug
BSO	drug
HIF1α	FACS

Metabolic Stress

Screen	Type
2-DG	drug
GLUT1	FACS

Proteotoxic and ER stress

Screen	Type
Bortezomib	drug
Onalespib	drug

Mitotic Stress

Screen	Type
Taxol	drug

DNA damage stress

Screen	Type
γH2AX	FACS
Entinostat	drug
p53	FACS

Sustained angiogenesis

Screen	Type
VEGF	FACS

Figure 4-13. Graphical Overview of Screens Performed. List of screens performed with the OG and TSG libraries according to screen type and the cancer hallmark phenotype to which the screen corresponds.

Because we were able to elicit this characterized antioxidant response pathway in a pilot screen for ROS stress, we proceeded to screen both the TSG and OG libraries under a broader variety of conditions corresponding to the stress phenotypes of cancer (Figure 4-13). All of these screens have been completed, and the results are pending sequencing and analysis. In parallel, we screened both the TSG and OG libraries by FACS-sorting HMECs for cell-surface and intracellular markers that provide information

about these hallmark phenotypic properties, including molecules upregulated by tumors to evade the immune system (Figure 4-13).

An *in vivo* metastasis screen suggests that *TP53* and *IRF6* promote breast cancer metastasis

Among the 12 hallmark phenotypes of cancer, by far the hallmark feature that causes the greatest morbidity and mortality for patients is the phenomenon of metastasis. Primary tumors shower micro-emboli into the vasculature, beginning fairly early in tumorigenesis, and rarely these circulating cells can seed metastatic clones in a foreign microenvironment. Once multiple or inoperable metastases grow large enough to become symptomatic, therapeutic options for patients are very limited; metastatic growth and its consequences are responsible for about 90% of cancer-related mortality (*338*).

Although the success rate of seeding a metastatic colony from a primary tumor is extremely low, it has been suggested that there are no "metastasis genes" (*12*); in other words, there are no additional rare genetic alterations that endow cells with the capacity to accomplish this exceptional feat. Rather, cancer driver genes altered in the primary tumor may promote metastasis, and success of metastatic seeding may be purely stochastic, increasing in probability as the primary tumor grows in size. To test if any genes in our TSG library can promote metastasis, we performed an *in vivo* screen in collaboration with the laboratory of Kornelia Polyak.

First, we performed a pilot experiment to determine if our library could maintain adequate representation during tumor formation. We chose a breast cancer cell line that does not involute after orthotopic injection into mammary fat pads, with the hope that cells would survive and maintain representation of the full complexity of the library. We transduced MDA-MB-468 cells with a doxycycline-inducible library containing 5000 unique reagents, orthotopically implanted them into mammary fat pads, and analyzed the complexity of the library maintained in tumors after 3.5 weeks of growth in the absence of doxycycline. We found that the complexity of the library detected in the cells we injected orthotopically compared to the tumors we harvested after 3.5 weeks were comparable, regardless of the number of reads per reagent we required as a cutoff for detection (Figure 4-14A).

Under these experimental conditions, we proceeded to test our TSG CRISPR library in MDA-MB-468 cells, in two batches: the top 100 TSGs (originally ranked 1-100 by TUSON) and the bottom 400 TSGs (ranked 101-500 by TUSON). We analyzed primary tumors along with lungs, which are the main site of metastasis for this cell line, at both an early and late time point; the early and late time point experiments were performed in series, not in parallel. At the late time point, the lungs of all mice were riddled with micrometastases, though no macrometastases were visible (Figure 4-14B). Macrometastases were visible in the axillary lymph nodes of two mice carrying the bottom 400 TSG library; these lymph nodes were harvested and analyzed separately.

Figure 4-14. *In vivo* **screen for TSGs that enhance metastasis. (A)** Bar graph depicting the number of shRNAs detected from a doxycycline-inducible library of 5000 shRNAs by sequencing either the pre-injection cells, or the tumor resulting from orthotopic injection of cells into mammary fat pads and growth for 3.5 weeks in the absence of doxycycline. The number of reagents from the library that was detected according to the indicated read count cutoff threshold is plotted for both pre-injection cells and harvested tumors. **(B)** Representative H&E-stained lung sections from mice orthotopically injected with MDA-MB-M468 cells containing the CRISPR TSG library. **(C-D)** Heatmap depicting genes with the most significant p-values by edgeR for enrichment in MDA-MB-468 cells. MDA-MB-468 cells were transduced with the TSG CRISPR library in two pools: either the 100 genes with the best original ranking by TUSON (Top 100 TSGs) or the next 400-ranked genes (Bottom 400 TSGs). Both libraries were screened *in vitro* and orthotopically implanted into mammary fat pads, after which tumors and lungs were examined at both an early and late time point. The p-values for genes with the most statistically significant enrichments are plotted as color-coded values on a -$\log_{10}$ scale.
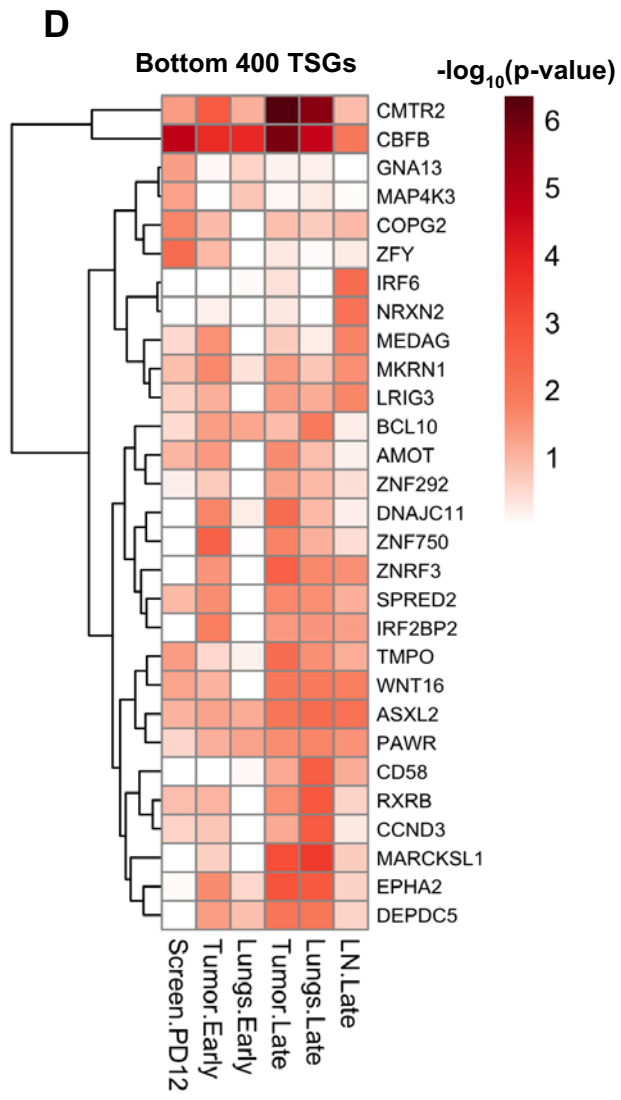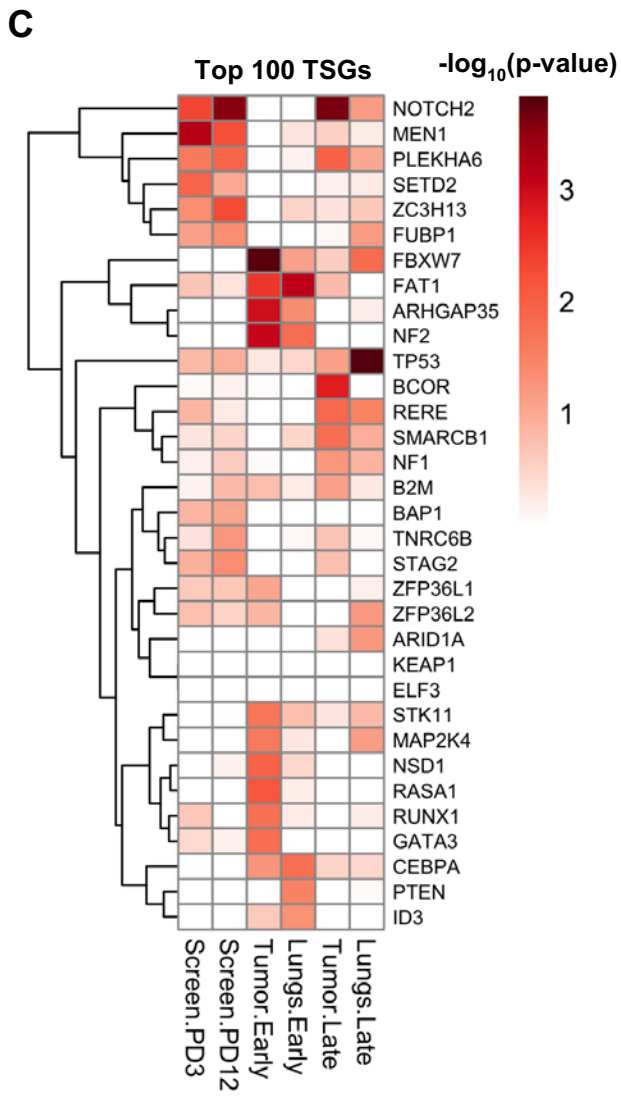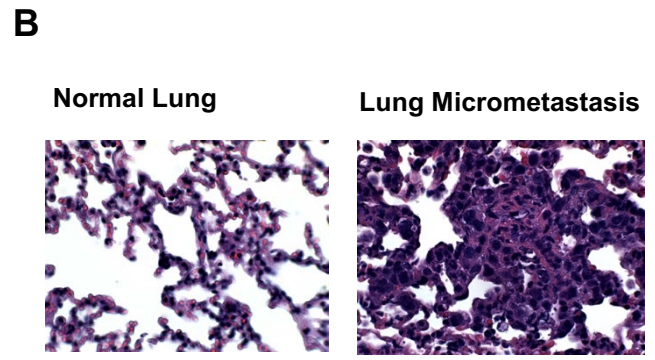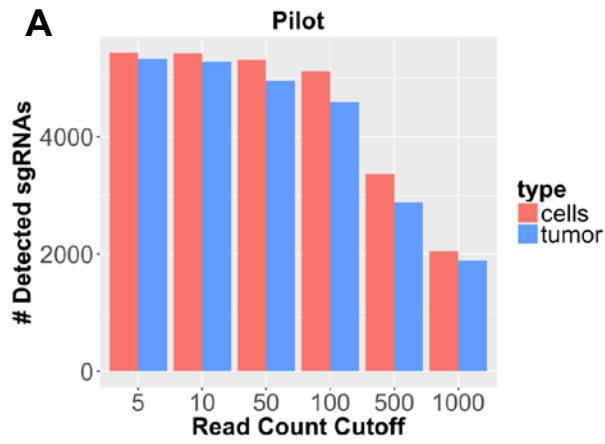
Figure 4-14 (Continued). *In vivo* screen for TSGs that enhance metastasis.

We also performed an *in vitro* screen to compare *in vitro* versus *in vivo* growth for this cell line.

Overall, primary tumor read counts in this experiment were dominated by one or two gRNAs; read counts for all other gRNAs were consequently compressed by this phenomenon. Nonetheless, *CBFB* and *CMTR2* consistently enhanced growth of this cell line both *in vitro* and *in vivo* (Figure 4-14D). *CBFB* is frequently mutated in breast cancer and is a well-established breast cancer driver that binds *RUNX* transcription factors in a heterodimeric fashion (*339*). In terms of metastasis enhancers, the strongest overall driver of metastasis to the lungs was *TP53* (Figure 4-14C) and the strongest driver of metastasis to axillary lymph node macrometastases was *IRF6* (Figure 4-14D). p53 loss, and dominant negative p53 mutations in particular, are known to enhance metastasis in mouse models (*340*). Interestingly, *IRF6* knockout also potentiated bypass of anoikis stress *in vitro*, as discussed above. Survival in the absence of attachment to basement membrane is a prerequisite step in the successful completion of the metastatic cascade, and thus these two phenotypic behaviors are related.

III.   Discussion

We designed genetic sublibraries targeting known and predicted driver genes of two classes: TSGs, as CRISPR and shRNA-based libraries, and OGs, as a barcoded mixture of wild-type and mutant ORFs. We systematically screened these libraries in cell lines derived from different tissue types, and for a variety of hallmark cancer

145

phenotypes, with multiple aims: to uncover novel drivers from the predicted driver gene set, to study differences in genes that drive proliferation in different cell lines, and to ask which driver genes promote hallmark phenotypes of cancer.

25% of our TSG library and 26% of our OG library genes are annotated as Cancer Driver Genes by the Pan-Cancer Atlas (*26*). By screening our libraries in four immortalized cell lines derived from different tissue types, we found 8 genes that strongly drive proliferation but are not annotated as drivers by the Pan-Cancer Atlas: *AMBRA1, TRAF3, KDM3B, PAWR, ZFX, USP28, TAOK1,* and *KIRREL.* This list becomes longer depending on the cutoff set for p-value and log2-fold enrichment. Requiring a p-value<0.01 and a log2-fold-change increase>1, we observed 20 genes driving proliferation that are not annotated by the Pan-Cancer Atlas. Regardless of the cutoff value chosen, our data suggests that functional studies can augment computational approaches in identifying cancer driver genes.

From our OG library screens, we found that the *MYCN* P44L mutation is functionally activating and potently drives proliferation, specifically in epithelial cell lines. This point mutation increases the log2-fold-change compared to wild-type *MYCN* by 3-fold in colon cells, 4.45-fold in breast cells, and 2-fold in pancreas cells. The P44L mutation has been recurrently observed in multiple tumor types but neither functional nor biochemically characterized previously. This point mutation is located in the N-terminal transactivation domain (TAD) of *MYCN,* which interacts with Aurora-A, a kinase that promotes N-Myc stability by reducing the proportion of K48 linkages in its

poly-ubiquitin chains (*341*). We plan investigate how the P44L mutation activates N-Myc and specifically test if this mutation enhances the stability of N-Myc.

The P44L *MYCN* mutation is an epithelial-specific driver of proliferation, and both wild-type *MYCN* and wild-type *MYC* also selectively enhance proliferation of epithelial cells, but not fibroblasts. We observed this pattern incidentally, and independently confirmed it with an unbiased analysis for tissue-specific pathways that promote proliferation. Through this analysis we also saw that the TGFβ pathway exerts a tissue-specific inhibitory effect on proliferation of breast and colon cells. We classified genes according to the 10 canonical signaling pathways of cancer characterized in the Pan-Cancer Atlas by Bailey *et al.* (*1*). These authors also profiled how frequently these 10 pathways are altered across 33 different tumor types. Our functional data is partially concordant with their findings, and we plan to formally analyze the statistical significance of this concordance, as well as investigate possible explanations for points of discrepancy.

Among the genes that drove proliferation in most cell lines, *AMBRA1* enhanced proliferation of every cell line except for HMECs. *AMBRA1* is a negative regulator of c-Myc, promoting its dephosphorylation and degradation by acting as a scaffold with the phosphatase PP2A (*342*). It also promotes autophagy, directly binding Beclin-1 and increasing its interaction with VPS34, thus fostering autophagosome formation (*343*); the latter function may be more important in fibroblasts, which should be relatively insensitive to Myc stabilization. Its failure to promote HMEC proliferation, and similar

147

discrepancies of TSG performance more broadly, may be due to differences in protein expression levels. We plan to compare RNA-seq data with our functional proliferation data, to ask if expression differences can explain some of the tissue-specific proliferation effects we observed in driver performance, particularly with the TSG library.

When we compared shRNA and CRISPR-based TSG libraries in HMEC proliferation screens, we noted overall stronger effects with CRISPR-based reagents. An exception to this trend was the pair of ribosomal proteins *RPL5* and *RPL18*, which produced a more robust dropout phenotype with the shRNA library than the CRISPR library. Prior studies have shown that shRNA-based libraries identify different categories of essential genes more effectively than CRISPR-based libraries (*344*). It may be that cell death following CRISPR-based cutting of ribosomal gene loci occurs so rapidly that it partially occurs during puromycin selection, before the initial cell pellet is collected, thus limiting the dynamic range of the CRISPR screen to detect dropout.

The driving hypothesis for our subsequent series of screens was the notion that, in addition to proliferation, the 12 hallmark cancer phenotypes may be driven by genes that are frequently mutated in tumors. For cancer stress phenotypes, mutated genes may promote tolerance of stress conditions that are endemic to the cancer condition. Thus, we screened 30 independent conditions with both the OG and TSG libraries, probing each of the hallmark phenotypes associated with cancer; the results of many of these screens are pending sequencing and analysis. The size of these small, biologically-

motivated genetic libraries allowed high-throughput parallel screening for multiple phenotypes of interest, which would be infeasible with a whole-genome library.

From the results we have obtained, we found and validated a specific driver of anoikis bypass, *IRF6*, which also promoted metastasis of breast cancer cells to axillary lymph nodes in an *in vivo* screen. *IRF6*, when mutated in humans, can lead to familial syndromes that include cleft lip and palate features (*345*), and mutational variants have been strongly associated with syndromic cleft lip (*346*). Downregulation of *IRF6* has been shown to promote invasion of keratinocytes *in vitro*, which is concordant with our results. We plan to explore the mechanism of *IRF6*'s role in potentiating anoikis bypass by studying the transcriptional consequences of its by RNA-seq.

We have identified a specific antioxidant stress response pathway, the Nrf2-Keap1-Cul3 pathway, by exposing cells to reactive oxygen species stress. The results of the rest of our screens are pending sequencing and analysis, but we hope to elicit similar stress response pathways from our other perturbations. We screened cells for a variety of cell surface and intracellular markers of hallmark cancer phenotypes by FACS, and we plan to analyze the results, with an eye toward identifying drivers involved in multiple hallmark phenotypes.

IV. Methods

Cell culture

Human mammary epithelial cells (HMECs), originally derived from reduction mammoplasty tissue, were purchased from Lonza and immortalized by transduction with a retroviral pBabe-hygro vector expressing human telomerase (hTERT). hTERT-HMECs were maintained in Mammary Epithelial Cell Growth Medium (MEGM, Lonza). hTERT-immortalized human nestin-expressing pancreatic epithelial cells (HPNEs) (*347*) and clonal derivatives were maintained in a 1:4 mixture of M3:BaseF™ culture medium (InCell) and high-glucose Dulbecco's Modified Eagle's Medium (DMEM, Gibco), supplemented with 10% fetal bovine serum (FBS), 100 units/mL penicillin, and 100 μg/mL streptomycin.

Human colonic epithelial cells (HCECs) immortalized with retroviral Cdk4 and hTERT were generously shared by Jerry Shay (*308*). HCECs and clonal derivatives were maintained in a 1:4 mixture of Medium 199 and high-glucose DMEM with the following supplements, at the indicated concentrations: EGF (20ng/mL, Invitrogen), hydrocortisone (1μg/mL, Sigma), insulin (10μg/mL, Sigma), transferrin (2μg/mL, Sigma), sodium selenite (5nM, Sigma), 10% FBS, 100 units/mL penicillin, and 100 μg/mL streptomycin. Human fetal lung fibroblast cells (IMR90) were obtained from the American Tissue Type Collection (ATCC) and immortalized by transduction with a retroviral pBabe-hygro vector expressing hTERT. hTERT-IMR90s were maintained at

3% O₂ in high-glucose DMEM with Glutamax™ supplement (Gibco), and additionally supplemented with 15% FBS, 0.1mM non-essential amino acids, 120 units/mL penicillin, and 120 μg/mL streptomycin. HEK293T cells utilized for virus production were grown in DMEM supplemented with 10% FBS, 100 units/mL penicillin, and 100 μg/mL streptomycin. U2-OS osteosarcoma cells from ATCC, utilized for titering virus, were cultured in DMEM supplemented with 10% FBS, 100 units/mL penicillin, and 100 μg/mL streptomycin.

Clonal p53-knockout cells were generated from parental HCEC and HPNE cell lines by Timothy Martin. To generate these lines, HCEC and HPNE parental cells were co-transfected with a pENTR-U6 gRNA expression construct and pcDNA3.3-TOPO-hCas9 (Addgene Plasmid #41815). sgRNA sequences used to targeted TP53 were: 5'-GCAGTCACAGCACATGACGG-3' and 5'- GAATCAACCCACAGCTGCAC-3'. One week after transfection, cells were treated with 10μM nutlin-3 (Cayman Chemical) and grown until a non-sg*TP53* control plate stopped growing completely. Subsequently, individual clones were isolated from a sparsely plated 15cm tissue culture dish, and loss of p53 protein in isolated clones was verified by Western blotting.

Lentivirus production and titering

To produce lentivirus, HEK293T cells were seeded in tissue culture dishes at a density equivalent to 6×10^5 cells per 0.9 cm^2 of tissue culture surface area. Plasmid

DNA was diluted into serum-free medium with a lentiviral packaging plasmid mixture of SV40 VSVg, Gag/Pol, Tat, and Rev, and transfected with PolyJet (SignaGen). Cell culture media was changed 24 hours later. After 48 h, the supernatant was harvested, filtered through a low-protein-binding HT Tuffryn® membrane with 0.45 µm pores (Pall, cat. #4184), aliquoted, and stored at -80°C. Lentiviral titer was determined by transducing U2-OS cells plated at clonogenic density with serial dilutions of virus in the presence of 4 µg/mL polybrene. After selecting with puromycin, colonies were stained with methylene blue and counted manually to determine viral titer.

TSG Library Construction

DNA oligonucleotides encoding shRNA and sgRNA sequences designed to target the top 500 TUSON-predicted TSGs were synthesized on a custom microarray (Agilent). These oligonucleotides were PCR-amplified separately with specific sets of primers in four batches: the top 100 shRNAs were PCR-amplified separately from the bottom 400 shRNAs, and likewise the sgRNAs were PCR-amplified in two pools.

PCR-amplified gRNA libraries were digested with BbsI and purified on a 10% TBE PAGE gel. Purified, digested fragments were cloned into BsmBI-digested pLentiCRISPRv2 (Addgene Plasmid #52961). A negative control non-cutting gRNA library comprised of 500 gRNAs targeting the *E. coli* genome was designed and cloned in parallel.

PCR-amplified shRNA oligonucleotide libraries were digested with XhoI and

EcoRI, and subsequently purified on a 3% Nusieve gel. Gel purified digestion products

were cloned into a XhoI/EcoRI digested pHAGE-pInducer10-miRE-pheS(ΔEcoRI)

plasmid. To create the pHAGE-pInducer10miRE-pheS(ΔEcoRI) vector, the pInducer10

mir30 shRNA construct (*248*) was moved to the pHAGE backbone, with pertinent

mir30 elements being replaced by miR-E elements by PCR (*348*). In addition, the sole

EcoRI restriction site in the pHAGE backbone was mutated to facilitate cloning. A

negative control shRNA library targeting the *E. coli* genome was designed and cloned in

parallel, in the same fashion.

OG Library Construction

An OG library gene list was compiled by combining the top 150 TUSON-

predicted OGs with a hand-curated list of ~150 putative OGs culled from prior screen

data (*255, 314*), genes found in amplification peaks in tumors (*21*), and annotation as

oncogenes in prior studies (*22, 315*). ORFs corresponding to these genes were

cherrypicked from the ORFeome v8.1 and Ultimate ORF Collection libraries. Genes

obtained from the ORFeome v8.1 collection do not have endogenous STOP codons, so

endogenous STOP codons were introduced by PCR. The 150 TUSON-predicted OGs

were mutated by site-directed mutagenesis to introduce the most frequent point

mutation observed for that gene across all tumor types (QuikChange II XL Site-Directed

Mutagenesis Kit, Agilent). Each point mutant was sequence-verified, arrayed with its

corresponding wild-type ORF, and pooled into a total of 6 pools. Each pool was cloned by LR reaction into a pre-barcoded pHAGE-EFS-DEST-3'BC V2 library vector (LR Clonase II, Thermo Fisher), along with two additional pools of wild-type OG ORFs. The resulting set of 8 pools was then sequenced to link the barcodes with corresponding genes: on average, 5 barcodes were mapped to each mutant or wild-type ORF.

In vitro TSG and OG Library Screens

CRISPR TSG and ORF OG proliferation screens were performed in HMECs, hTERT-IMR90s, HCECs, p53-/- HCECs, HPNEs, and p53-/- HPNEs using the following procedure. Cells were transduced with lentivirus containing either the OG or CRISPR TSG library at a multiplicity of infection (MOI) of 0.2 and representation of 500 cellular integrations per barcode or 1000 cellular integrations per gRNA. All OG library screens were performed in triplicate, and CRISPR screens were performed in triplicate with the exception of HPNE and p53-/- HPNE screens, which were performed in duplicate.

After puromycin selection, initial reference cell pellets were collected. Cells were passaged for 12 population doublings (PDs), and final cell pellets were collected. In many cases, cell pellets were also collected at intermediate time points. Whenever possible, all cell pellets contained a 2-fold excess of cells required to represent the original complexity of the library present at the time of infection.

The CRISPR TSG, shRNA TSG, and ORF OG libraries were also screened in HMECs for several growth phenotypes of interest. For this battery of screens, the same set of infected cells was exposed to several growth stresses in parallel. In all cases, HMECs were transduced with the corresponding library at an MOI of 0.2, in triplicate, at a representation of at least 500 cellular integrations per barcode or 1000 cellular integrations per gRNA or shRNA. Cells were subsequently selected in 1.5-2µg/mL puromycin, and a starting cell population was collected for each replicate, to serve as the initial reference populations for the subsequent screens performed. Unless indicated, a representation of 500 cellular integration events per barcode or 1000 cellular integration events per gRNA or shRNA was maintained throughout the screen.

For the clonogenic screen, HMECs were seeded on 15cm tissue culture plates at a density of 50K cells per plate, maintaining a representation of 200 cellular integration events per ORF barcode, gRNA, or shRNA. Cells were grown to enough confluence to collect a representative pellet, and then serially re-plated at this density two more times. For the anoikis screen, cells were seeded overnight on ultra-low attachment plates (Westnet) at a density of 1×10^6 cells per 10cm plate and recovered on 15cm plates. After sufficient growth to collect a representative pellet, cells were re-plated on ultra-low attachment plates and recovered twice more. For low growth factor conditions, HMECs were cultured in modified MEGM medium containing 5% of the concentration of bovine pituitary extract (BPE), hydrocortisone, EGF, and insulin typically provided with

MEGM medium. For the H_2O_2 stress screen, HMECs were serially subjected to 0.5mM H_2O_2 (Sigma), supplemented in the media every 3-4 days.

For all screens, genomic DNA was isolated from the initial and final cell pellets from each screen by two rounds of phenol:chloroform extraction using Phaselock tubes (5 PRIME), followed by two rounds of chloroform extraction. RNase A was added at a final concentration of 25μg/mL and incubated for at least 4 hours at 37°C before two additional rounds of phenol:chloroform and one additional round of chloroform extraction. DNA was ethanol-precipitated, recovered by centrifugation, washed three times with 70% ethanol, and resuspended in 10mM Tris-Cl pH 8.5.

shRNA half-hairpin (*219*), gRNA, or barcode sequences were PCR-amplified from resuspended genomic DNA and adapted for Illumina sequencing. This adaptation involved the addition of a P5 adapter, a standard Illumina primer binding site, and a stagger sequence of variable length (0-8bp) 5' to the variable sequence of interest (either half-hairpin, gRNA, or barcode), followed by a 3' Illumina index sequence primer binding site, a 7 base pair index sequence, and a P7 adaptor. The relative representation of library reagents in each sample was determined by Illumina sequencing on a NextSeq500 or HiSeq2000 System.

In vivo TSG Library Screens

First, a pilot screen was performed to determine if representation of the TSG library could be maintained during tumor formation. Breast adenocarcinoma MDA-

MB-468 cells were transduced in triplicate in the presence of 8μg/mL polybrene with pooled shRNA lentiviral libraries at an MOI of 0.162 and a representation of 1000 cellular integrations per shRNA. Negative control shRNA library lentivirus was mixed with TSG shRNA library virus prior to transduction. After 48 hours, cells were split into 1μg/mL puro for selection. Cells were expanded in culture for several weeks in the absence of doxycycline prior to injection.

Tumors were induced by bilateral orthotopic injection into 8 x 6-week-old female NOD-*scid*-gamma mice (NSG, Jackson Labs) of $3x10^6$ MDA-MB-468 cells resuspended in 300μL of 50% Matrigel (BD Biosciences). Mice were not administered doxycycline at any point during this pilot experiment. After 3 weeks, small tumors (~4x5mm) were harvested, manually diced with a razor blade, and digested overnight at 55°C in 10mM Tris pH 8.0, 10mM EDTA, 0.5% SDS, and 0.5mg/mL Proteinase K. Genomic DNA was isolated and PCRs were performed as described above for *in vitro* screens.

Next, two rounds of *in vivo* CRISPR screens were performed, terminating at either an early (4.5 weeks) or late (12 weeks) timepoint. For both rounds of screening, performed in series (not in parallel), MDA-MB-468 cells were transduced in triplicate with lentiviral libraries in the presence of 8 μg/mL polybrene at an MOI of 0.2 and a representation of 1000 cellular integrations per gRNA. The top 100 TSG CRISPR library and bottom 400 TSG CRISPR library were transduced separately and never mixed, though negative control sgRNAs were spiked into each lentiviral pool. After 48 hours, cells were split into 1ug/mL puromycin for selection. After selection, all library-

infected cells were transduced with an mCherry/Luciferase-expressing lentiviral construct (obtained from K. Polyak laboratory via C. Mitsiades laboratory, DFCI) at an MOI of 2.

For both rounds of CRISPR screens, 9 x 6-week-old NSG mice were handled in two groups: 3 mice received the top 100 TSG CRISPR library, and 6 mice received the bottom 400 TSG CRISPR library. In each case, tumors were induced by unilateral orthotopic injection of $2x10^6$ MDA-MB-468 cells resuspended in 200μL of 50% Matrigel (BD Biosciences). Tumor size was measured and luciferase signal was imaged once per week. After either 4.5 weeks (early timepoint) or 12 weeks (late timepoint), tumors were harvested. For the 12-week time point, additional organs were harvested to assess possible metastases: lungs, bone marrow from the femur, brain, liver, spleen, and two axillary lymph nodes that harbored visible macrometastases. For 12-week time point samples, small sections of tumors and lungs were fixed, embedded in parafilm, and stained with hematoxylin and eosin (H&E). All tumors were manually diced with a razor blade and digested overnight at 55°C in 10mM Tris pH 8.0, 10mM EDTA, 0.5% SDS, and 0.5mg/mL Proteinase K. Organs were digested in the same solution overnight at 55°C with the addition of type 2 collagenase (Worthington). Subsequently, genomic DNA was isolated and PCRs were performed as described above for *in vitro* screens.

Analysis of Screen Results

Sequencing reads were trimmed to isolate half-hairpin, gRNA, or ORF barcode sequences as appropriate. Trimmed reads were aligned to the corresponding reference library using Bowtie (*308*). CRISPR and shRNA read counts were analyzed by MAGeCK (*223*) to calculate gene rank lists, p-values, and FDRs for each gene. ORF read counts were analyzed by edgeR (*224*) utilizing camera gene set enrichment analysis to calculate gene rank lists, p-values, and FDRs for each gene.

The log2-fold-change for each TSG and OG in each screen was calculated using the following process. First, all samples were normalized for total read counts. Next, barcodes, gRNAs or shRNAs with low starting read counts were eliminated (typically, reagents with the lowest 10% starting read counts were discarded). The fold-change of each barcode, gRNA, or shRNA was then calculated on a log2 scale, averaged for each gene, and mean-normalized to facilitate comparison across screens. When plotted as a heatmap, genes were hierarchically clustered using the complete linkage method for hierarchical clustering.

Pathway analysis was performed by downloading the canonical pathway annotations from the Pan-Cancer Atlas (*1*). These annotations were used to classify all genes in the TSG and OG libraries into the 10 canonical signaling pathways. Gene performance in each cell line was scored as a hit if the gene performed with a p-value cutoff<0.01 by MAGeCK (*223*) for CRISPR or edgeR (*224*) for OGs. The significance of the pathway was calculated with a two-tailed Fisher's exact test for the proportion of

genes in the pathway that were hits, compared to the proportion of genes that were hits but not in the pathway. This calculation was repeated for all pathways and all cell lines.

Chapter 5: Conclusions and Perspective

Advances in Genetic Screening Technology

Genetic screens provide an unbiased method for addressing biological questions of interest. Although a variety of model systems can serve as the contextual basis for genetic screens, cultured mammalian cells offer several advantages as a model system, particularly in posing questions pertinent to the field of cancer biology. For example, the alternative eukaryotic genetic model *S. cerevisiae* lacks orthologs of many human genes and cellular processes comparable to senescence, oncogenic transformation, and tissue-specific differentiation (*349, 350*).

Unbiased genetic screening involves systematically perturbing gene function, recognizing or selecting for a phenotype of interest, and characterizing the genetic perturbations that cause an effect on that phenotype. This process relies on technologies that induce gene perturbations, which represent a broad toolkit including transposon-mediated insertional mutagenesis, RNA-interference (RNAi) delivered through antisense oligonucleotides (siRNA), lentiviral constructs expressing short hairpin RNA (shRNA), chemical inhibition with small molecules, and overexpression of open reading frames (ORFs). To conduct these screens in a pooled format, these perturbations must be identifiable from the bulk population, for example through PCR amplification of genomically-integrated gene-perturbing reagents or barcodes linked to these reagents. Pooled genetic screens are more efficient and cost-effective than arrayed

libraries, and protocols for pooled genetic screens have been streamlined and optimized (*351-354*).

The watershed discovery that the CRISPR (clustered regularly interspaced short palindromic repeat)-Cas9 bacterial immune system can be harnessed to edit the mammalian genome (*355, 356*) significantly expanded this toolkit of genetic perturbation technologies. The Cas9 endonuclease can be directed to create a double-strand break (DSB) in DNA at a specific genomic locus of interest by a small guide RNA (gRNA), resulting predominantly in repair by non-homologous end-joining (NHEJ) which generates small insertions and deletions (indels) at the induced DSB breakpoint. Particularly if Cas9 introduces this DSB at a region of the gene that is important for gene function, the resulting indels impair function of the gene product (*357, 358*).

This technological advance added a highly penetrant, specific (*359*) gene-targeting tool to the arsenal of machinery that can be used to create perturbations in genetic screens. Rollout of this easily-applied, effective, and inexpensive technology occurred rapidly (*360*). Shortly thereafter, multiple efforts resulted in variants of the original CRISPR-Cas9 system that can also be applied to genetic screening. For example, CRISPRi, a related loss-of-function screening technology, involves fusion of an enzymatically dead Cas9 (dCas9) to a Krüppel-associated box (KRAB) transcriptional repression domain; this dCas9-KRAB complex does not cleave its target gene, but rather transcriptionally represses it by localizing KRAB to the target gene's transcriptional start site (TSS) with a gRNA (*361, 362*). Alternatively, fusion of dCas9

162

to transcriptional activators such as VP64 or the p65 activation domain (p65AD) allows adaptation of the CRISPR-Cas9 system to gain-of-function screening (*361, 363*).

Comparison of Tools in the Genetic Screening Toolkit

Evers *et al.* performed a direct comparison of three of the aforementioned loss-of-function screening technologies to interrogate off-target effects, signal-to-noise, and consistency of performance (*364*). Compared to shRNA and CRISPRi, CRISPR exhibited better signal-to-noise (measured by statistical significance determined by MAGeCK) and consistency of performance between cell lines. On the other hand, shRNA showed significantly more off-target effects than CRISPRi and classical CRISPR; off-targets were estimated by testing a set of *a priori* known essential and non-essential genes.

We also performed comparisons between shRNA and CRISPR libraries, in our *BRCA2* synthetic lethal screens and in our HMEC TSG proliferation screens. Like Evers *et al.*, we found better signal-to-noise (statistical significance as determined by MAGeCK) from screens with CRISPR libraries versus shRNA libraries. Likewise, the magnitude of log2-fold-change effects we observed from the CRISPR library was greater in both cases, and we observed better linear correlations between log2-fold-changes of gene performances between replicates (*data not shown*).

However, there were certain groups of essential genes from which only shRNA libraries could elicit a phenotype. For example, in our *BRCA2* synthetic lethal screens,

splicing factors such as *SF1* and *SF3B2* scored with our shRNA but not our CRISPR DNA damage sublibrary. In our HMEC proliferation screen, the microRNA machinery component *TNRC6B* enhanced proliferation only when depleted by shRNA, not by CRISPR. We also saw stronger depletion of *RPL5* and *RPL18* in the HMEC TSG proliferation screen with shRNA compared with CRISPR-based depletion.

Morgens *et al.* previously compared the ability of CRISPR versus shRNA to detect essential genes (*344*). They found that shRNA outperforms CRISPR at detecting certain classes of essential genes including proteasome components, the spliceosome, and nuclear pore proteins, while CRISPR outperformed shRNA at detection of metabolic genes, mediator, and RNA polymerase. Our results agree with their observations, in that ribosomal and spliceosome genes are detectable with shRNA only.

Several groups have found that gene copy number affects the results of genome-scale CRISPR-Cas9 proliferation screens (*358, 365*). Increased genomic copy number results in increased generation of DSBs, inducing gene-independent cell-cycle arrest and false positives in proliferation screen dropouts. In our BRCA2 synthetic lethality study, we performed CRISPR screens in both the pseudodiploid DLD-1 cell line and the aneuploid PEO1 cell line. PEO1 cells have a modal number of 41, but a large number of chromosome arm-level copy number alterations (*data not shown*). Perhaps unrelated to copy number, the DLD-1 cell line was a better application for our CRISPR library, yielding stronger log2-fold-changes for both dropout and enrichment.

In addition to these loss-of-function technologies, gain-of-function genetic libraries can offer a complementary modality to enable unbiased genetic screening approaches. An alternative to CRISPRa, discussed above, is ORF overexpression. Our laboratory recently published a genome-scale barcoded human ORFeome library, in a Gateway-compatible doxycycline-inducible lentiviral expression system (*255*). Barcoding ORFs obviates the substantial issue of PCR bias that otherwise introduces significant noise into data from ORFeome screens (*366*).

We applied this genome-scale barcoded ORFeome library to conduct a screen that had previously been performed with loss-of-function reagents (*262, 367*). In screening for genes that positively regulate PD-L1, our gain-of-function screen recovered different hits compared with either a CRISPR-based screen (*262*) or an insertional mutagenesis screen in haploid cells (*367*), although we also recovered *CMTM4*, a family member of the top shared hit from these two screens. We chose the inverse approach of selecting for up-regulation of PD-L1 cell surface expression with a gain-of-function library, rather than selecting for down-regulation of PD-L1 cell surface expression with a loss-of-function library. The fact that we uncovered a different gene set highlights the complementarity of screening with loss-of-function as well as gain-of-function reagents. Our recovery of *CLK2* from this screen underscores the utility of gain-of-function reagents in addressing cancer-relevant screening questions: genes identified by these reagents can be amplified and expressed in tumors.

Towards a Comprehensive List of Cancer Drivers

A wealth of tumor sequencing data has accumulated over the past decade and has provided incredibly fine resolution topography of the cancer genome landscape. Mutation frequency of driver genes been described as hills and mountains, with mountains representing the most frequently mutated genes. As the number of publicly available tumor sequencing datasets from TCGA has grown to include over 11K tumors and 33 tumor types, it has been come clear how staggering these mountains truly are; sequencing efforts repeatedly recover alterations in the same highly mutated driver genes across cancer.

But with a finer resolution map, one can study the hills: the genes whose mutation frequency is not as prominent, but which nonetheless play an important functional role in tumorigenesis. In asking how many genes can potentially drive cancer overall, or functionally contribute to tumorigenesis within a single tumor, one must face the issue of a cutoff value. We chose to screen the top 500 genes predicted by TUSON, which met an overall p-value<0.01 and q-value<0.38 for their likelihood of being a TSG based on tumor sequence data. However, a stringent threshold may have overlooked one of our strongest drivers of proliferation in multiple cell lines, *PAWR*, which had a TUSON p-value=0.003 and TUSON q-value=0.183 (rank=310).

In screens of TSG and OG libraries across multiple cell lines and hallmark cancer phenotypes, we recovered at least 20 driver genes not commonly annotated as drivers by public databases, including the final consensus list recently released by the Pan-Cancer

Atlas (*26*). Recovery of these genes validates TUSON's ability to predict driver genes and shows that functional studies can augment computational approaches in identifying hills across the cancer landscape. Furthermore, functional studies of hallmark cancer phenotypes in addition to simple cellular proliferation *in vitro* can unveil additional functional driver genes. For example, we elicited the Nrf2-Keap1-Cul3 pathway only by exposing cells to ROS stress, and we observed the functional importance of *IRF6* through an anoikis bypass screen.

Assembling a comprehensive list of driver genes may be an attainable goal, though our data suggests that functional characterization can augment computational approaches in aggregating this driver gene census. From our TSG and OG screens, we can add some genes to the list of commonly-annotated drivers, including *AMBRA1, TRAF3, KDM3B, PAWR, ZFX, USP28*, *TAOK1*, *KIRREL*, and *SKAP2.* While some of these have been reported as drivers in individual reports, with likely mechanisms for driving proliferation being known, others are less well characterized.

After identifying the hills and even the smaller mounds that define the cancer landscape, one should not discard the relevance of passenger mutations for in studying either tumorigenic mechanisms or questions of therapeutic relevance. As has been demonstrated by characterizing passenger mutations across 21 breast genomes, passengers are not merely mutational litter, but rather an accurate historical record of the process of tumorigenesis and the defects in DNA repair pathways. This mutational

archive can be used, for example, to predict *BRCA1* or *BRCA2* defects and response to PARP inhibitors, using an algorithm known as HRDetect (*368*).

In addition to wild-type drivers, our OG screens provided an opportunity to functionally validate tumor-associated point mutations in a systematic fashion. In general, we found that incorporating these point mutations into ORFs enhanced their ability to promote proliferation. A particularly striking example of this effect was the *MYCN* P44L mutation, which had not previously been functionally characterized. This mutant OG was in fact so potent that it partially limited the dynamic range of the rest of the reagents in the screen pool.

Although the compact nature of the TSG and OG libraries enables high-throughput screening applications, it is important to recall that they represent a concentrated set of potent drivers. Compared to a genome-scale library, in which most reagents should behave neutrally, the OG library in particular consists almost entirely of potent drivers of proliferation. Overall, the p-values and dynamic range that are observable with a genome-scale library tend to be more favorable. This represents a trade-off in the utilization of these small libraries of potent reagents, compared with genome-scale libraries.

Tumor suppressor genes have been conceptualized as gatekeepers and caretakers, with gatekeepers generally restricting growth of aberrant cells and caretakers maintaining the integrity of the genome. We note that we can interpret our TSG screens within this framework; gatekeepers such as TP53, *CDKN2A*, and *RB1* enrich in

our screens, while caretakers such as *ATR* and *SF1* drop out. Although most of the genes that drop out potently by our CRISPR-mediated TSG library are essential genes, we may be able to create a functional classification system for TSGs based on the results of our expanded set of phenotypic screens. It is not clear to us why TUSON predicted so many essential genes as potential TSGs, but this may represent a question for further investigation

Improving the Durability of Cancer Therapeutic Responses

Beginning with the incipient mutation that ultimately leads to the formation of a tumor, oncogenesis is a process of competitive clonal evolution. Tumors acquire phenotypic heterogeneity through successive genetic and epigenetic alterations, that ultimately endows the tumor with a great deal of plasticity. Studies of intratumor heterogeneity before and after chemotherapy have highlighted this plasticity, showing efficient clonal evolution in response to treatment with chemotherapeutics (*122*). This intrinsic heterogeneity and plasticity likely underlie the drug resistance that almost invariably develops to targeted therapies such as tyrosine kinase inhibitors (TKIs).

Targeting a gene such as PARP that is synthetic lethal with the tumor-specific alteration represents an alternative therapeutic strategy. In both TKI therapy and synthetic lethal therapy, the tumor is dependent upon the pathway targeted, through either oncogene addiction or non-oncogene addiction respectively. Yet exploiting synthetic lethal relationships offers access to drug targets that are not themselves

169

mutant in the tumor. *A priori* it is hard to know if this strategy would offer more durable clinical responses, but at least in the case of PARP inhibitors in metastatic breast cancer, acquired resistance appears to be a common problem.

It has been suggested that upfront rational polytherapy may be the only way to produce a durable clinical response with these types of drugs. For example, administering POLΘ and PARP inhibitors simultaneously, upfront, in patients with *BRCA*-deficient breast or ovarian cancer. Part of the rationale for this therapeutic approach lies in the fact that POLΘ is sometimes overexpressed in tumors (*66*). FEN1 is also sometimes overexpressed in breast and other tumors types, and more often amplified than POLΘ in tumors from the METABRIC study (*222*). FEN1 inhibitors may synergize even more strongly with PARP inhibitors than POLΘ inhibitors: FEN1 inhibitors should not only hamper MMEJ but may also block the repair of trapped PARP through long-patch base excision repair (BER). For this reason, we are currently optimizing drug concentrations to perform a drug synergy study between the FEN1 and PARP inhibitors.

Instead of combating the powerful evolutionary forces that drive drug-resistant clones to repopulate a tumor during or after targeted treatment, perhaps the more viable path toward attaining durable clinical responses to chemotherapy lies in the immunotherapeutic approach. The long-term study of pooled overall survival data from patients with advanced melanoma who received ipilimumab was startling: after 10 years of follow-up, the survival curve plateaued (*150*). ~20% of patients survived their case of

unresectable or metastatic melanoma for 10 years after immunotherapy, tempting one to invoke the word "cure."

A path toward durable clinical responses may be to expand this proportion of patients, by maximizing the immune system's capability to combat the tumor. Perhaps there will be a way to introduce benign neoantigenic diversity into tumors, to enhance the antitumor immune response. Personalized cancer vaccines may prompt T-cell based immunity toward greater efficacy. In the meanwhile, a thorough understanding of biomarkers that predict response to PD-1-PD-L1 blockade is of paramount importance in clinical application of this efficacious therapeutic agent. Our data suggests that *CLK2* amplification represents an important risk stratification criterion, and that therapeutic responses in these patients may improve with supplementation of *CLK2* inhibitors. We are pursuing these hypotheses with urgency.

References

1. F. Sanchez-Vega *et al.*, Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321-337 e310 (2018).

2. H. A. Shihab *et al.*, Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat* **34**, 57-65 (2013).

3. T. Boveri, Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. *J Cell Sci* **121 Suppl 1**, 1-84 (2008).

4. P. C. Nowell, D. A. Hungerford, Chromosome studies on normal and leukemic human leukocytes. *J Natl Cancer Inst* **25**, 85-109 (1960).

5. H. Fujiki, Gist of Dr. Katsusaburo Yamagiwa's papers entitled "Experimental study on the pathogenesis of epithelial tumors" (I to VI reports). *Cancer Sci* **105**, 143-149 (2014).

6. H. J. Muller, Artificial Transmutation of the Gene. *Science* **66**, 84-87 (1927).

7. B. N. Ames, Dietary carcinogens and anticarcinogens. Oxygen radicals and degenerative diseases. *Science* **221**, 1256-1264 (1983).

8. L. C. Strong, The induction of mutations by a carcinogen. *Br J Cancer* **3**, 97-108 (1949).

9. P. Rous, A Sarcoma of the Fowl Transmissible by an Agent Separable from the Tumor Cells. *J Exp Med* **13**, 397-411 (1911).

10. H. M. Temin, H. Rubin, Characteristics of an assay for Rous sarcoma virus and Rous sarcoma cells in tissue culture. *Virology* **6**, 669-688 (1958).

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

12. B. Vogelstein *et al.*, Cancer genome landscapes. *Science* **339**, 1546-1558 (2013).

13. C. Tomasetti, B. Vogelstein, G. Parmigiani, Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc Natl Acad Sci U S A* **110**, 1999-2004 (2013).

14.   F. Balaguer *et al.*, Colorectal cancers with microsatellite instability display unique miRNA profiles. *Clin Cancer Res* **17**, 6239-6249 (2011).

15.   T. Davoli *et al.*, Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155**, 948-962 (2013).

16.   P. Armitage, R. Doll, The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer* **8**, 1-12 (1954).

17.   B. Vogelstein, K. W. Kinzler, The Path to Cancer --Three Strikes and You're Out. *N Engl J Med* **373**, 1895-1898 (2015).

18.   C. Tomasetti, L. Marchionni, M. A. Nowak, G. Parmigiani, B. Vogelstein, Only three driver gene mutations are required for the development of lung and colorectal cancers. *Proc Natl Acad Sci U S A* **112**, 118-123 (2015).

19.   C. Kandoth *et al.*, Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333-339 (2013).

20.   R. Beroukhim *et al.*, The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899-905 (2010).

21.   C. H. Mermel *et al.*, GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41 (2011).

22.   P. A. Futreal *et al.*, A census of human cancer genes. *Nat Rev Cancer* **4**, 177-183 (2004).

23.   R. Kruger, Charting a Course to a Cure. *Cell* **173**, 277 (2018).

24.   C. Hutter, J. C. Zenklusen, The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell* **173**, 283-285 (2018).

25.   K. Tomczak, P. Czerwinska, M. Wiznerowicz, The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)* **19**, A68-77 (2015).

26.   M. H. Bailey *et al.*, Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371-385 e318 (2018).

27.   J. J. Michaelson *et al.*, Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* **151**, 1431-1442 (2012).

28. F. C. Nielsen, T. van Overeem Hansen, C. S. Sorensen, Hereditary breast and ovarian cancer: new genes in confined pathways. *Nat Rev Cancer* **16**, 599-612 (2016).

29. L. C. Strong, Genetic etiology of cancer. *Cancer* **40**, 438-444 (1977).

30. A. G. Knudson, Jr., Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**, 820-823 (1971).

31. S. Hodgson, Mechanisms of inherited cancer susceptibility. *J Zhejiang Univ Sci B* **9**, 1-4 (2008).

32. D. Malkin *et al.*, Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* **250**, 1233-1238 (1990).

33. O. M. Sieber, K. Heinimann, I. P. Tomlinson, Genomic instability--the engine of tumorigenesis? *Nat Rev Cancer* **3**, 701-708 (2003).

34. R. Wooster, B. L. Weber, Breast and ovarian cancer. *N Engl J Med* **348**, 2339-2347 (2003).

35. M. S. Brose *et al.*, Cancer risk estimates for BRCA1 mutation carriers identified in a risk evaluation program. *J Natl Cancer Inst* **94**, 1365-1372 (2002).

36. D. Thompson, D. F. Easton, C. Breast Cancer Linkage, Cancer Incidence in BRCA1 mutation carriers. *J Natl Cancer Inst* **94**, 1358-1365 (2002).

37. N. Petrucelli, M. B. Daly, G. L. Feldman, Hereditary breast and ovarian cancer due to mutations in BRCA1 and BRCA2. *Genet Med* **12**, 245-259 (2010).

38. R. Roy, J. Chun, S. N. Powell, BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat Rev Cancer* **12**, 68-78 (2011).

39. K. Kast *et al.*, Prevalence of BRCA1/2 germline mutations in 21 401 families with breast and ovarian cancer. *J Med Genet* **53**, 465-471 (2016).

40. F. Zhang *et al.*, PALB2 links BRCA1 and BRCA2 in the DNA-damage response. *Curr Biol* **19**, 524-529 (2009).

41. U. K. Westermark *et al.*, BARD1 participates with BRCA1 in homology-directed repair of chromosome breaks. *Mol Cell Biol* **23**, 7926-7936 (2003).

42. J. Bartkova *et al.*, Aberrations of the MRE11-RAD50-NBS1 DNA damage sensor complex in human breast cancer: MRE11 as a candidate familial cancer-predisposing gene. *Mol Oncol* **2**, 296-316 (2008).

43. N. Bogdanova *et al.*, NBS1 variant I171V and breast cancer risk. *Breast Cancer Res Treat* **112**, 75-79 (2008).

44. F. Damiola *et al.*, Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study. *Breast Cancer Res* **16**, R58 (2014).

45. H. Song *et al.*, Contribution of Germline Mutations in the RAD51B, RAD51C, and RAD51D Genes to Ovarian Cancer in the Population. *J Clin Oncol* **33**, 2901-2907 (2015).

46. C. Loveday *et al.*, Germline RAD51C mutations confer susceptibility to ovarian cancer. *Nat Genet* **44**, 475-476; author reply 476 (2012).

47. T. A. Knijnenburg *et al.*, Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell Rep* **23**, 239-254 e236 (2018).

48. L. Ding *et al.*, Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics. *Cell* **173**, 305-320 e310 (2018).

49. K. L. Huang *et al.*, Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell* **173**, 355-370 e314 (2018).

50. D. Hanahan, R. A. Weinberg, The hallmarks of cancer. *Cell* **100**, 57-70 (2000).

51. S. Friberg, S. Mattson, On the growth rates of human malignant tumors: implications for medical decision making. *J Surg Oncol* **65**, 284-297 (1997).

52. M. L. Taddei, E. Giannoni, T. Fiaschi, P. Chiarugi, Anoikis: an emerging hallmark in health and diseases. *J Pathol* **226**, 380-393 (2012).

53. J. W. Shay, W. E. Wright, Hayflick, his limit, and cellular ageing. *Nat Rev Mol Cell Biol* **1**, 72-76 (2000).

54. M. B. Sporn, The war on cancer. *Lancet* **347**, 1377-1381 (1996).

55. G. Kroemer, J. Pouyssegur, Tumor cell metabolism: cancer's Achilles' heel. *Cancer Cell* **13**, 472-482 (2008).

56. J. Luo, N. L. Solimini, S. J. Elledge, Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell* **136**, 823-837 (2009).

57. J. Cairns, Mutation selection and the natural history of cancer. *Nature* **255**, 197-200 (1975).

58. L. A. Loeb, Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res* **51**, 3075-3079 (1991).

59. C. Lengauer, K. W. Kinzler, B. Vogelstein, Genetic instabilities in human cancers. *Nature* **396**, 643-649 (1998).

60. K. W. Kinzler, B. Vogelstein, Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386**, 761, 763 (1997).

61. A. Ciccia, S. J. Elledge, The DNA damage response: making it safe to play with knives. *Mol Cell* **40**, 179-204 (2010).

62. J. W. Harper, S. J. Elledge, The DNA damage response: ten years after. *Mol Cell* **28**, 739-745 (2007).

63. S. Negrini, V. G. Gorgoulis, T. D. Halazonetis, Genomic instability--an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* **11**, 220-228 (2010).

64. L. A. Loeb, C. F. Springgate, N. Battula, Errors in DNA replication as a basis of malignant changes. *Cancer Res* **34**, 2311-2321 (1974).

65. P. C. Nowell, The clonal evolution of tumor cell populations. *Science* **194**, 23-28 (1976).

66. G. S. Higgins, S. J. Boulton, Beyond PARP-POLtheta as an anticancer target. *Science* **359**, 1217-1218 (2018).

67. J. E. Cleaver, Cancer in xeroderma pigmentosum and related disorders of DNA repair. *Nat Rev Cancer* **5**, 564-573 (2005).

68. R. D. Kennedy, A. D. D'Andrea, DNA repair pathways in clinical practice: lessons from pediatric cancer susceptibility syndromes. *J Clin Oncol* **24**, 3799-3808 (2006).

69. T. Ripperger, D. Gadzicki, A. Meindl, B. Schlegelberger, Breast cancer susceptibility: current knowledge and implications for genetic counselling. *Eur J Hum Genet* **17**, 722-731 (2009).

70. C. Z. Bachrati, I. D. Hickson, RecQ helicases: suppressors of tumorigenesis and premature aging. *Biochem J* **374**, 577-606 (2003).

71. R. Fishel *et al.*, The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75**, 1027-1038 (1993).

72. F. S. Leach *et al.*, Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**, 1215-1225 (1993).

73. S. Nik-Zainal, S. Morganella, Mutational Signatures in Breast Cancer: The Problem at the DNA Level. *Clin Cancer Res* **23**, 2617-2629 (2017).

74. S. Nik-Zainal *et al.*, Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979-993 (2012).

75. T. Lindahl, D. E. Barnes, Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol* **65**, 127-133 (2000).

76. J. H. Hoeijmakers, DNA damage, aging, and cancer. *N Engl J Med* **361**, 1475-1485 (2009).

77. D. H. Phillips, A. Hewer, C. N. Martin, R. C. Garner, M. M. King, Correlation of DNA adduct levels in human lung with cigarette smoking. *Nature* **336**, 790-792 (1988).

78. J. Jiricny, The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol* **7**, 335-346 (2006).

79. K. L. Limpose, A. H. Corbett, P. W. Doetsch, BERing the burden of damage: Pathway crosstalk and posttranslational modification of base excision repair proteins regulate DNA damage management. *DNA Repair (Amst)* **56**, 51-64 (2017).

80. A. B. Robertson, A. Klungland, T. Rognes, I. Leiros, DNA repair in mammalian cells: Base excision repair: the long and short of it. *Cell Mol Life Sci* **66**, 981-993 (2009).

81. J. A. Marteijn, H. Lans, W. Vermeulen, J. H. Hoeijmakers, Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol* **15**, 465-481 (2014).

82. R. L. Flynn, L. Zou, ATR: a master conductor of cellular responses to DNA replication stress. *Trends Biochem Sci* **36**, 133-140 (2011).

83. A. Marechal, L. Zou, DNA damage sensing by the ATM and ATR kinases. *Cold Spring Harb Perspect Biol* **5**, (2013).

84. A. N. Blackford, S. P. Jackson, ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. *Mol Cell* **66**, 801-817 (2017).

85. M. McVey, S. E. Lee, MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* **24**, 529-538 (2008).

86. B. J. Lamarche, N. I. Orazio, M. D. Weitzman, The MRN complex in double-strand break repair and telomere maintenance. *FEBS Lett* **584**, 3682-3695 (2010).

87. S. F. Bunting *et al.*, 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* **141**, 243-254 (2010).

88. S. Panier, S. J. Boulton, Double-strand break repair: 53BP1 comes into focus. *Nat Rev Mol Cell Biol* **15**, 7-18 (2014).

89. J. R. Chapman, M. R. Taylor, S. J. Boulton, Playing the end game: DNA double-strand break repair pathway choice. *Mol Cell* **47**, 497-510 (2012).

90. N. Shima, R. J. Munroe, J. C. Schimenti, The mouse genomic instability mutation chaos1 is an allele of Polq that exhibits genetic interaction with Atm. *Mol Cell Biol* **24**, 10381-10389 (2004).

91. R. Ceccaldi *et al.*, Homologous-recombination-deficient tumours are dependent on Poltheta-mediated repair. *Nature* **518**, 258-262 (2015).

92. T. N. Schumacher, R. D. Schreiber, Neoantigens in cancer immunotherapy. *Science* **348**, 69-74 (2015).

93. S. Vranic, Microsatellite instability status predicts response to anti-PD-1/PD-L1 therapy regardless the histotype: A comment on recent advances. *Bosn J Basic Med Sci* **17**, 274-275 (2017).

94. S. H. Baumeister, G. J. Freeman, G. Dranoff, A. H. Sharpe, Coinhibitory Pathways in Immunotherapy for Cancer. *Annu Rev Immunol* **34**, 539-573 (2016).

95. M. J. Smyth *et al.*, Perforin-mediated cytotoxicity is critical for surveillance of spontaneous lymphoma. *J Exp Med* **192**, 755-760 (2000).

96. M. J. Smyth *et al.*, Differential tumor surveillance by natural killer (NK) and NKT cells. *J Exp Med* **191**, 661-668 (2000).

97.   M. J. Smyth, D. I. Godfrey, NKT cells and tumor immunity--a double-edged
       sword. *Nat Immunol* **1**, 459-460 (2000).

98.   V. Shankaran *et al.*, IFNgamma and lymphocytes prevent primary tumour
       development and shape tumour immunogenicity. *Nature* **410**, 1107-1111 (2001).

99.   G. P. Dunn, A. T. Bruce, H. Ikeda, L. J. Old, R. D. Schreiber, Cancer
       immunoediting: from immunosurveillance to tumor escape. *Nat Immunol* **3**, 991-
       998 (2002).

100.   J. Galon, H. K. Angell, D. Bedognetti, F. M. Marincola, The continuum of cancer
       immunosurveillance: prognostic, predictive, and mechanistic signatures. *Immunity*
       39, 11-26 (2013).

101.   C. Sun, R. Mezzadra, T. N. Schumacher, Regulation and Function of the PD-L1
       Checkpoint. *Immunity* **48**, 434-452 (2018).

102.   M. E. Keir, M. J. Butte, G. J. Freeman, A. H. Sharpe, PD-1 and its ligands in
       tolerance and immunity. *Annu Rev Immunol* **26**, 677-704 (2008).

103.   E. N. Rozali, S. V. Hato, B. W. Robinson, R. A. Lake, W. J. Lesterhuis,
       Programmed death ligand 2 in cancer-induced immune suppression. *Clin Dev
       Immunol* **2012**, 656340 (2012).

104.   Y. Tanaka *et al.*, Expression pattern of PD-L1 and PD-L2 in classical Hodgkin
       lymphoma, primary mediastinal large B-cell lymphoma, and gray zone
       lymphoma. *Eur J Haematol*, (2018).

105.   W. G. Kaelin, Jr., Synthetic lethality: a framework for the development of wiser
       cancer therapeutics. *Genome Med* **1**, 99 (2009).

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

107.   V. C. Jordan, Tamoxifen: a most unlikely pioneering medicine. *Nat Rev Drug
       Discov* **2**, 205-213 (2003).

108.   S. Farber, L. K. Diamond, Temporary remissions in acute leukemia in children
       produced by folic acid antagonist, 4-aminopteroyl-glutamic acid. *N Engl J Med*
       238, 787-793 (1948).

109. M. J. Osborn, M. Freeman, F. M. Huennekens, Inhibition of dihydrofolic reductase by aminopterin and amethopterin. *Proc Soc Exp Biol Med* **97**, 429-431 (1958).

110. A. de Klein *et al.*, A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukaemia. *Nature* **300**, 765-767 (1982).

111. B. J. Druker *et al.*, Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N Engl J Med* **344**, 1031-1037 (2001).

112. A. Hochhaus *et al.*, Long-Term Outcomes of Imatinib Treatment for Chronic Myeloid Leukemia. *N Engl J Med* **376**, 917-927 (2017).

113. D. J. Slamon *et al.*, Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* **344**, 783-792 (2001).

114. A. E. Wakeling *et al.*, ZD1839 (Iressa): an orally active inhibitor of epidermal growth factor signaling with potential for cancer therapy. *Cancer Res* **62**, 5749-5754 (2002).

115. R. S. Herbst *et al.*, Gefitinib in combination with paclitaxel and carboplatin in advanced non-small-cell lung cancer: a phase III trial--INTACT 2. *J Clin Oncol* **22**, 785-794 (2004).

116. G. Giaccone *et al.*, Gefitinib in combination with gemcitabine and cisplatin in advanced non-small-cell lung cancer: a phase III trial--INTACT 1. *J Clin Oncol* **22**, 777-784 (2004).

117. T. J. Lynch *et al.*, Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350**, 2129-2139 (2004).

118. T. S. Mok *et al.*, Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* **361**, 947-957 (2009).

119. R. Rosell *et al.*, Screening for epidermal growth factor receptor mutations in lung cancer. *N Engl J Med* **361**, 958-967 (2009).

120. F. Morgillo, C. M. Della Corte, M. Fasano, F. Ciardiello, Mechanisms of resistance to EGFR-targeted drugs: lung cancer. *ESMO Open* **1**, e000060 (2016).

121. I. B. Weinstein, Cancer. Addiction to oncogenes--the Achilles heal of cancer. *Science* **297**, 63-64 (2002).

122. D. A. Landau *et al.*, Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714-726 (2013).

123. A. Swaika, J. A. Crozier, R. W. Joseph, Vemurafenib: an evidence-based review of its clinical utility in the treatment of metastatic melanoma. *Drug Des Devel Ther* **8**, 775-787 (2014).

124. W. Pao *et al.*, Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med* **2**, e73 (2005).

125. E. L. Kwak *et al.*, Irreversible inhibitors of the EGF receptor may circumvent acquired resistance to gefitinib. *Proc Natl Acad Sci U S A* **102**, 7665-7670 (2005).

126. D. S. Neel, T. G. Bivona, Resistance is futile: overcoming resistance to targeted therapies in lung adenocarcinoma. *NPJ Precis Oncol* **1**, (2017).

127. N. L. Solimini, J. Luo, S. J. Elledge, Non-oncogene addiction and the stress phenotype of cancer cells. *Cell* **130**, 986-988 (2007).

128. P. G. Richardson, C. Mitsiades, T. Hideshima, K. C. Anderson, Bortezomib: proteasome inhibition as an effective anticancer therapy. *Annu Rev Med* **57**, 33-47 (2006).

129. D. B. Doroshow, J. P. Eder, P. M. LoRusso, BET inhibitors: a novel epigenetic approach. *Ann Oncol* **28**, 1776-1787 (2017).

130. J. D. Kessler *et al.*, A SUMOylation-dependent transcriptional subprogram is required for Myc-driven tumorigenesis. *Science* **335**, 348-353 (2012).

131. T. D. Martin *et al.*, A Role for Mitochondrial Translation in Promotion of Viability in K-Ras Mutant Cells. *Cell Rep* **20**, 427-438 (2017).

132. J. Luo *et al.*, A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137**, 835-848 (2009).

133. T. Davoli *et al.*, Functional genomics reveals that tumors with activating phosphoinositide 3-kinase mutations are dependent on accelerated protein turnover. *Genes Dev* **30**, 2684-2695 (2016).

134. C. E. Strom *et al.*, Poly (ADP-ribose) polymerase (PARP) is not involved in base excision repair but PARP inhibition traps a single-strand intermediate. *Nucleic Acids Res* **39**, 3166-3175 (2011).

135. T. Helleday, The underlying mechanism for the PARP and BRCA synthetic lethality: clearing up the misunderstandings. *Mol Oncol* **5**, 387-393 (2011).

136. J. Murai *et al.*, Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. *Cancer Res* **72**, 5588-5599 (2012).

137. M. Robson, C. Goessl, S. Domchek, Olaparib for Metastatic Germline BRCA-Mutated Breast Cancer. *N Engl J Med* **377**, 1792-1793 (2017).

138. M. Robson *et al.*, Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. *N Engl J Med* **377**, 523-533 (2017).

139. T. Evans, U. Matulonis, PARP inhibitors in ovarian cancer: evidence, experience and clinical potential. *Ther Adv Med Oncol* **9**, 253-267 (2017).

140. B. G. Bitler, Z. L. Watson, L. J. Wheeler, K. Behbakht, PARP inhibitors: Clinical utility and possibilities of overcoming resistance. *Gynecol Oncol* **147**, 695-704 (2017).

141. C. J. Lord, A. Ashworth, PARP inhibitors: Synthetic lethality in the clinic. *Science* **355**, 1152-1158 (2017).

142. S. Guillemette *et al.*, Resistance to therapy in BRCA2 mutant cells due to loss of the nucleosome remodeling factor CHD4. *Genes Dev* **29**, 489-494 (2015).

143. D. R. Leach, M. F. Krummel, J. P. Allison, Enhancement of antitumor immunity by CTLA-4 blockade. *Science* **271**, 1734-1736 (1996).

144. M. J. Selby *et al.*, Anti-CTLA-4 antibodies of IgG2a isotype enhance antitumor activity through reduction of intratumoral regulatory T cells. *Cancer Immunol Res* **1**, 32-42 (2013).

145. F. S. Hodi *et al.*, Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* **363**, 711-723 (2010).

146. T. J. Lynch *et al.*, Ipilimumab in combination with paclitaxel and carboplatin as first-line treatment in stage IIIB/IV non-small-cell lung cancer: results from a randomized, double-blind, multicenter phase II study. *J Clin Oncol* **30**, 2046-2054 (2012).

147. B. Sangro *et al.*, A clinical trial of CTLA-4 blockade with tremelimumab in patients with hepatocellular carcinoma and chronic hepatitis C. *J Hepatol* **59**, 81-88 (2013).

148. K. Y. Chung *et al.*, Phase II study of the anti-cytotoxic T-lymphocyte-associated antigen 4 monoclonal antibody, tremelimumab, in patients with refractory metastatic colorectal cancer. *J Clin Oncol* **28**, 3485-3490 (2010).

149. L. Calabro *et al.*, Tremelimumab for patients with chemotherapy-resistant advanced malignant mesothelioma: an open-label, single-arm, phase 2 trial. *Lancet Oncol* **14**, 1104-1111 (2013).

150. D. Schadendorf *et al.*, Pooled Analysis of Long-Term Survival Data From Phase II and Phase III Trials of Ipilimumab in Unresectable or Metastatic Melanoma. *J Clin Oncol* **33**, 1889-1894 (2015).

151. S. J. Antonia *et al.*, Durvalumab after Chemoradiotherapy in Stage III Non-Small-Cell Lung Cancer. *N Engl J Med* **377**, 1919-1929 (2017).

152. H. Borghaei *et al.*, Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med* **373**, 1627-1639 (2015).

153. J. Brahmer *et al.*, Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. *N Engl J Med* **373**, 123-135 (2015).

154. L. Fehrenbacher *et al.*, Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *Lancet* **387**, 1837-1846 (2016).

155. R. S. Herbst *et al.*, Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet* **387**, 1540-1550 (2016).

156. M. Reck *et al.*, Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med* **375**, 1823-1833 (2016).

157. C. J. Langer *et al.*, Carboplatin and pemetrexed with or without pembrolizumab for advanced, non-squamous non-small-cell lung cancer: a randomised, phase 2 cohort of the open-label KEYNOTE-021 study. *Lancet Oncol* **17**, 1497-1508 (2016).

158. J. R. Brahmer *et al.*, Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity,

pharmacodynamics, and immunologic correlates. *J Clin Oncol* **28**, 3167-3175 (2010).

159. J. Larkin, F. S. Hodi, J. D. Wolchok, Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *N Engl J Med* **373**, 1270-1271 (2015).

160. A. Ribas *et al.*, Pembrolizumab versus investigator-choice chemotherapy for ipilimumab-refractory melanoma (KEYNOTE-002): a randomised, controlled, phase 2 trial. *Lancet Oncol* **16**, 908-918 (2015).

161. J. S. Weber *et al.*, Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial. *Lancet Oncol* **16**, 375-384 (2015).

162. J. Weber *et al.*, Adjuvant Nivolumab versus Ipilimumab in Resected Stage III or IV Melanoma. *N Engl J Med* **377**, 1824-1835 (2017).

163. C. Robert *et al.*, Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med* **372**, 320-330 (2015).

164. C. Robert *et al.*, Pembrolizumab versus Ipilimumab in Advanced Melanoma. *N Engl J Med* **372**, 2521-2532 (2015).

165. D. T. Le *et al.*, PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* **372**, 2509-2520 (2015).

166. D. T. Le *et al.*, Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409-413 (2017).

167. M. J. Overman *et al.*, Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet Oncol* **18**, 1182-1191 (2017).

168. R. J. Motzer *et al.*, Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med* **373**, 1803-1813 (2015).

169. S. M. Ansell *et al.*, PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma. *N Engl J Med* **372**, 311-319 (2015).

170. R. Chen *et al.*, Phase II Study of the Efficacy and Safety of Pembrolizumab for Relapsed/Refractory Classic Hodgkin Lymphoma. *J Clin Oncol* **35**, 2125-2132 (2017).

171. R. L. Ferris *et al.*, Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. *N Engl J Med* **375**, 1856-1867 (2016).

172. T. Y. Seiwert *et al.*, Safety and clinical activity of pembrolizumab for treatment of recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-012): an open-label, multicentre, phase 1b trial. *Lancet Oncol* **17**, 956-965 (2016).

173. H. L. Kaufman *et al.*, Avelumab in patients with chemotherapy-refractory metastatic Merkel cell carcinoma: a multicentre, single-group, open-label, phase 2 trial. *Lancet Oncol* **17**, 1374-1385 (2016).

174. A. B. El-Khoueiry *et al.*, Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet* **389**, 2492-2502 (2017).

175. A. V. Balar *et al.*, First-line pembrolizumab in cisplatin-ineligible patients with locally advanced and unresectable or metastatic urothelial cancer (KEYNOTE-052): a multicentre, single-arm, phase 2 study. *Lancet Oncol* **18**, 1483-1492 (2017).

176. A. V. Balar *et al.*, Atezolizumab as first-line treatment in cisplatin-ineligible patients with locally advanced and metastatic urothelial carcinoma: a single-arm, multicentre, phase 2 trial. *Lancet* **389**, 67-76 (2017).

177. J. Bellmunt *et al.*, Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma. *N Engl J Med* **376**, 1015-1026 (2017).

178. M. R. Patel *et al.*, Avelumab in metastatic urothelial carcinoma after platinum failure (JAVELIN Solid Tumor): pooled results from two expansion cohorts of an open-label, phase 1 trial. *Lancet Oncol* **19**, 51-64 (2018).

179. T. Powles *et al.*, Efficacy and Safety of Durvalumab in Locally Advanced or Metastatic Urothelial Carcinoma: Updated Results From a Phase 1/2 Open-label Study. *JAMA Oncol* **3**, e172411 (2017).

180. J. E. Rosenberg *et al.*, Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with

platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* **387**, 1909-1920 (2016).

181. P. Sharma *et al.*, Nivolumab in metastatic urothelial carcinoma after platinum therapy (CheckMate 275): a multicentre, single-arm, phase 2 trial. *Lancet Oncol* **18**, 312-322 (2017).

182. C. S. Fuchs *et al.*, Safety and Efficacy of Pembrolizumab Monotherapy in Patients With Previously Treated Advanced Gastric and Gastroesophageal Junction Cancer: Phase 2 Clinical KEYNOTE-059 Trial. *JAMA Oncol*, (2018).

183. S. Iwama *et al.*, Pituitary expression of CTLA-4 mediates hypophysitis secondary to administration of CTLA-4 blocking antibody. *Sci Transl Med* **6**, 230ra245 (2014).

184. K. E. Beck *et al.*, Enterocolitis in patients with cancer after antibody blockade of cytotoxic T-lymphocyte-associated antigen 4. *J Clin Oncol* **24**, 2283-2289 (2006).

185. A. B. Schulze, L. H. Schmidt, PD-1 targeted Immunotherapy as first-line therapy for advanced non-small-cell lung cancer patients. *J Thorac Dis* **9**, E384-E386 (2017).

186. J. Gong, A. Chehrazi-Raffle, S. Reddi, R. Salgia, Development of PD-1 and PD-L1 inhibitors as a form of cancer immunotherapy: a comprehensive review of registration trials and future considerations. *J Immunother Cancer* **6**, 8 (2018).

187. B. Liu, Y. Song, D. Liu, Recent development in clinical applications of PD-1 and PD-L1 antibodies for cancer immunotherapy. *J Hematol Oncol* **10**, 174 (2017).

188. L. Gandhi *et al.*, Pembrolizumab plus Chemotherapy in Metastatic Non-Small-Cell Lung Cancer. *N Engl J Med*, (2018).

189. R. H. Vonderheide, S. M. Domchek, A. S. Clark, Immunotherapy for Breast Cancer: What Are We Missing? *Clin Cancer Res* **23**, 2640-2646 (2017).

190. M. Janiczek *et al.*, Immunotherapy as a Promising Treatment for Prostate Cancer: A Systematic Review. *J Immunol Res* **2017**, 4861570 (2017).

191. A. M. Goodman *et al.*, Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol Cancer Ther* **16**, 2598-2608 (2017).

192. N. A. Rizvi *et al.*, Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124-128 (2015).

193. P. C. Tumeh *et al.*, PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568-571 (2014).

194. B. Seliger, The link between MHC class I abnormalities of tumors, oncogenes, tumor suppressor genes, and transcription factors. *J Immunotoxicol* **11**, 308-310 (2014).

195. T. Noguchi *et al.*, Temporally Distinct PD-L1 Expression by Tumor and Host Cells Contributes to Immune Escape. *Cancer Immunol Res* **5**, 106-117 (2017).

196. J. Lau *et al.*, Tumour and host cell PD-L1 is required to mediate suppression of anti-tumour immunity in mice. *Nat Commun* **8**, 14572 (2017).

197. V. R. Juneja *et al.*, PD-L1 on tumor cells is sufficient for immune evasion in immunogenic tumors and inhibits CD8 T cell cytotoxicity. *J Exp Med* **214**, 895-904 (2017).

198. G. Lin *et al.*, Prognostic significance of PD-L1 expression and tumor infiltrating lymphocyte in surgically resectable non-small cell lung cancer. *Oncotarget* **8**, 83986-83994 (2017).

199. S. Mariathasan *et al.*, TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544-548 (2018).

200. K. A. Schalper, E. Kaftan, R. S. Herbst, Predictive Biomarkers for PD-1 Axis Therapies: The Hidden Treasure or a Call for Research. *Clin Cancer Res* **22**, 2102-2104 (2016).

201. J. M. Zaretsky *et al.*, Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *N Engl J Med* **375**, 819-829 (2016).

202. V. Cianfanelli *et al.*, AMBRA1 links autophagy to cell proliferation and tumorigenesis by promoting c-Myc dephosphorylation and degradation. *Nat Cell Biol* **17**, 20-30 (2015).

203. P. A. Mateos-Gomez *et al.*, Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination. *Nature* **518**, 254-257 (2015).

204. H. Kobayashi, S. Ohno, Y. Sasaki, M. Matsuura, Hereditary breast and ovarian cancer susceptibility genes (review). *Oncol Rep* **30**, 1019-1029 (2013).

205. P. L. Welcsh, M. C. King, BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum Mol Genet* **10**, 705-713 (2001).

206. S. A. Narod, Modifiers of risk of hereditary breast cancer. *Oncogene* **25**, 5832-5836 (2006).

207. N. Petrucelli, M. B. Daly, T. Pal, in *GeneReviews((R)),* M. P. Adam *et al.*, Eds. (Seattle (WA), 1993).

208. K. Schlacher *et al.*, Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* **145**, 529-542 (2011).

209. J. Spies *et al.*, Nek1 Regulates Rad54 to Orchestrate Homologous Recombination and Replication Fork Stability. *Mol Cell* **62**, 903-917 (2016).

210. A. M. Kolinjivadi *et al.*, Smarcal1-Mediated Fork Reversal Triggers Mre11-Dependent Degradation of Nascent DNA in the Absence of Brca2 and Stable Rad51 Nucleofilaments. *Mol Cell* **67**, 867-881 e867 (2017).

211. M. Lomonosov, S. Anand, M. Sangrithi, R. Davies, A. R. Venkitaraman, Stabilization of stalled DNA replication forks by the BRCA2 breast cancer susceptibility protein. *Genes Dev* **17**, 3017-3022 (2003).

212. Y. Doksani, T. de Lange, The role of double-strand break repair pathways at functional and dysfunctional telomeres. *Cold Spring Harb Perspect Biol* **6**, a016576 (2014).

213. V. Bhatia *et al.*, BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**, 362-365 (2014).

214. H. E. Bryant *et al.*, Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913-917 (2005).

215. H. Farmer *et al.*, Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917-921 (2005).

216. S. Narod, C. M. Booth, W. D. Foulkes, Olaparib for Metastatic Germline BRCA-Mutated Breast Cancer. *N Engl J Med* **377**, 1792 (2017).

217. M. R. Mirza *et al.*, Niraparib Maintenance Therapy in Platinum-Sensitive, Recurrent Ovarian Cancer. *N Engl J Med* **375**, 2154-2164 (2016).

218. T. Hucl *et al.*, A syngeneic variance library for functional annotation of human variation: application to BRCA2. *Cancer Res* **68**, 5023-5030 (2008).

219. M. R. Schlabach *et al.*, Cancer proliferation gene discovery through functional genomics. *Science* **319**, 620-624 (2008).

220. W. Sakai *et al.*, Functional restoration of BRCA2 protein by secondary BRCA2 mutations in BRCA2-mutated ovarian carcinoma. *Cancer Res* **69**, 6381-6386 (2009).

221. A. Carreira, S. C. Kowalczykowski, Two classes of BRC repeats in BRCA2 promote RAD51 nucleoprotein filament function by distinct mechanisms. *Proc Natl Acad Sci U S A* **108**, 10448-10453 (2011).

222. B. Pereira *et al.*, The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun* **7**, 11479 (2016).

223. W. Li *et al.*, MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **15**, 554 (2014).

224. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).

225. K. E. Zahn, A. M. Averill, P. Aller, R. D. Wood, S. Doublie, Human DNA polymerase theta grasps the primer terminus to mediate DNA repair. *Nat Struct Mol Biol* **22**, 304-311 (2015).

226. C. Cretu *et al.*, Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations. *Mol Cell* **64**, 307-319 (2016).

227. A. Smogorzewska *et al.*, Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell* **129**, 289-301 (2007).

228. W. Sakai *et al.*, Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. *Nature* **451**, 1116-1120 (2008).

229. K. A. Effenberger, V. K. Urabe, B. E. Prichard, A. K. Ghosh, M. S. Jurica, Interchangeable SF3B1 inhibitors interfere with pre-mRNA splicing at multiple stages. *RNA* **22**, 350-359 (2016).

230. H. Nilsen, H. E. Krokan, Base excision repair in a network of defence and tolerance. *Carcinogenesis* **22**, 987-998 (2001).

231. P. Burkovics, V. Szukacsov, I. Unk, L. Haracska, Human Ape2 protein has a 3'-5' exonuclease activity that acts preferentially on mismatched base pairs. *Nucleic Acids Res* **34**, 2508-2515 (2006).

232. M. Z. Hadi, K. Ginalski, L. H. Nguyen, D. M. Wilson, 3rd, Determinants in nuclease specificity of Ape1 and Ape2, human homologues of Escherichia coli exonuclease III. *J Mol Biol* **316**, 853-866 (2002).

233. S. Xanthoudakis, G. G. Miao, T. Curran, The redox and DNA-repair activities of Ref-1 are encoded by nonoverlapping domains. *Proc Natl Acad Sci U S A* **91**, 23-27 (1994).

234. G. Rai *et al.*, Synthesis, biological evaluation, and structure-activity relationships of a novel class of apurinic/apyrimidinic endonuclease 1 inhibitors. *J Med Chem* **55**, 3101-3112 (2012).

235. G. M. Zou, A. Maitra, Small-molecule inhibitor of the AP endonuclease 1/REF-1 E3330 inhibits pancreatic cancer cell growth and migration. *Mol Cancer Ther* **7**, 2012-2021 (2008).

236. R. M. Meyers *et al.*, Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet* **49**, 1779-1784 (2017).

237. L. Balakrishnan, R. A. Bambara, Flap endonuclease 1. *Annu Rev Biochem* **82**, 119-138 (2013).

238. J. C. Exell *et al.*, Cellularly active N-hydroxyurea FEN1 inhibitors block substrate entry to the active site. *Nat Chem Biol* **12**, 815-821 (2016).

239. T. A. Ward, P. J. McHugh, S. T. Durant, Small molecule inhibitors uncover synthetic genetic interactions of human flap endonuclease 1 (FEN1) with DNA damage response genes. *PLoS One* **12**, e0179278 (2017).

240. T. Kent, G. Chandramouly, S. M. McDevitt, A. Y. Ozdemir, R. T. Pomerantz, Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta. *Nat Struct Mol Biol* **22**, 230-237 (2015).

241. A. Sfeir, L. S. Symington, Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends Biochem Sci* **40**, 701-714 (2015).

242. A. Ray Chaudhuri *et al.*, Replication fork stability confers chemoresistance in BRCA-deficient cells. *Nature* **535**, 382-387 (2016).

243. J. C. Saldivar, D. Cortez, K. A. Cimprich, The essential kinase ATR: ensuring faithful duplication of a challenging genome. *Nat Rev Mol Cell Biol* **18**, 622-636 (2017).

244. Y. Ide, D. Tsuchimoto, Y. Tominaga, Y. Iwamoto, Y. Nakabeppu, Characterization of the genomic structure and expression of the mouse Apex2 gene. *Genomics* **81**, 47-57 (2003).

245. D. Tsuchimoto *et al.*, Human APE2 protein is mostly localized in the nuclei and to some extent in the mitochondria, while nuclear APE2 is partly associated with proliferating cell nuclear antigen. *Nucleic Acids Res* **29**, 2349-2360 (2001).

246. M. Tanikawa, K. Sanjiv, T. Helleday, P. Herr, O. Mortusewicz, The spliceosome U2 snRNP factors promote genome stability through distinct mechanisms; transcription of repair factors and R-loop processing. *Oncogenesis* **5**, e280 (2016).

247. F. Stegmeier, G. Hu, R. J. Rickles, G. J. Hannon, S. J. Elledge, A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. *Proc Natl Acad Sci U S A* **102**, 13212-13217 (2005).

248. K. L. Meerbrey *et al.*, The pINDUCER lentiviral toolkit for inducible RNA interference in vitro and in vivo. *Proc Natl Acad Sci U S A* **108**, 3665-3670 (2011).

249. B. Langmead, Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics* **Chapter 11**, Unit 11 17 (2010).

250. W. Li *et al.*, Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol* **16**, 281 (2015).

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

252. B. Chen *et al.*, Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479-1491 (2013).

253. O. Shalem *et al.*, Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87 (2014).

254. J. R. Pritchard *et al.*, Defining principles of combination drug mechanisms of action. *Proc Natl Acad Sci U S A* **110**, E170-179 (2013).

255. L. M. Sack *et al.*, Profound Tissue Specificity in Proliferation Control Underlies Cancer Drivers and Aneuploidy Patterns. *Cell* **173**, 499-514 e423 (2018).

256. A. V. Balar, J. S. Weber, PD-1 and PD-L1 antibodies in cancer: current status and future directions. *Cancer Immunol Immunother* **66**, 551-564 (2017).

257. S. Y. Liu, Y. L. Wu, Ongoing clinical trials of PD-1 and PD-L1 inhibitors for lung cancer in China. *J Hematol Oncol* **10**, 136 (2017).

258. A. Garcia-Diaz *et al.*, Interferon Receptor Signaling Pathways Regulating PD-L1 and PD-L2 Expression. *Cell Rep* **19**, 1189-1201 (2017).

259. R. Schoop *et al.*, Suppressed T-cell activation by IFN-gamma-induced expression of PD-L1 on renal tubular epithelial cells. *Nephrol Dial Transplant* **19**, 2713-2720 (2004).

260. K. Kataoka *et al.*, Aberrant PD-L1 expression through 3'-UTR disruption in multiple cancers. *Nature* **534**, 402-406 (2016).

261. A. Y. Gong *et al.*, MicroRNA-513 regulates B7-H1 translation and is involved in IFN-gamma-induced B7-H1 expression in cholangiocytes. *J Immunol* **182**, 1325-1333 (2009).

262. M. L. Burr *et al.*, CMTM6 maintains the expression of PD-L1 and regulates anti-tumour immunity. *Nature* **549**, 101-105 (2017).

263. R. Mezzadra *et al.*, Identification of CMTM6 and CMTM4 as PD-L1 protein regulators. *Nature* **549**, 106-110 (2017).

264. J. Zhang *et al.*, Cyclin D-CDK4 kinase destabilizes PD-L1 via cullin 3-SPOP to control cancer immune surveillance. *Nature* **553**, 91-95 (2018).

265. J. Gao *et al.*, Loss of IFN-gamma Pathway Genes in Tumor Cells as a Mechanism of Resistance to Anti-CTLA-4 Therapy. *Cell* **167**, 397-404 e399 (2016).

266. R. T. Manguso *et al.*, In vivo CRISPR screening identifies Ptpn2 as a cancer immunotherapy target. *Nature* **547**, 413-418 (2017).

267. M. J. Eppihimer *et al.*, Expression and regulation of the PD-L1 immunoinhibitory molecule on microvascular endothelial cells. *Microcirculation* **9**, 133-145 (2002).

268. B. Schreiner *et al.*, Interferon-beta enhances monocyte and dendritic cell expression of B7-H1 (PD-L1), a strong inhibitor of autologous T-cell activation: relevance for the immune modulatory effect in multiple sclerosis. *J Neuroimmunol* **155**, 172-182 (2004).

269. J. W. Choi *et al.*, LPA receptors: subtypes and biological actions. *Annu Rev Pharmacol Toxicol* **50**, 157-186 (2010).

270. J. W. Moon *et al.*, IFNgamma induces PD-L1 overexpression by JAK2/STAT1/IRF-1 signaling in EBV-positive gastric carcinoma. *Sci Rep* **7**, 17810 (2017).

271. K. Mimura *et al.*, PD-L1 expression is mainly regulated by interferon gamma associated with JAK-STAT pathway in gastric cancer. *Cancer Sci* **109**, 43-53 (2018).

272. T. Doi *et al.*, The JAK/STAT pathway is involved in the upregulation of PD-L1 expression in pancreatic cancer cell lines. *Oncol Rep* **37**, 1545-1554 (2017).

273. J. N. Ou, A. E. Wiedeman, A. M. Stevens, TNF-alpha and TGF-beta counter-regulate PD-L1 expression on monocytes in systemic lupus erythematosus. *Sci Rep* **2**, 295 (2012).

274. A. Starke, R. P. Wuthrich, Y. Waeckerle-Men, TGF-beta treatment modulates PD-L1 and CD40 expression in proximal renal tubular epithelial cells and enhances CD8 cytotoxic T-cell responses. *Nephron Exp Nephrol* **107**, e22-29 (2007).

275. X. Y. Ni *et al.*, TGF-beta of lung cancer microenvironment upregulates B7H1 and GITRL expression in dendritic cells and is associated with regulatory T cell generation. *Oncol Rep* **28**, 615-621 (2012).

276. S. Song *et al.*, Dendritic cells with an increased PD-L1 by TGF-beta induce T cell anergy for the cytotoxicity of hepatocellular carcinoma cells. *Int Immunopharmacol* **20**, 117-123 (2014).

277. M. Baas *et al.*, TGFbeta-dependent expression of PD-1 and PD-L1 controls CD8(+) T cell anergy in transplant tolerance. *Elife* **5**, e08133 (2016).

278. A. C. Poncelet, M. P. de Caestecker, H. W. Schnaper, The transforming growth factor-beta/SMAD signaling pathway is present and functional in human mesangial cells. *Kidney Int* **56**, 1354-1365 (1999).

279. H. Ikedo *et al.*, Smad protein and TGF-beta signaling in vascular smooth muscle cells. *Int J Mol Med* **11**, 645-650 (2003).

280. Y. E. Zhang, Non-Smad pathways in TGF-beta signaling. *Cell Res* **19**, 128-139 (2009).

281. D. P. Brazil, R. H. Church, S. Surae, C. Godson, F. Martin, BMP signalling: agony and antagony in the family. *Trends Cell Biol* **25**, 249-264 (2015).

282. W. Huang da, B. T. Sherman, R. A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).

283. A. Riaz, Y. Huang, S. Johansson, G-Protein-Coupled Lysophosphatidic Acid Receptors and Their Regulation of AKT Signaling. *Int J Mol Sci* **17**, 215 (2016).

284. L. Tong, V. Tergaonkar, Rho protein GTPases and their interactions with NFkappaB: crossroads of inflammation and matrix biology. *Biosci Rep* **34**, (2014).

285. M. A. Coelho *et al.*, Oncogenic RAS Signaling Promotes Tumor Immunoresistance by Stabilizing PD-L1 mRNA. *Immunity* **47**, 1083-1099 e1086 (2017).

286. K. J. Lastwika *et al.*, Control of PD-L1 Expression by Oncogenic Activation of the AKT-mTOR Pathway in Non-Small Cell Lung Cancer. *Cancer Res* **76**, 227-238 (2016).

287. S. C. Casey *et al.*, MYC regulates the antitumor immune response through CD47 and PD-L1. *Science* **352**, 227-231 (2016).

288. E. Y. Kim, A. Kim, S. K. Kim, Y. S. Chang, MYC expression correlates with PD-L1 expression in non-small cell lung cancer. *Lung Cancer* **110**, 63-67 (2017).

289. S. G. Pai *et al.*, Wnt/beta-catenin pathway: modulating anticancer immune response. *J Hematol Oncol* **10**, 101 (2017).

290. T. Dasgupta, A. N. Ladd, The importance of CELF control: molecular and biological roles of the CUG-BP, Elav-like family of RNA-binding proteins. *Wiley Interdiscip Rev RNA* **3**, 104-121 (2012).

291. J. M. Gallo, C. Spickett, The role of CELF proteins in neurological disorders. *RNA Biol* **7**, 474-479 (2010).

292. A. N. Ladd, N. Charlet, T. A. Cooper, The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol Cell Biol* **21**, 1285-1296 (2001).

293. P. J. Good, Q. Chen, S. J. Warner, D. C. Herring, A family of human RNA-binding proteins related to the Drosophila Bruno translational regulator. *J Biol Chem* **275**, 28583-28592 (2000).

294. A. N. Ladd, N. H. Nguyen, K. Malhotra, T. A. Cooper, CELF6, a member of the CELF family of RNA-binding proteins, regulates muscle-specific splicing enhancer-dependent alternative splicing. *J Biol Chem* **279**, 17756-17764 (2004).

295. B. A. Weir *et al.*, Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450**, 893-898 (2007).

296. S. Chauhan *et al.*, Evolution of the Cdk-activator Speedy/RINGO in vertebrates. *Cell Mol Life Sci* **69**, 3835-3850 (2012).

297. L. A. Porter *et al.*, Human Speedy: a novel cell cycle regulator that enhances proliferation through activation of Cdk2. *J Cell Biol* **157**, 357-366 (2002).

298. A. Dinarina *et al.*, Negative regulation of cell-cycle progression by RINGO/Speedy E. *Biochem J* **410**, 535-542 (2008).

299. M. J. Butte, V. Pena-Cruz, M. J. Kim, G. J. Freeman, A. H. Sharpe, Interaction of human PD-L1 and B7-1. *Mol Immunol* **45**, 3567-3572 (2008).

300. T. Yoshida *et al.*, CLK2 Is an Oncogenic Kinase and Splicing Regulator in Breast Cancer. *Cancer Res* **75**, 1516-1526 (2015).

301. E. Cerami *et al.*, The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404 (2012).

302. F. M. Moeslein, M. P. Myers, G. E. Landreth, The CLK family kinases, CLK1 and CLK2, phosphorylate and activate the tyrosine phosphatase, PTP-1B. *J Biol Chem* **274**, 26697-26704 (1999).

303. M. Kleppe *et al.*, PTPN2 negatively regulates oncogenic JAK1 in T-cell acute lymphoblastic leukemia. *Blood* **117**, 7090-7098 (2011).

304. M. Kleppe *et al.*, Deletion of the protein tyrosine phosphatase gene PTPN2 in T-cell acute lymphoblastic leukemia. *Nat Genet* **42**, 530-535 (2010).

305. X. H. He, L. H. Xu, Y. Liu, Identification of a novel splice variant of human PD-L1 mRNA encoding an isoform-lacking Igv-like domain. *Acta Pharmacol Sin* **26**, 462-468 (2005).

306. S. Kaur *et al.*, Dual regulatory roles of phosphatidylinositol 3-kinase in IFN signaling. *J Immunol* **181**, 7316-7323 (2008).

307. E. M. Van Allen *et al.*, Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207-211 (2015).

308. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).

309. J. Gao *et al.*, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1 (2013).

310. M. Imielinski *et al.*, Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107-1120 (2012).

311. N. Cancer Genome Atlas, Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).

312. M. S. Lawrence *et al.*, Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495-501 (2014).

313. K. A. Hoadley *et al.*, Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304 e296 (2018).

314. N. N. Pavlova *et al.*, A role for PVRL4-driven cell-cell interactions in tumorigenesis. *Elife* **2**, e00358 (2013).

315. F. Sanchez-Garcia *et al.*, Integration of genomic data enables selective discovery of breast cancer drivers. *Cell* **159**, 1461-1475 (2014).

316. T. A. Ince *et al.*, Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes. *Cancer Cell* **12**, 160-170 (2007).

317. P. Feijoo *et al.*, Breast primary epithelial cells that escape p16-dependent stasis enter a telomere-driven crisis state. *Breast Cancer Res* **18**, 7 (2016).

318. A. E. Mazzucco *et al.*, Genetic interrogation of replicative senescence uncovers a dual role for USP28 in coordinating the p53 and GATA4 branches of the senescence program. *Genes Dev* **31**, 1933-1938 (2017).

319. C. S. Fong *et al.*, 53BP1 and USP28 mediate p53-dependent cell cycle arrest in response to centrosome loss and prolonged mitosis. *Elife* **5**, (2016).

320. W. Tao, A. J. Levine, P19(ARF) stabilizes p53 by blocking nucleo-cytoplasmic shuttling of Mdm2. *Proc Natl Acad Sci U S A* **96**, 6937-6941 (1999).

321. Chio, II *et al.*, TRADD contributes to tumour suppression by regulating ULF-dependent p19Arf ubiquitylation. *Nat Cell Biol* **14**, 625-633 (2012).

322. G. B. Silberstein, C. W. Daniel, Reversible inhibition of mammary gland growth by transforming growth factor-beta. *Science* **237**, 291-293 (1987).

323. R. D. Williams *et al.*, Multiple mechanisms of MYCN dysregulation in Wilms tumour. *Oncotarget* **6**, 7232-7243 (2015).

324. N. Cancer Genome Atlas Research, Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061-1068 (2008).

325. D. T. Jones *et al.*, Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**, 100-105 (2012).

326. J. Wu *et al.*, Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. *Proc Natl Acad Sci U S A* **108**, 21188-21193 (2011).

327. S. A. Forbes *et al.*, COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* **39**, D945-950 (2011).

328. T. J. Pugh *et al.*, The genetic landscape of high-risk neuroblastoma. *Nat Genet* **45**, 279-284 (2013).

329. D. S. Rickman, J. H. Schulte, M. Eilers, The Expanding World of N-MYC-Driven Tumors. *Cancer Discov* **8**, 150-163 (2018).

330. D. J. Shields *et al.*, Oncogenic Ras/Src cooperativity in pancreatic neoplasia. *Oncogene* **30**, 2123-2134 (2011).

331. R. Schmitz *et al.*, Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature* **490**, 116-120 (2012).

332. N. Xiao *et al.*, Ott1 (Rbm15) regulates thrombopoietin response in hematopoietic stem cells through alternative splicing of c-Mpl. *Blood* **125**, 941-948 (2015).

333. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).

334. S. K. Rhie *et al.*, ZFX acts as a transcriptional activator in multiple types of human tumors by binding downstream of transcription start sites at the majority of CpG island promoters. *Genome Res*, (2018).

335. C. Y. Chen, D. Zheng, Z. Xia, A. B. Shyu, Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nat Struct Mol Biol* **16**, 1160-1166 (2009).

336. A. Munshi, M. Hobbs, R. E. Meyn, Clonogenic cell survival assay. *Methods Mol Med* **110**, 21-28 (2005).

337. P. H. Vachon, Integrin signaling, cell survival, and anoikis: distinctions, differences, and differentiation. *J Signal Transduct* **2011**, 738137 (2011).

338. T. N. Seyfried, L. C. Huysentruyt, On the origin of cancer metastasis. *Crit Rev Oncog* **18**, 43-73 (2013).

339. M. P. van Bragt, X. Hu, Y. Xie, Z. Li, RUNX1, a transcription factor mutated in breast cancer, controls the fate of ER-positive mammary luminal cells. *Elife* **3**, e03881 (2014).

340. F. Vogiatzi *et al.*, Mutant p53 promotes tumor progression and metastasis by the endoplasmic reticulum UDPase ENTPD5. *Proc Natl Acad Sci U S A* **113**, E8433-E8442 (2016).
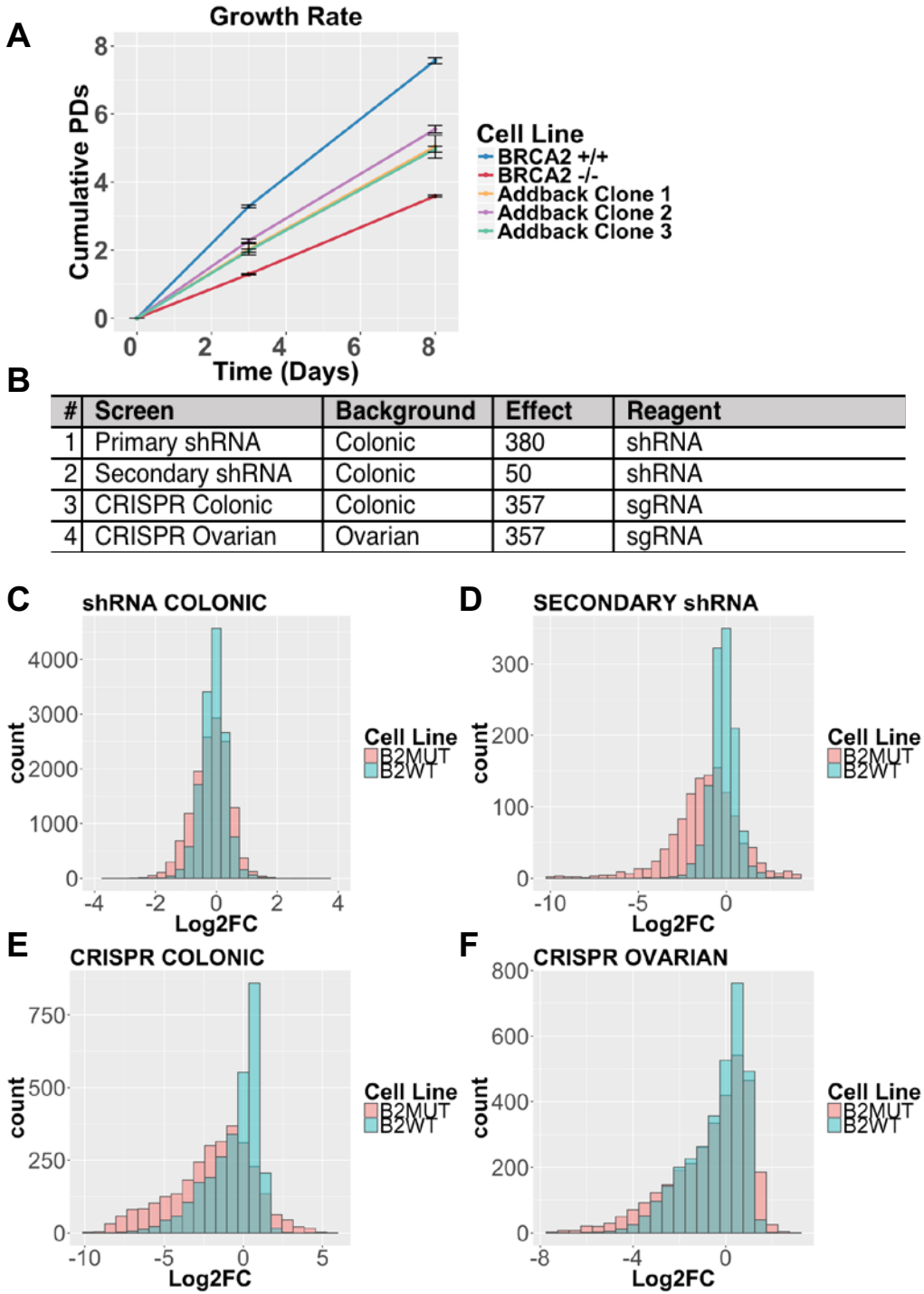
341. M. W. Richards *et al.*, Structural basis of N-Myc binding by Aurora-A and its destabilization by kinase inhibitors. *Proc Natl Acad Sci U S A* **113**, 13726-13731 (2016).

342. V. Cianfanelli *et al.*, AMBRA1 links autophagy to cell proliferation and tumorigenesis by promoting c-Myc dephosphorylation and degradation. *Nat Cell Biol* **17**, 706 (2015).

343. V. Cianfanelli, F. Nazio, F. Cecconi, Connecting autophagy: AMBRA1 and its network of regulation. *Mol Cell Oncol* **2**, e970059 (2015).

344. D. W. Morgens, R. M. Deans, A. Li, M. C. Bassik, Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat Biotechnol* **34**, 634-636 (2016).

345. H. J. Little *et al.*, Missense mutations that cause Van der Woude syndrome and popliteal pterygium syndrome affect the DNA-binding and transcriptional activation functions of IRF6. *Hum Mol Genet* **18**, 535-545 (2009).

346. F. Rahimov *et al.*, Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. *Nat Genet* **40**, 1341-1347 (2008).

347. Z. Chang *et al.*, Cooperativity of oncogenic K-ras and downregulated p16/INK4A in human pancreatic tumorigenesis. *PLoS One* **9**, e101452 (2014).

348. C. Fellmann *et al.*, An optimized microRNA backbone for effective single-copy RNAi. *Cell Rep* **5**, 1704-1713 (2013).

349. L. Aravind, V. M. Dixit, E. V. Koonin, Apoptotic molecular machinery: vastly increased complexity in vertebrates revealed by genome comparisons. *Science* **291**, 1279-1284 (2001).

350. S. Grimm, The art and design of genetic screens: mammalian culture cells. *Nat Rev Genet* **5**, 179-189 (2004).

351. D. Sims *et al.*, High-throughput RNA interference screening using pooled shRNA libraries and next generation sequencing. *Genome Biol* **12**, R104 (2011).

352. H. Hoshiyama *et al.*, Development of methods for quantitative comparison of pooled shRNAs by mass sequencing. *J Biomol Screen* **17**, 258-265 (2012).

353. Z. Strezoska *et al.*, Optimized PCR conditions and increased shRNA fold representation improve reproducibility of pooled shRNA screens. *PLoS One* **7**, e42341 (2012).

354. W. S. Hu, H. M. Temin, Retroviral recombination and reverse transcription. *Science* **250**, 1227-1233 (1990).

355. M. Jinek *et al.*, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).

356. L. Cong *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).

357. J. Shi *et al.*, Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol* **33**, 661-667 (2015).

358. D. M. Munoz *et al.*, CRISPR Screens Provide a Comprehensive Assessment of Cancer Vulnerabilities but Generate False-Positive Hits for Highly Amplified Genomic Regions. *Cancer Discov* **6**, 900-913 (2016).

359. J. G. Doench *et al.*, Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191 (2016).

360. J. G. Doench, Am I ready for CRISPR? A user's guide to genetic screens. *Nat Rev Genet* **19**, 67-80 (2018).

361. L. A. Gilbert *et al.*, CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451 (2013).

362. M. H. Larson *et al.*, CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* **8**, 2180-2196 (2013).

363. P. Perez-Pinera *et al.*, RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* **10**, 973-976 (2013).

364. B. Evers *et al.*, CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat Biotechnol* **34**, 631-633 (2016).

365. A. J. Aguirre *et al.*, Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov* **6**, 914-929 (2016).

366. L. M. Sack, T. Davoli, Q. Xu, M. Z. Li, S. J. Elledge, Sources of Error in Mammalian Genetic Screens. *G3 (Bethesda)* **6**, 2781-2790 (2016).

367. E. Mamessier, D. J. Birnbaum, P. Finetti, D. Birnbaum, F. Bertucci, CMTM6 stabilizes PD-L1 expression and refines its prognostic value in tumors. *Ann Transl Med* **6**, 54 (2018).

368. H. Davies *et al.*, HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat Med* **23**, 517-525 (2017).

Appendix I. Supplemental Figures

Supplemental Figure S2-1 (related to Figure 2-1). (A) The indicated cell lines were passaged in culture for 8 days and counted to determine cumulative population doublings (PDs) at an early (3 days) and late (8 days) time point. **(B)** List of screens performed including the isogenic cell line pair background and the number of genes targeted by the indicated sublibrary. **(C)** Distribution of log2-fold-changes (Log2FC) for each shRNA in the primary shRNA screen, performed in colonic *BRCA2* isogenic cell line background. **(D)** Distribution of log2-fold-changes (Log2FC) for each shRNA in the secondary shRNA screen, performed in colonic *BRCA2* isogenic cell line background. **(E)** Distribution of log2-fold-changes (Log2FC) for each sgRNA in the CRISPR colonic B2SL screen. **(F)** Distribution of log2-fold-changes (Log2FC) for each sgRNA in the CRISPR ovarian B2SL screen.

A Growth Rate

Cumulative PDs vs Time (Days)

Cell Line
- BRCA2 +/+
- BRCA2 -/-
- Addback Clone 1
- Addback Clone 2
- Addback Clone 3

B

#	Screen	Background	Effect	Reagent
1	Primary shRNA	Colonic	380	shRNA
2	Secondary shRNA	Colonic	50	shRNA
3	CRISPR Colonic	Colonic	357	sgRNA
4	CRISPR Ovarian	Ovarian	357	sgRNA

C shRNA COLONIC

D SECONDARY shRNA

E CRISPR COLONIC

F CRISPR OVARIAN

Supplemental Figure S2-1 (related to Figure 2-1) (Continued).
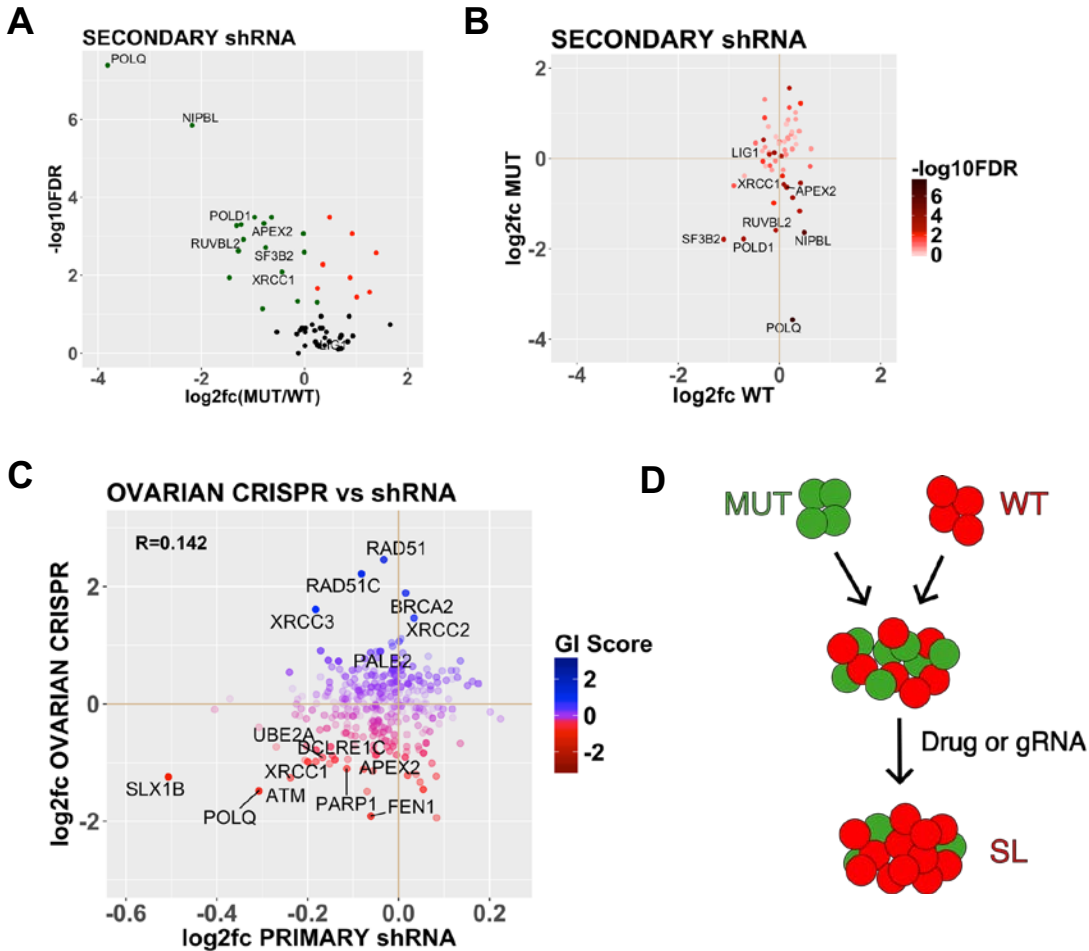
Supplemental Figure S2-2 (related to Figure 2-2). A) Volcano plot for the secondary shRNA screen, performed with the colonic *BRCA2* isogenic cell line pair. Significance (-log$_{10}$FDR) is plotted against the genetic interaction (GI) score (average log2-fold-change for each gene in BRCA2 MUT versus BRCA2 WT cells). Genes that met a significance threshold of -log$_{10}$FDR>1 are color-coded as green for relative dropout or red for relative enrichment in BRCA2 MUT vs BRCA2 WT cells. **(B)** Results from the secondary shRNA screen in the colonic *BRCA2* isogenic background, plotted as the log2-fold-change in BRCA2 MUT cells against the log2-fold-change in BRCA2 WT cells. **(C)** Comparison of GI score (log2-fold-change for each gene in BRCA2 MUT versus BRCA2 WT cells) in the ovarian CRISPR screen versus the colonic shRNA screen. **(D)** Schematic depicting the multicolor competition assay (MCA). GFP-labeled BRCA2 MUT cells are mixed with DsRed-labeled BRCA2 WT cells. The cell mixture is passaged in the presence of a drug or transduced with a virus expressing a gRNA and Cas9. The change in percent in GFP^{+} cells measured by FACS indicates the relative growth of BRCA2 MUT versus BRCA2 WT cells in the presence of the relevant treatment.

Supplemental Figure S2-3 (related to Figures 2-4 and 2-5). (A-C) MCA assays in which ovarian GFP-labeled BRCA2 MUT cells and DsRed-labeled BRCA2 WT cells were mixed and co-infected with Cas9 and negative control gRNAs (A), three individual gRNAs targeting *POLQ* (B), or three individual gRNAs targeting *UBE2A* (C). Negative control gRNAs included gRNAs that do not induce cutting by Cas9 in the human genome (targeting *E. coli*), and gRNAs that cut intergenic regions with low predicted off-target cutting (NC-1, NC-2, and NC-3). **(D)** MCA assay in which colonic GFP-labeled BRCA2 MUT cells and DsRed-labeled BRCA2 WT cells were mixed and co-treated with the PARP inhibitor olaparib. The percent of GFP$^+$ cells was quantified by FACS after 12 days and normalized to DMSO. **(E-I)** MCA assays in which colonic GFP-labeled BRCA2 MUT cells and DsRed-labeled BRCA2 WT cells were mixed and co-infected with Cas9 and three individual gRNAs to the indicated genes. **(J-N)** MCA assays in which GFP-labeled BRCA1 MUT cells and DsRed-labeled BRCA1 WT cells were mixed and co-infected with Cas9 and three individual gRNAs to the indicated genes. **(P).** Dependency of *BRCA1/2* MUT or *BRCA1/2* WT cell lines on *POLQ*, processed by the CERES computational method from genome-scale CRISPR-Cas9 essentiality screens across 324 cancer cell lines (*236*).
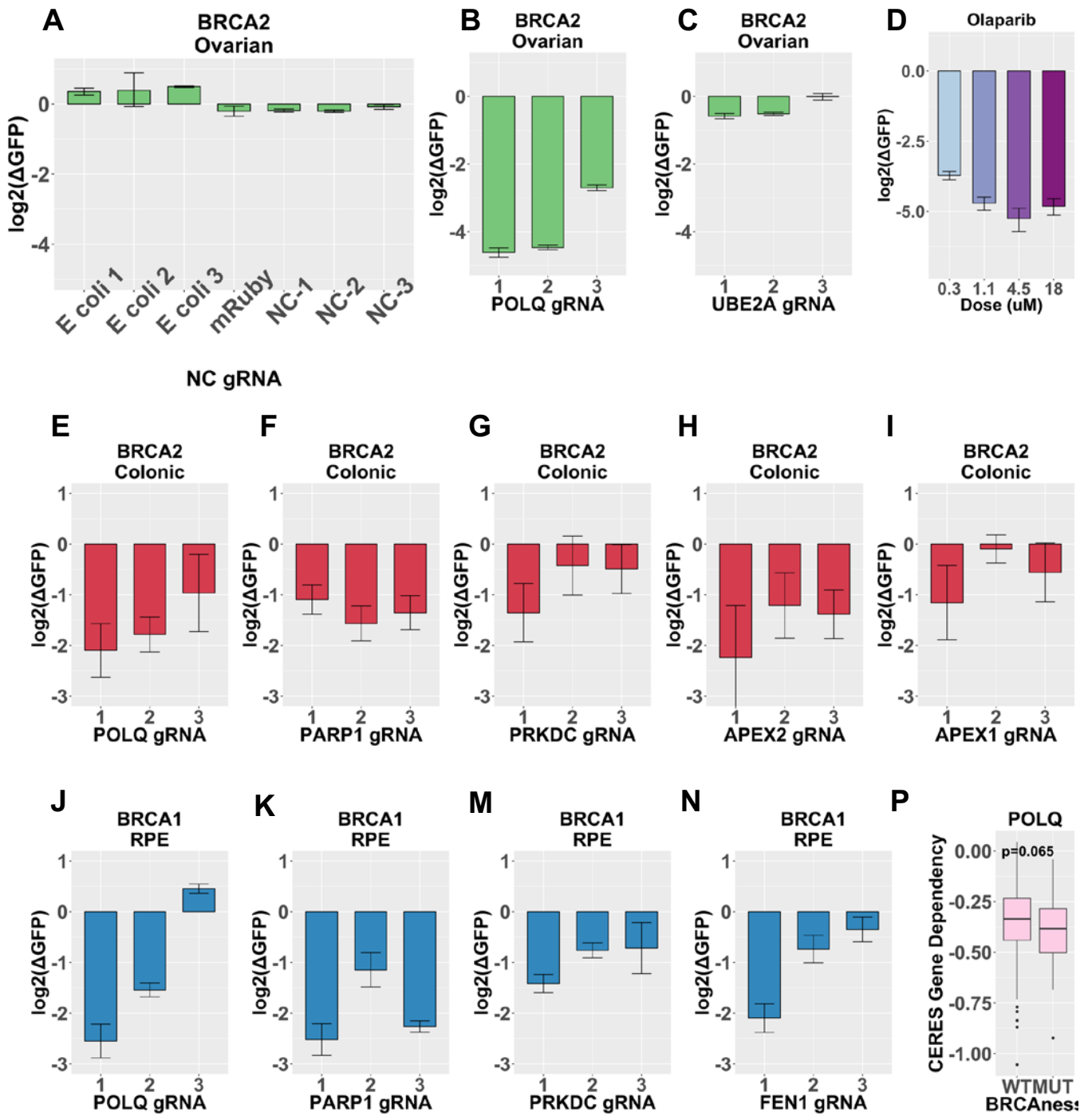
Supplemental Figure S2-3 (related to Figures 2-4 and 2-5) (Continued).
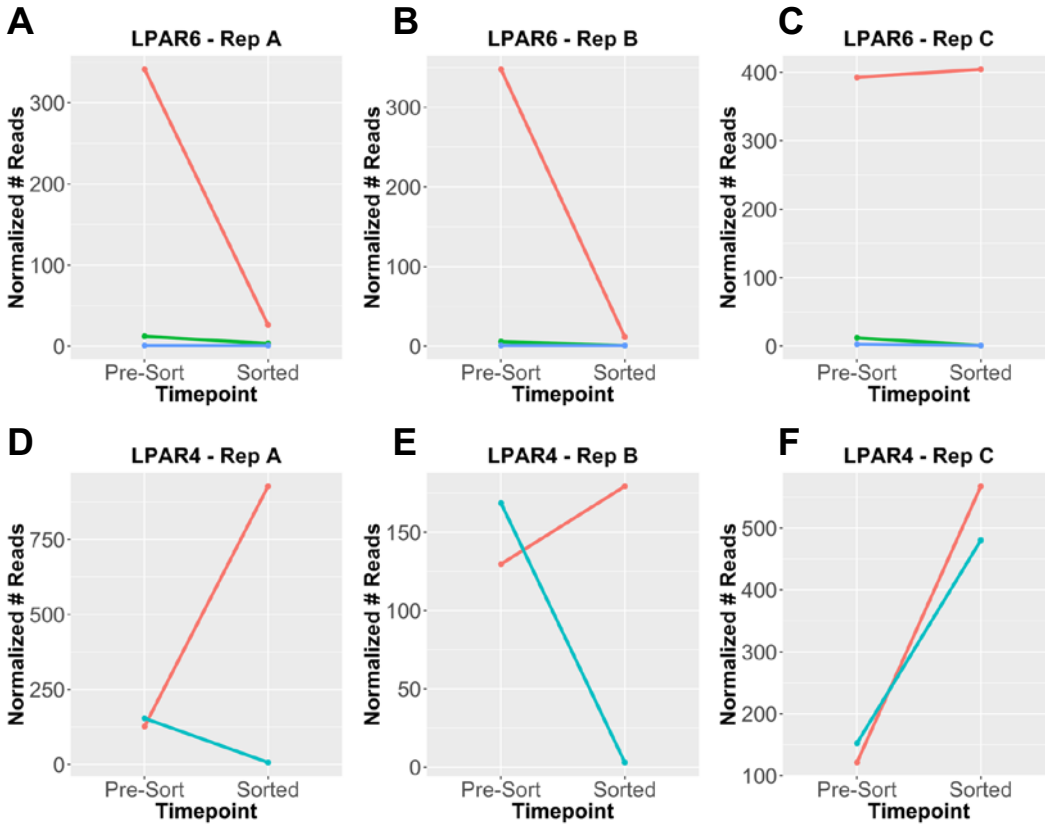
Supplemental Figure S3-1. Performance of barcodes for LPAR family members that did not score in the PD-L1 Enrichment screen. (A-C) Normalized read counts of barcodes corresponding to *LPAR6* from the pre-sorted unstained library population and the PD-L1[high] sorted cell population, for all three replicates. Each line represents an individual barcode. **(D-F)** Normalized read counts as described above, for barcodes corresponding to *LPAR4*.

Supplemental Figure S3-2. Ingenuity Pathway Analysis of the PD-L1 Induction Screen. (A) Genes that significantly enriched in the PD-L1 Induction Screen (with a p-value of <0.01, determined by edgeR) were submitted to Ingenuity for pathway analysis, after filtering out all type I interferons (which otherwise predominate the ranked list of pathways). The top pathways are listed, ordered and color-coded by the p-value returned by DAVID for their enrichment. The percent of genes in the pathway that scored in our screen is plotted on the x-axis. **(B)** A top network identified by Ingenuity as enriched in the PD-L1 induction screen. Red shading indicates genes that scored with a p-value <0.01 by edgeR in the original screen, and the legend identifies the type of protein in the network its shape.

A

Ingenuity Pathway Analysis

Pathway (y-axis):
- Role of Lipids-Lipid Rafts in the Pathogenesis of Influenza
- Role of Wnt-GSK-3b Signaling in the Pathogenesis of Influenza
- Human Embryonic Stem Cell Pluripotency
- Role of Cytokines in Mediating Communication between Immune Cells
- Retinoic acid Mediated Apoptosis Signaling
- Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses
- Role of JAK1, JAK2 and TYK2 in Interferon Signaling
- Role of PI3K-AKT Signaling in the Pathogenesis of Influenza
- Th1 and Th2 Activation Pathway
- Interferon Signaling
- ILK Signaling
- NADH Repair
- Osteoarthritis Pathway
- Role of Hypercytokinemia-hyperchemokinemia in the Pathogenesis of Influenza
- Catecholamine Biosynthesis
- RhoA Signaling
- Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis
- Tetrapyrrole Biosynthesis II
- HMGB1 Signaling
- Ga12/13 Signaling
- Th1 Pathway
- Colorectal Cancer Metastasis Signaling
- IL-12 Signaling and Production in Macrophages
- Molecular Mechanisms of Cancer
- Mitotic Roles of Polo-Like Kinase

x-axis: % of Genes in Pathway (0.0, 0.1, 0.2, 0.3)

-log p-value: 3.5, 3.0, 2.5, 2.0, 1.5

B

Legend:
- Complex
- Enzyme
- Ion Channel
- Kinase
- Phosphatase
- Transcription Regulator
- Translation Regulator
- Transporter
- Other
- — Relationship
- --- Relationship

Supplemental Figure S3-2 (Continued). Ingenuity Pathway Analysis of the PD-L1 Induction Screen.
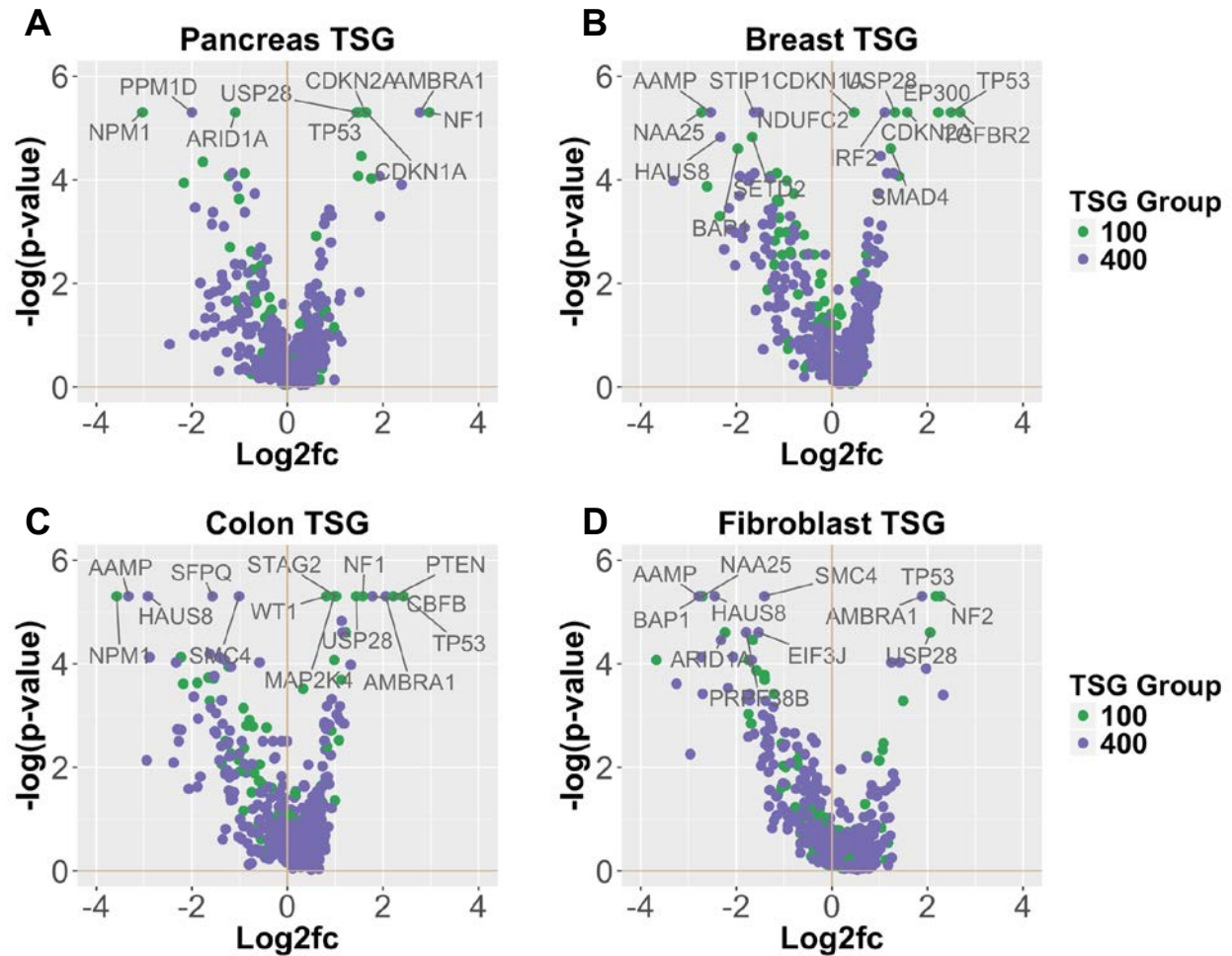
Supplemental Figure S4-1. Volcano Plots Summarizing TSG Proliferation Screens.
Volcano plots depicting MAGECK -$\log_{10}$(p-value) versus log2-fold-change are shown for
proliferation screens conduced with the CRISPR TSG library in immortalized cell lines of four
tissue types: pancreas **(A)**, breast **(B)**, colon **(C)**, and fetal lung fibroblasts **(D)**. Each dot
represents the performance of a single gene and is color-coded according to whether the gene was
originally ranked in the top 100 TSGs by TUSON or the next 400 TSGs (101-500) by TUSON.
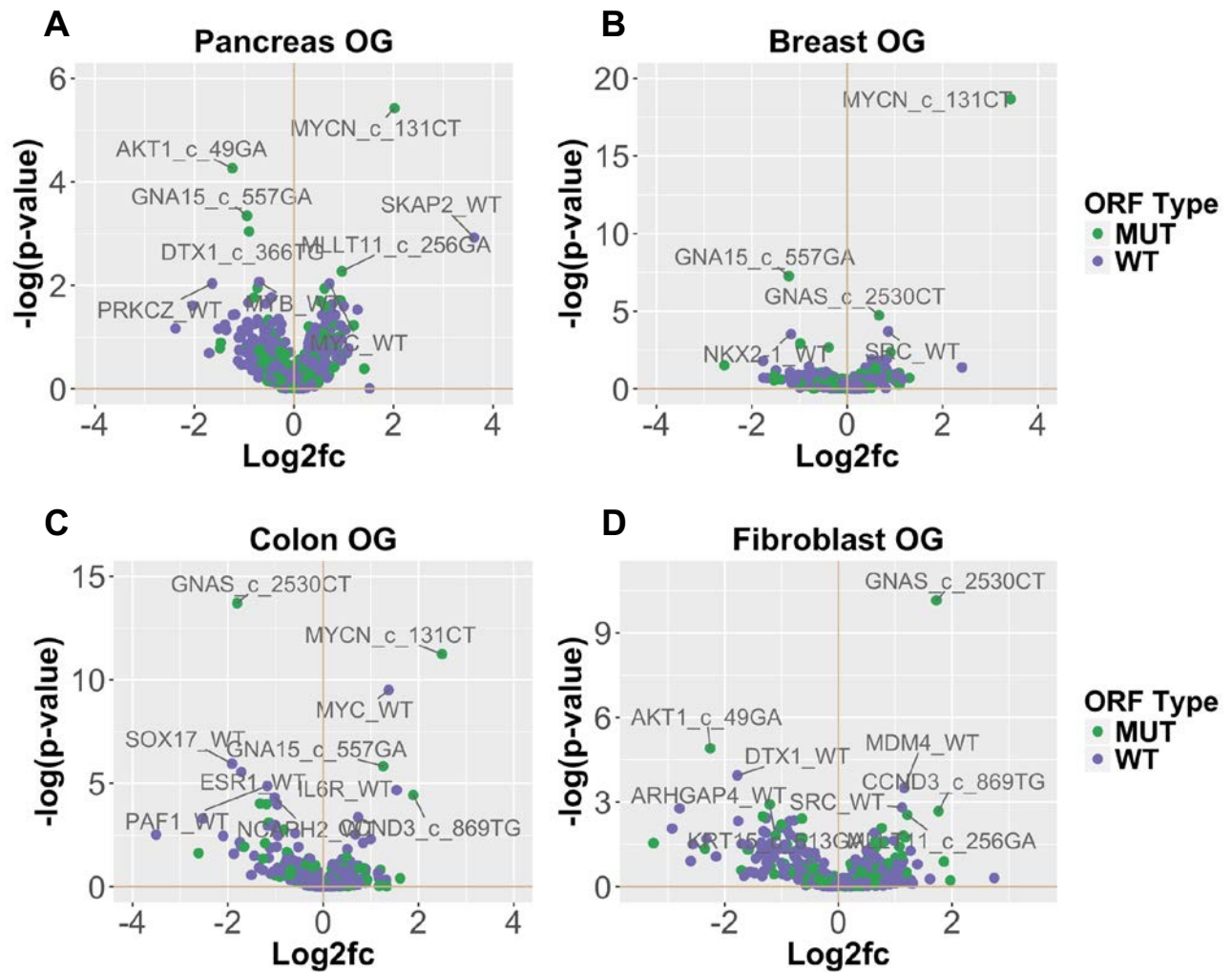
Supplemental Figure S4-2. Volcano Plots Summarizing OG Proliferation Screens. Volcano plots depicting edgeR -$\log_{10}$(p-value) versus log2-fold-change are shown for proliferation screens conduced with 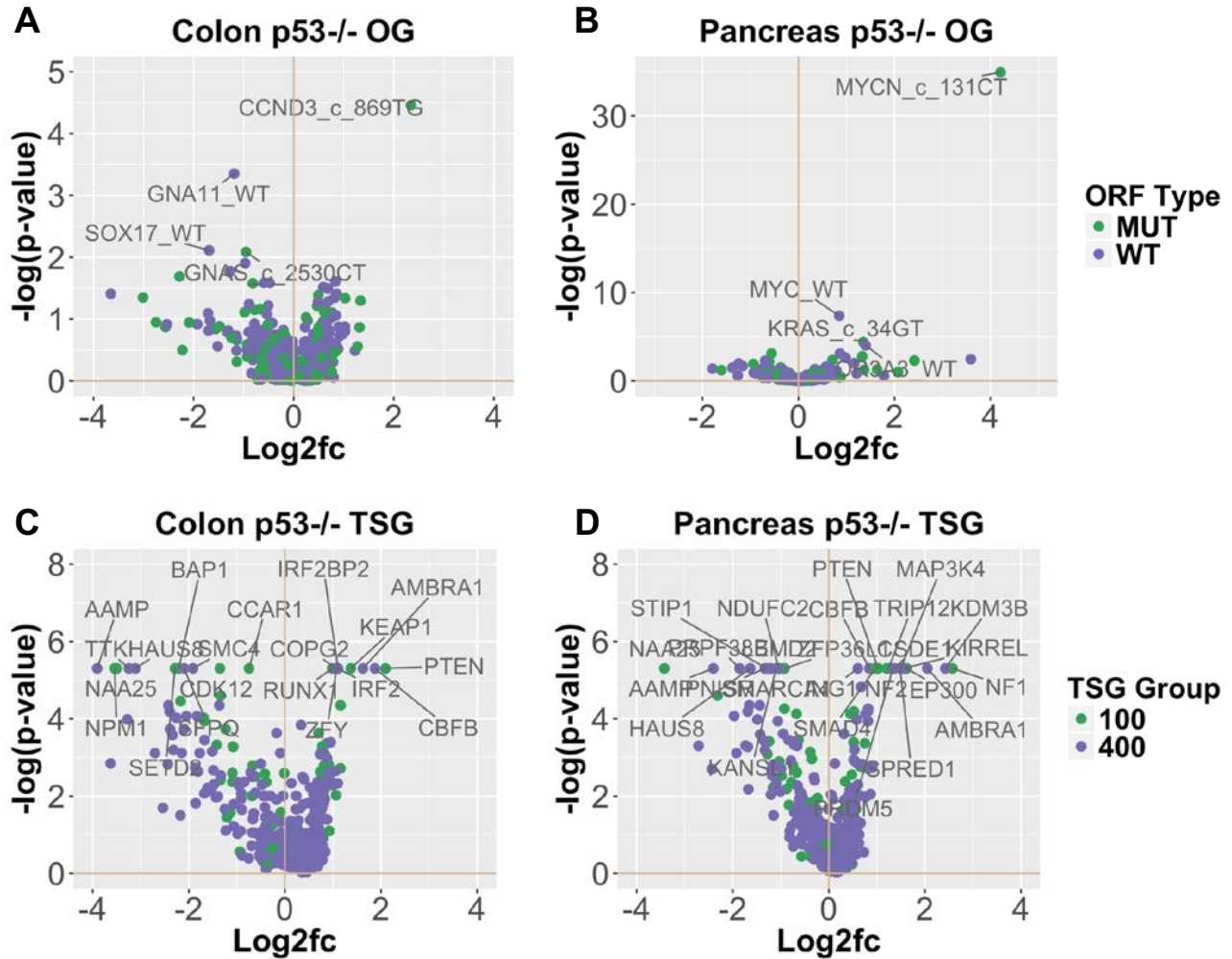the OG library in immortalized cell lines of four tissue types: pancreas **(A)**, breast **(B)**, colon **(C)**, and fetal lung fibroblasts **(D)**. Each dot represents the performance of a single gene and is color-coded according to whether the ORF was mutated by site-directed mutagenesis or remained wild-type in the library.

Supplemental Figure S4-3. Volcano Plots Summarizing p53 knockout proliferation Screens. Volcano plots depicting edgeR (A-B) or MAGeCK (C-D) -log$_{10}$(p-value) versus log2-fold-change are shown for proliferation screens conduced in p53 knockout cell lines with either the OG library (A-B) or CRISPR TSG library (C-D). p53 knockout lines were generated in both the colon **(A, C)** and pancreas **(B, D)** backgrounds. Each dot represents the performance of a single gene and is color-coded according to (A-B) whether ORFs were mutated by site-directed mutagenesis or (C-D) whether TSGs were among the top 100 TSGs originally predicted by TUSON Explorer.

OG and TSG Tissue-Specific Proliferation

filtered log2fc

Gene rows (top to bottom): MYCN_MUT, AMBRA1, NF1, TRAF3, CDKN1A, CCND3_MUT, RBM15, ASXL2, SPRED1, MYC_WT, KEAP1, KDM3B, KMT2B, IL6R_WT, TGFBR2, NEDD9_WT, MAP4K4, RHEB_WT, C15orf53_WT, GNG4_MUT, PPIL1_MUT, DKK2_MUT, MDM4_WT, MLLT11_MUT, RB1, WT1, PAWR, SRC_WT, MAP4K3, C15orf53_MUT, MYD88_MUT, SPRED2, AKT1_WT, DUSP16, PRDM5, MYCN_WT, HIST1H2BD_WT, RIT1_MUT, AQP2_MUT, BET1_WT, EBPL_MUT, TRIP12, ZFY, BCL10, DYRK1A, CDK6_WT, IL21_WT, PHF20L1, PSCA, MAP2K4, ACVR2A, OR3A3_WT, RHEB_MUT, CDK4_WT, RRAS2_WT, NRAS_MUT, RXRA_MUT, TXNRD1_MUT, MFAP5_MUT, IRF2, MAP3K4, PTEN, SKAP2_WT, TP53, CDKN2A, TAOK1, USP28, NF2, KIRREL

Columns: Breast, Colon, Pancreas, Fibroblast

Color scale: 3, 2, 1, 0, -1

Supplemental Figure S4-4. Combined TSG and OG proliferation drivers from cell lines of four different tissue types. TSGs and OGs with the most significant p-values in proliferation screens across four immortalized cell lines are depicted as a combined heatmap. The log2-fold-change of each gene is plotted by color-coding on the indicated color scale, for statistically significant genes only; log2-fold-changes that were not statistically significant are defaulted to 0 (white). Genes were hierarchically clustered using the complete linkage method.
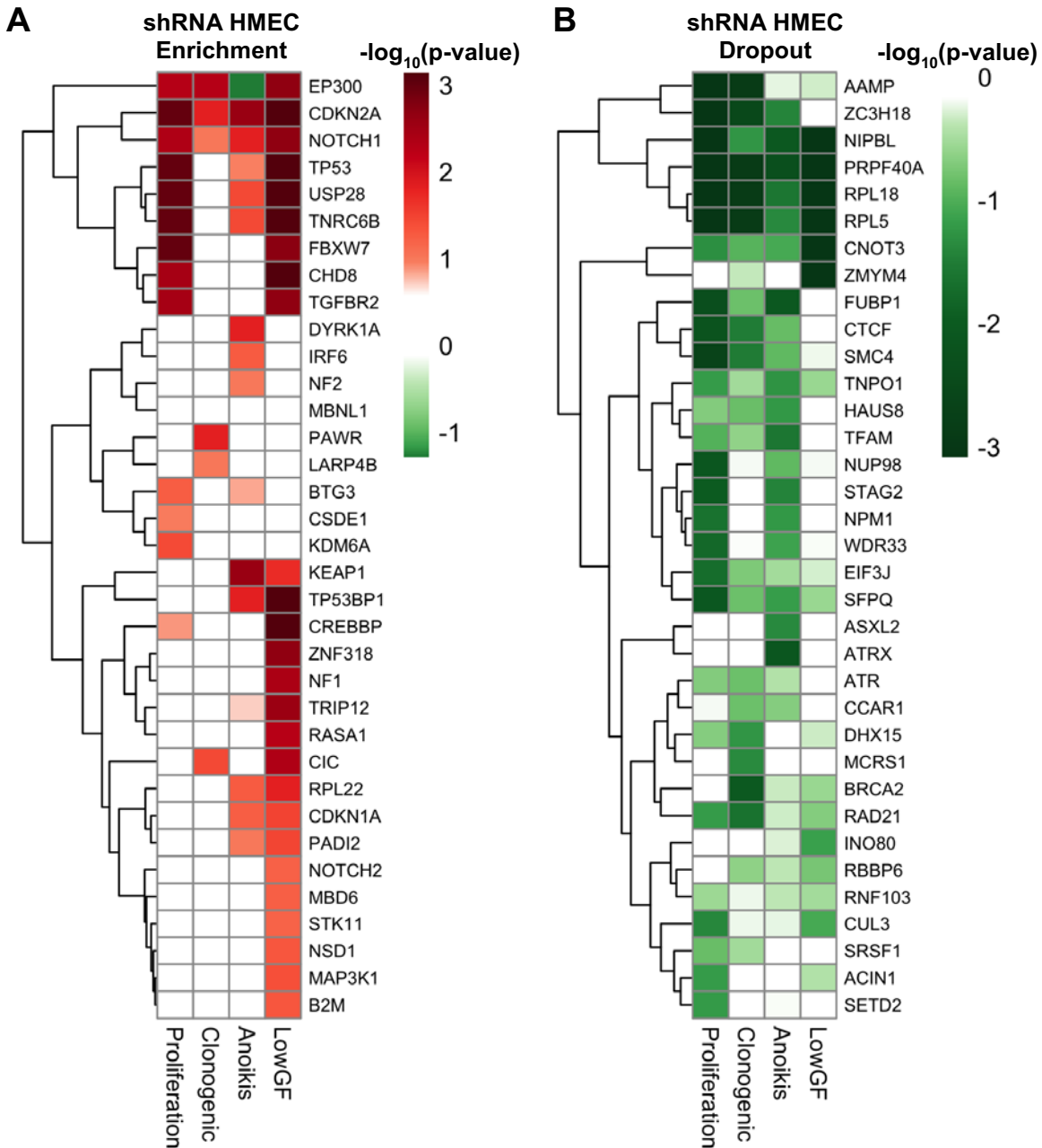
Supplemental Figure S4-5. Heatmaps Portraying the Performance of the shRNA TSG Library Across Four Different Growth Phenotypes. Heatmap depicting genes with the most significant p-values by MAGeCK for enrichment **(A)** or dropout **(B)** in HMEC CRISPR TSG screens for proliferation, clonogenic growth, anoikis bypass, and growth factor deprivation. The p-value of each gene is plotted by color-coding according to the adjacent color legend, on a -$\log_{10}$scale. To clarify directionality, this value is plotted as positive for genes that enriched in the screen (with a positive log2-fold change) and negative for genes that dropped out in the screen (with a negative log2-fold-change. Genes were hierarchically clustered using the complete linkage method.
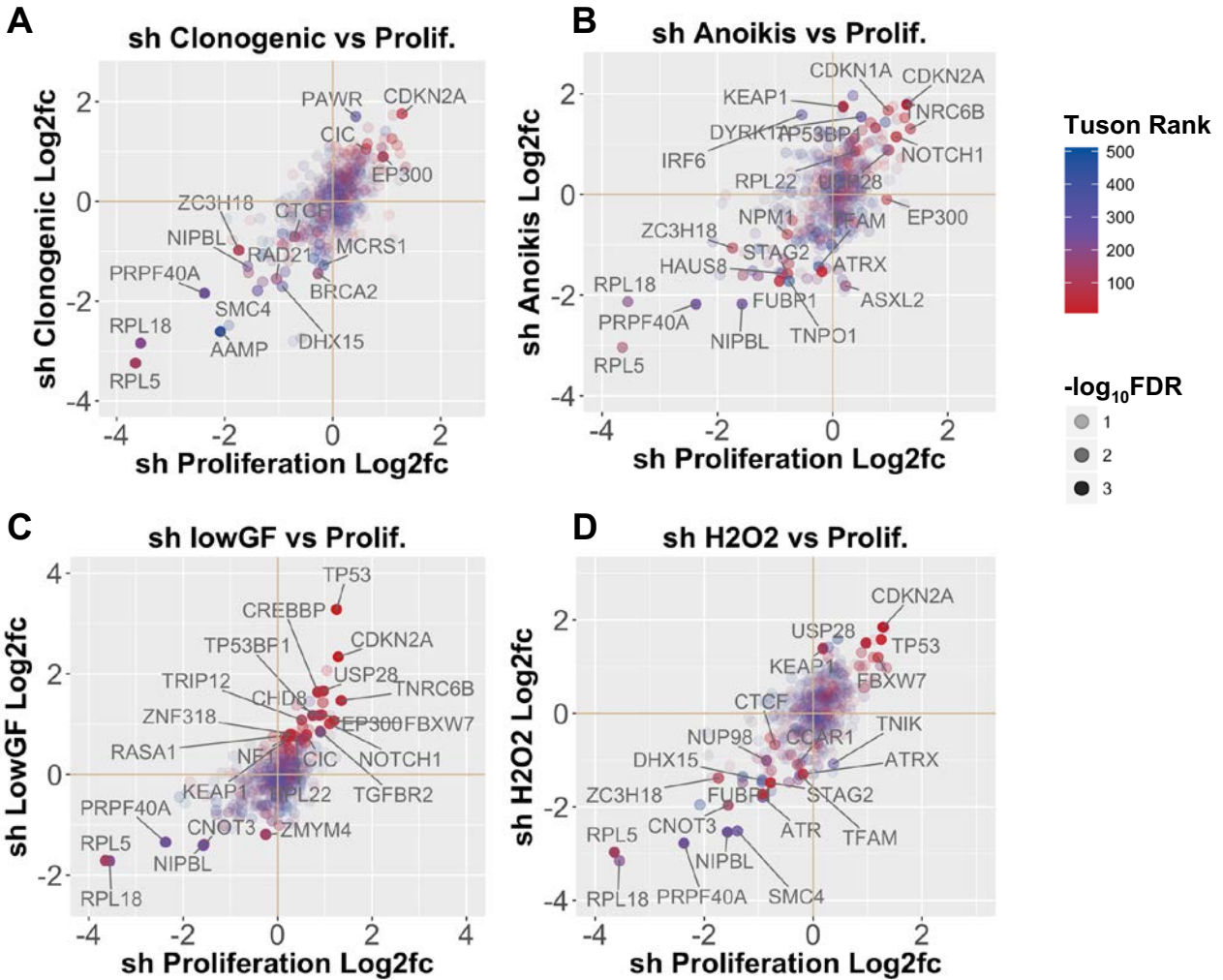
Supplemental Figure S4-6. Comparison of Gene Performance in shRNA HMEC Phenotypic Screens. Scatterplots comparing the log2-fold-change of each TSG in the shRNA TSG library in the HMEC clonogenic (**A**), anoikis bypass (**B**), low GF (**C**), or H2O2 bypass (**D**) screen versus its performance in HMEC proliferation. Each gene is represented by a single dot, color-coded by its rank in the original TUSON Explorer prediction list and opacified according to the significance of its performance as determined by MAGeCK.
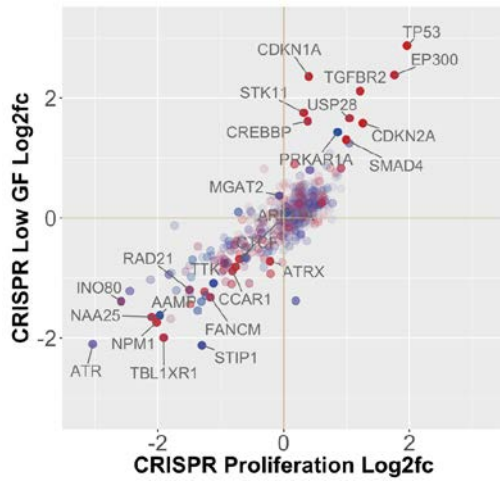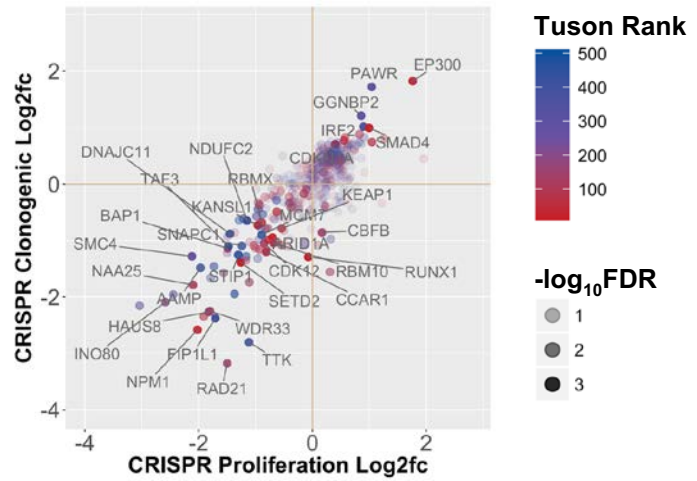
Supplemental Figure S4-7. Comparison of Gene Performance in CRISPR HMEC Phenotypic Screens. Scatterplots comparing the log2-fold-change of each TSG in the CRISPR TSG library in the low GF (**A**) and clonogenic (**B**) screen versus its performance in proliferation. Each gene is represented by a single dot, color-coded by its rank in the original TUSON Explorer prediction list and opacified according to the significance of its performance as determined by MAGeCK.

Appendix II. Supplemental Tables

Supplemental Table S3-1. Complete List of PD-L1 Induction Screen Hits. All genes achieving a p-value <0.01 are listed, along with their rank by edgeR, their p-value as determined by edgeR, and their full gene name.

Rank	Gene	P-Value	Gene Name
1	IFNA14	2.04E-09	interferon alpha 14
2	IFNA17	1.19E-07	interferon alpha 17
3	MRGPRX4	2.23E-07	MAS related GPR family member X4
4	IFNA6	3.00E-07	interferon alpha 6
5	IFNA4	3.86E-07	interferon alpha 4
6	SFXN5	9.18E-06	sideroflexin 5
7	IFNB1	1.32E-05	interferon beta 1
8	IFNA10	2.83E-05	interferon alpha 10
9	CELF3	3.37E-05	CUGBP Elav-like family member 3
10	SERBP1	5.67E-05	SERPINE1 mRNA binding protein 1
11	LPAR1	7.64E-05	lysophosphatidic acid receptor 1
12	CD274	8.95E-05	CD274 molecule
13	CWF19L1	1.31E-04	CWF19 like 1, cell cycle control (S. pombe)
14	IFNA8	1.43E-04	interferon alpha 8
15	S1PR2	2.40E-04	sphingosine-1-phosphate receptor 2
16	CELF5	2.44E-04	CUGBP Elav-like family member 5
17	BHLHA15	3.17E-04	basic helix-loop-helix family member a15
18	CRTAC1	3.32E-04	cartilage acidic protein 1
19	LPAR2	4.00E-04	lysophosphatidic acid receptor 2
20	ZIC3	4.61E-04	Zic family member 3
21	PPP4R2	4.77E-04	protein phosphatase 4 regulatory subunit 2
22	IFNA2	4.88E-04	interferon alpha 2
23	FCGR2A	5.28E-04	Fc fragment of IgG receptor IIa
24	LPAR5	5.57E-04	lysophosphatidic acid receptor 5
25	IFNA1	7.83E-04	interferon alpha 1
26	ANXA2R	9.31E-04	annexin A2 receptor
27	POLR3D	9.50E-04	RNA polymerase III subunit D
28	GPR107	9.51E-04	G protein-coupled receptor 107
29	IFNA5	1.24E-03	interferon alpha 5
30	EIF1AY	1.36E-03	eukaryotic translation initiation factor 1A, Y-linked
31	C15orf59	1.39E-03	chromosome 15 open reading frame 59
32	C9orf24	1.59E-03	chromosome 9 open reading frame 24
33	CLDN14	1.73E-03	claudin 14
34	ILF2	1.88E-03	interleukin enhancer binding factor 2
35	SERPINA4	1.91E-03	serpin family A member 4
36	MRPL47	1.98E-03	mitochondrial ribosomal protein L47

37	BZW2	2.02E-03	basic leucine zipper and W2 domains 2
38	TNRC6A	2.16E-03	trinucleotide repeat containing 6A
39	PXK	2.26E-03	PX domain containing serine/threonine kinase like
40	CTDSP2	2.31E-03	CTD small phosphatase 2
41	CLEC18A	2.35E-03	C-type lectin domain family 18 member A
42	ARHGEF40	2.61E-03	Rho guanine nucleotide exchange factor 40
43	FLYWCH2	2.68E-03	FLYWCH family member 2
44	VIM	2.71E-03	vimentin
45	EDIL3	2.73E-03	EGF like repeats and discoidin domains 3
46	P2RY8	2.86E-03	P2Y receptor family member 8
47	EEF1G	3.05E-03	eukaryotic translation elongation factor 1 gamma
48	PPP2R2C	3.18E-03	protein phosphatase 2 regulatory subunit B gamma
49	PIGO	3.44E-03	phosphatidylinositol glycan anchor biosynthesis class O
50	KIAA1456	3.53E-03	tRNA methyltransferase 9B (putative)
51	DBH	3.78E-03	dopamine beta-hydroxylase
52	MAGED4B	3.78E-03	MAGE family member D4B
53	PGK2	3.98E-03	phosphoglycerate kinase 2
54	IL22RA1	4.02E-03	interleukin 22 receptor subunit alpha 1
55	BGN	4.05E-03	biglycan
56	RHOC	4.34E-03	ras homolog family member C
57	IFNW1	4.34E-03	interferon omega 1
58	IFNG	4.38E-03	interferon gamma
59	MRPS15	4.48E-03	mitochondrial ribosomal protein S15
60	PPAN	4.50E-03	peter pan homolog (Drosophila)
61	BMP4	4.72E-03	bone morphogenetic protein 4
62	RFPL3	4.79E-03	ret finger protein like 3
63	INSL3	4.93E-03	insulin like 3
64	DRD3	4.97E-03	dopamine receptor D3
65	RBM42	5.01E-03	RNA binding motif protein 42
66	SCNN1B	5.03E-03	sodium channel epithelial 1 beta subunit
67	RPS10P7	5.08E-03	ribosomal protein S10 pseudogene 7
68	SYT17	5.11E-03	synaptotagmin 17
69	ADCK1	5.19E-03	aarF domain containing kinase 1
70	USE1	5.25E-03	ubiquitin conjugating enzyme E2 Z
71	SYTL1	5.32E-03	synaptotagmin like 1
72	SLC25A10	5.44E-03	solute carrier family 25 member 10
73	LHX4	5.48E-03	LIM homeobox 4
74	DMTN	5.50E-03	dematin actin binding protein
75	HHAT	5.53E-03	hedgehog acyltransferase
76	UNC5CL	5.71E-03	unc-5 family C-terminal like
77	ATF4	5.80E-03	activating transcription factor 4
78	C1orf74	6.00E-03	chromosome 1 open reading frame 74
79	HMBS	6.32E-03	hydroxymethylbilane synthase
80	ZNF513	6.44E-03	zinc finger protein 513

81	TBC1D3C	6.68E-03	TBC1 domain family member 3C
82	ZNRF1	6.69E-03	zinc and ring finger 1
83	GPR68	6.74E-03	G protein-coupled receptor 68
84	GPBP1L1	6.99E-03	GC-rich promoter binding protein 1 like 1
85	RPL3	7.11E-03	ribosomal protein L3
86	OR2T33	7.13E-03	olfactory receptor family 2 subfamily T member 33
87	GTDC2	7.18E-03	protein O-linked mannose N-acetylglucosaminyltransferase 2 (beta 1,4-)
88	TUBGCP4	7.23E-03	tubulin gamma complex associated protein 4
89	RBMS1	7.36E-03	RNA binding motif single stranded interacting protein 1
90	NAT16	7.46E-03	N-acetyltransferase 16 (putative)
91	PLEKHA8P1	7.55E-03	pleckstrin homology domain containing A8 pseudogene 1
92	MRPL41	7.65E-03	mitochondrial ribosomal protein L41
93	WNT3A	7.69E-03	Wnt family member 3A
94	SPSB4	7.71E-03	splA/ryanodine receptor domain and SOCS box containing 4
95	DPP7	7.77E-03	dipeptidyl peptidase 7
96	CARKD	8.02E-03	NAD(P)HX dehydratase
97	CYP2A7	8.12E-03	cytochrome P450 family 2 subfamily A member 7
98	FUT1	8.29E-03	fucosyltransferase 1 (H blood group)
99	IFNA13	8.30E-03	interferon alpha 13
100	RASL12	8.37E-03	RAS like family 12
101	FAM96A	8.50E-03	family with sequence similarity 96 member A
102	GATA2	8.58E-03	GATA binding protein 2
103	BPHL	8.59E-03	biphenyl hydrolase like
104	ZNF581	8.62E-03	zinc finger protein 581
105	DPY19L2P2	8.68E-03	DPY19L2 pseudogene 2
106	HIST1H2BA	8.77E-03	histone cluster 1 H2B family member a
107	FGG	8.78E-03	fibrinogen gamma chain
108	KRTAP22-1	8.93E-03	keratin associated protein 22-1
109	HEXIM2	9.05E-03	hexamethylene bisacetamide inducible 2
110	TGFB1	9.39E-03	transforming growth factor beta 1
111	GPR152	9.67E-03	G protein-coupled receptor 152
112	TEX28P1	9.68E-03	testis expressed 28 pseudogene 1
113	CMTM4	9.78E-03	CKLF like MARVEL transmembrane domain containing 4
114	TLE6	9.86E-03	transducin like enhancer of split 6