Investigations of Loss of Heterozygosity–Associated Dependencies in Cancer

Investigations of Loss of Heterozygosity–Associated Dependencies in Cancer

A dissertation presented

by

Caitlin Adair Nichols

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

April 2019

Dissertation Advisor: Rameen Beroukhim Caitlin Adair Nichols

Investigations of Loss of Heterozygosity–Associated Dependencies in Cancer

Abstract

Alterations in non-driver genes represent an emerging class of potential therapeutic targets in cancer. Most tumors undergo loss of heterozygosity (LOH) of hundreds to thousands of non-driver genes, generating discrete differences between tumor and normal cells. However, only a small body of research examines how these differences can be used as the basis for cancer therapies. To address this issue, we interrogated LOH of polymorphisms in essential genes as a novel class of therapeutic targets. We hypothesized that monoallelic inactivation of the allele retained in tumors can selectively kill cancer cells but not somatic cells, which retain both alleles. We term this class GEMINI vulnerabilities after the twins from Greek mythology, Castor and Pollux, one of whom was mortal and other immortal. To study these vulnerabilities, we first performed a large-scale analysis of germline variant, gene essentiality, and LOH data to identify potential GEMINI vulnerabilities. This analysis identified 5664 variants in 1278 essential genes that undergo LOH in cancer. Second, we performed proof-of-concept validation of GEMINI vulnerabilities in two essential genes, *PRIM1*and *EXOSC8*. We demonstrated that allele-specific inactivation of either of these genes reduced growth of cells harboring the targeted allele, while cells harboring the non-targeted allele remain intact. Finally, we evaluated the potential for each GEMINI variant to be targeted using allele-specific gene-editing, RNAi, or small-molecule approaches. These analyses indicated that not all GEMINI vulnerabilities may be ideal candidates for RNAi-based therapeutics. Nonetheless, we identified several promising GEMINI vulnerabilities for future efforts to develop allele-specific small-molecule inhibitors. This work

rigorously validates GEMINI vulnerabilities as a potential class of therapeutic targets and defines its potential scope. These findings provide a strong starting point for future development of therapeutic agents targeting GEMINI vulnerabilities.

Table of Contents

This page intentionally left blank.

List of Figures

This page intentionally left blank.

Acknowledgements

I would like to thank those who have helped me through this process—you have encouraged me, enlightened me, and helped me to become a better scientist and person.

First, to Rameen Beroukhim. Thank you for the opportunity to work in your lab. You have built a community of people who are not only smart and hardworking, but also kind and collaborative, and the atmosphere in the Beroukhim lab is a tribute to you as a person and a scientist. It's been a privilege to work with people I like and respect, especially during the inevitable difficult times of graduate school. I am also grateful for your dedicated mentorship and reliable sense of humor. The rigor of the scientific training you have provided is a gift that will continue to bless my life and hopefully the lives of others as I apply what I have learned. Thank you for supporting me when I needed it the most.

To my former postdoctoral mentor, Brenton Paolella—the 'good cop' to my 'bad cop' of data pessimism. You have been my day-to-day mentor, emotional support, and friend. Thank you for all of your help in getting this to the finish line, even after you left the lab. This thesis would not have been possible without you.

A big thanks to Beroukhim, Bandopadhayay, and Stiles lab members for all the personal and professional support you have provided for the last five years. In particular, this work would not be possible without the computational expertise of Will Gibson. I've also had the privilege of working with several talented and hardworking technicians—Laura Urbanski, John Busanovich, Naomi Currimjee, and especially Meredith Brown. Thanks to all of you for keeping the lab running (not a small feat) and helping me multiply my efforts at the bench.

I would also like to express special appreciation to Sirano Dhe-Paganon and Jack Kosmicki. Sirano, your structural biology expertise has been invaluable to this work, and your

excitement about the project has helped all of us believe in its potential. And Jack, while I'm grateful for all your ExAC knowledge, your most important contribution has been as my friend. Thanks for being there through all the ups and downs of the last six years.

To my dissertation committee members, Benjamin Ebert, David Weinstock, and Matthew Meyerson—thank you for your feedback and encouragement. In particular, thank you, Matthew, for your refreshing honesty about your graduate school experience and your continued confidence in my work and my abilities. Your expressions of support have meant a great deal.

Steve Sowa, thank you for your invaluable support with the template and formatting for this document.

To my Strathmore family—thank you for being there through the highs and lows of the last five years.

Aaron Nichols, thank you for being my mentor and editor-in-chief.

To Alexander Gimelbrant, Florian Muller, and David Barbie—thank you for your time and consideration in serving as exam committee members.

Finally, to the unnamed. Thank you for each time you taught, helped, encouraged, and listened.

Attributions

William J. Gibson: generated essential gene list and wrote associated methods (Chapter 2); analyzed CCLE data to identify cell line models (Chapter 3)

Meredith S. Brown, John P. Busanovich, Hope Wei, Laura M. Urbanski, Naomi Curimjee: assisted with cell culture, cell growth assays, and molecular biology experiments (PCR, immunoblotting) (Chapter 3)

Jack A. Kosmicki: provided assistance with ExAC variant list (Chapter 2)

Ashton C. Berger, Galen F. Gao, Andrew D. Cherniack: generated LOH calls (Chapter 2)

Sirano Dhe-Paganon: performed p-blast of targets against structures in PDB and visually scored hits; wrote associated methods (Chapter 4)

Brenton R. Paolella: provided some code for genomic analyses and wrote associated figure legends (Chapter 2); provided conceptual and technical guidance; assisted with dissertation review

Rameen Beroukhim: provided conceptual and technical guidance; assisted with dissertation drafting and review

Some passages have been quoted verbatim from the bioRxiv manuscript "Loss of heterozygosity of essential genes represents a widespread class of potential cancer vulnerabilities" (© 2019 Caitlin A. Nichols)[1].

This page intentionally left blank.

CHAPTER 1: INTRODUCTION

This page intentionally left blank.

CHAPTER 1: INTRODUCTION

Cancer as a disease derives from the uncontrolled growth and spread of abnormal cells. Cancerous cells are marked by several features, including unlimited proliferative capacity, dysregulated cell growth and division, and escape from programmed cell death[2].

Prior to the 20th century, cancer therapy had remained much the same for millennia. In antiquity, tumors were treated by surgery or the application of various (largely ineffective) natural remedies (e.g., bacterial toxins, animal organ extracts, starvation). The use of surgery and other attempted therapies continued largely unchallenged into modernity, until the discovery of x-rays led to the introduction of radiation therapy in the early 20th century. This breakthrough resulted in significant improvement in clinical outcomes, especially when coupled with chemotherapy[3]. The use of potassium arsenite to treat chronic myelogenous leukemia in 1865 represented the first effective chemotherapeutic, yet the widespread use of chemotherapy did not begin until after World Wars I and II, with the discoveries that mustard gas (sulfur mustard) and its chemical analogs exhibited antitumor activity[4]. While these and later chemotherapeutics have clearly benefited patients, the irony that chemical weapons for a time represented the most promising cancer therapies cannot be overstated.

Breakthroughs in understanding the molecular underpinnings of cancer would shift the approach to cancer therapeutics. The discovery that cancers are caused by genetic alterations of proto-oncogenes[5,6] and tumor suppressor genes[7,8] in the 1970s and 1980s led to an intense hunt for these initiating events and drugs that could target them. For the first time, rationally targeting tumors based on discrete differences from normal tissue—the molecular drivers of oncogenesis—became possible. The completion of the

Human Genome Project in 2003 and subsequent large-scale cancer sequencing efforts have further increased the rate of discovery of tumor-initiating events[9].

Yet as the cancer genome has been characterized with greater and greater granularity, there remain relatively few highly selective therapies in clinical use. In practice, identifying drivers of oncogenesis has proved perhaps the easiest task—with 'undruggable' targets, tumor heterogeneity, drug resistance, and other challenges complicating what initially appeared to be a clear path to therapeutic success. Since targeting solely cancer-causing alterations has proven insufficient, new strategies are required to target and eliminate malignant cells. This thesis explores one such alternate strategy: leveraging the widespread loss of genetic diversity in cell-essential genes that occurs as a 'collateral damage of normal oncogenic processes.

Cancer as a genomic disease

Cancers arise through a process of Darwinian evolution acting on a single genomically altered cell[10,11]. Precancerous cells gain random genetic or epigenetic alterations, which may exert beneficial, neutral, or detrimental effects on cell proliferation and survival. Beneficial alterations may lead to the expansion a particular cellular clone through processes of natural (e.g., hypoxia, host immune response) or artificial selection (e.g., therapy)[12]. On the other hand, clones containing beneficial, neutral, or detrimental alterations may remain stable in the growing tumor, or they may be lost or expand through genetic drift[10,11,13]. Alterations that confer a substantial growth or survival advantage to a cell are known as driver alterations, and genes that harbor these changes are known as driver genes. In contrast, alterations that have neutral or negative effects on cancer cell growth are known as passenger alterations. Both driver and

passenger events can lead to alteration of non-driver genes, also termed bystander genes[14].

The interplay of random mutation, genetic drift, and selection can generate significant intratumoral heterogeneity[10], with a single neoplasm potentially containing multiple evolutionarily divergent clones[11]. Multiple-sample sequencing studies across tumor types have identified a wide range in the number of coding mutations (between 0 and 8000) that varied between different regions of the same primary tumor or between primary and metastatic/recurrent lesions[15,16]. This intratumoral genetic diversity can provide the basis for drug resistance. For example, in non-small cell lung cancer, preexisting subclonal mutations have been shown to underlie the evolution of resistance to rociletinib, a third-generation EGFR inhibitor targeting the EGFR T790M mutation[12,17]. Similarly, analyses of circulating tumor DNA identified up to 12 distinct resistant subclones in patients whose colorectal cancers had stopped responding to anti-EGFR therapy[12,18].

Driver genes

Cancer cells derive their growth advantage from driver alterations that improve fitness. These alterations affect function of proto-oncogenes and tumor suppressor genes, often called ‘driver genes’. Oncogenes arise from proto-oncogenes, genes responsible for spurring normal processes of cell growth; the increased activity or expression of these genes contributes to transformation[19]. In contrast, tumor-suppressor genes (TSGs) oppose the process of tumorigenesis by negatively regulating cell growth, division, and survival. Functions of TSGs include maintaining genomic integrity through DNA repair and triggering cell cycle arrest or apoptosis if DNA damage remains unresolved[19]. The loss of

TSG function releases cells from these controls on cell growth and survival. Unlike proto-oncogenes, which typically gain oncogenic potential through a single activating alteration, TSGs generally require biallelic inactivation in order to contribute to oncogenesis[19,20].

Somatic copy-number alterations

Driver genes can be activated or inactivated by point mutations[21–25]; small insertions or deletions[26,27]; larger rearrangements[28–31]; introduction of exogenous DNA[32,33]; or changes in copy number (amplifications resulting in more than two copies or deletions resulting in fewer copies)[34–37]. Among these alterations, somatic copy-number alterations (SCNAs) affect more genes than any other class of genetic alteration in cancer[38–40]. In the case of genomic deletions, most tumor types undergo loss of at least 11% of their genomes, with the average gene being deleted from 16% of cancers[40].

Frequent SCNAs observed in human tumors include amplifications of oncogenes (e.g., *EGFR*, *ERBB2*, *MYC*) and deletion of TSGs (e.g., *CDKN2A/B*, *PTEN*, *RB1*)[39,40]. Deletion of the TSG *TP53* represents one of the most common alterations in cancer, with losses observed in a fifth of ovarian and renal carcinomas; a quarter of breast, gastric, and liver carcinomas; and a third of small-cell lung carcinomas[41]. SCNAs affecting driver genes can also contribute to oncogenesis in a lineage-specific manner; for example, a vast majority (85–98%) of renal clear cell carcinomas exhibit deletion of the TSG *VHL*[42]. SCNAs occurring as passenger events can also affect large portions of the genome.

SCNAs can be small but are often quite large. Small (focal) events occur at a rate inversely proportional to their size[39]. In contrast, arm-level alterations affecting entire chromosomes or chromosome arms account for 60% of deleted regions and can affect tens

to over 1000 genes in a single event[43]. For instance, the most frequently lost chromosome arms, 17p and 8p, are deleted in 35% and 33% of cancers, respectively[44], with 17p containing nearly 500 genes and 8p containing ~400 genes[39]. The high frequency of such large SCNAs means that a majority of genes affected by copy-number changes represent non-driver (bystander) genes[45].

SCNAs arise from dysfunctional processes of genome maintenance, including those involved in proper mitosis and DNA replication and repair[13]. Chromosome-level SCNAs can result from chromosome missegregation caused by centrosomal aberrations[46] or failure of the spindle assembly checkpoint (SAC)[47]. Failures in DNA damage repair can also lead to structural alterations in cancer genomes. Both normal and oncogene-induced replicative stress can cause collapsed replication forks and double-strand breaks (DSBs), which when addressed through error-prone repair mechanisms can generate structural alterations including rearrangements, amplifications, and deletions[48].

SCNAs can also be caused by instances of catastrophic DNA breakage and repair, such as chromothripsis ('chromosome shattering') or chromoplexy[49]. While the precise mechanisms behind these complex structural alterations remain unclear, chromothripsis is thought to arise from one of two causes[49]: breakage-fusion-bridge cycles acting on dicentric chromosomes formed during telomere crisis[50,51], or missegregated chromosomes sequestered in micronuclei and subsequently subjected to extensive mutation, breakage, and error-prone repair[52,53].

Whole genome doubling (WGD) events generate SCNAs on a large scale. WGD can be caused by cell-cell fusions or cytokinesis failure (e.g., due to blockage of the cytokinesis furrow by a chromatin bridge). WGD can also arise when cells successfully

proceed through S phase but fail to enter or complete mitosis, a process known as mitotic slippage[54]. Cells that have undergone WGD tend to accumulate additional SCNAs relative to non-WGD cells[44,54]. In particular, WGD is associated with an increased frequency of gene deletions, perhaps as a result of decreased selective pressure provided by additional gene copies[40].

Therapeutics targeting SCNAs of driver genes have been introduced to the clinic, with varying success. Trastuzumab, a monoclonal antibody targeting the receptor tyrosine kinase HER2[55], has improved outcomes for patients with HER2-overexpressing breast cancers (30% of which harbor amplifications in the HER2 gene *ERBB2*[56]) and is now considered the standard therapy for the disease[57]. However, resistance to trastuzumab limits its effectiveness, with only 35% of patients responding to initial treatment and 70% of initial responders experiencing disease progression within a year. Trastuzumab has also shown limited efficacy against brain metastases as well as the potential for substantial cardiotoxicity[57]. *EGFR*-amplified tumors have also been targeted by EGFR inhibitors (e.g., the small-molecule tyrosine-kinase inhibitors erlotinib and gefitinib in non–small-cell lung cancer and lapatinib in HER2-amplified metastatic breast cancer; the monoclonal antibody nimotuzumab in glioblastoma). Yet like trastuzumab, EGFR inhibitors are also subject to inherent and required resistance[58].

Aneuploidy—a state of deviation from the normal complement of chromosomes[44]—may itself be therapeutically targetable. Genomic gains can lead to increased protein production, causing a strain on cellular protein-folding and -degradation machinery. The resulting proteotoxic and energetic stress has been shown in some instances to uniquely sensitize aneuploid cells to drugs that further augment these stresses,

generating a potential therapeutic window between tumor and normal cells[59,60]. Aneuploidy-induced stress may also be targetable by a class of collateral lethalities known as CYCLOPS genes[43], which will be described below. This thesis explores an alternative to these methods of targeting aneuploidy, based upon loss of heterozygosity (LOH).

Loss of heterozygosity (LOH)

LOH occurs when a cancer cell originally heterozygous for a particular genetic variant loses one of the two copies of the surrounding genomic locus[34]. Because deletion events are frequent and often quite large, any given genomic locus will often experience such losses; thus, LOH represents a common mechanism of generating one of the 'hits' in the two-hit inactivation of a TSG[61]. For instance, LOH was first identified in cases of familial retinoblastoma, where individuals who inherited one inactivated and one wild-type allele of *RB1* developed the disease after loss of the wild-type allele[34]. Additional studies of LOH events in familial cancers identified loss of TSGs including *WT1*[62], *NF1*[63], *APC*[64], and *PTEN*[35] as initiating events in transformation[14]. LOH events have also been described across a wide variety of sporadic cancers, including some for which the LOH event represents a distinguishing driver alteration[14,65–67]. However, LOH events can also occur as a result of general genomic instability (passenger events).

LOH can take two forms: simple loss of a single allele (copy-loss LOH) or loss of one allele accompanied by duplication of the remaining allele (copy-neutral LOH)[68]. Copy-loss LOH may be generated through unbalanced translocations, arm-level or focal deletions, or whole chromosomal loss; copy-neutral LOH may arise through gene conversion, mitotic recombination between homologous chromosomes, or deletion of one allele paired with an amplification of the other allele[34,69–73].

Both copy-loss and copy-neutral LOH represent a loss of genetic diversity within a tumor. Cancer cells that undergo LOH of a gene are left relying on solely one allele for any necessary gene products, in contrast to normal cells, which retain both alleles. This thesis explores an approach to take advantage of this allelic imbalance.

Targeted therapeutics

Direct targeting of driver genes

Toward the end of the 20th century and beginning of the 21st, an increasing number of therapies were discovered through rational design as *targeted* therapies. In contrast to chemotherapy, these therapies target unique molecular characteristics of tumor cells to obtain efficacy and specificity. The first such targeted therapy was tamoxifen, a failed contraceptive which was found to antagonize estrogen binding to the estrogen receptor (ER) and was subsequently used to treat ER-positive breast cancers[74,75]. Tamoxifen is thought to be effective in treating these tumors because it decreases proliferative signaling downstream of ER[76].

Most efforts toward development of targeted therapies have attempted to reverse the oncogenic effects of driver events, either through targeting oncogenes activated by these events or targeting their downstream pathways. The first example of this approach was the use of all-trans-retinoic acid (ATRA) to target the *PML-RARA* fusion protein in acute premyelocytic leukemia (APL)[77] (although the identity of the fusion as the molecular target of ATRA was unknown at the time)[28].

The first targeted therapy developed from an understanding of the underlying molecular mechanism of a driver event[78] was trastuzumab (Herceptin), developed to target HER2 (the gene product of *ERBB2*) in HER2-positive breast cancers[55].

Development of trastuzumab also highlighted another important feature of targeted therapeutics: that their application tends to be most effective in patients with the genetic events that they target. Indeed, Genentech developed diagnostic tools to identify trastuzumab-benefitting patient populations alongside the development of the therapeutic antibody itself[79].

Following these, the first rationally discovered therapeutic kinase inhibitor was imatinib (Gleevec), targeting the BCR-ABL fusion protein in chronic myelogenous leukemia (CML)[80]. This 'miracle drug' transformed CML treatment, increasing five-year survival rates from only 30% to at least 89%[81].

Because driver alterations such as those affecting *PML-RARA*, *ERBB2*, and *BCR-ABL* directly contribute to oncogenesis, they represent logical targets for cancer drug discovery programs. Continued efforts to inhibit driver genes have produced several successes, including vemurafenib and dabrafenib (targeting BRAF V600E mutations in melanoma and brain metastases); and erlotinib, cetuximab, and panitumumab (targeting mutant EGFR in colorectal, pancreatic, and non-small cell lung cancers)[82].

Yet despite these notable examples, there remain few targeted therapies in comparison to the number of identified driver genes[82], because many classes of driver genes have been difficult to drug. Some driver proteins (such as kinases, growth factor receptors, G-protein coupled receptors, and receptor ligands) are referred to as 'druggable' because they contain deep, hydrophobic pockets that are amenable to small-molecule binding or they are exposed to the extracellular environment and can therefore be targeted by antibodies or other biologics[83]. However, other driver classes, including transcription factors (e.g., STAT3) and *RAS* family proteins, lack such favorable

properties for small-molecule intervention, and are typically labelled as 'undruggable'. Furthermore, in the case of tumor suppressor genes, recapitulating the function of a lost or inactivated protein is challenging[83].

Moreover, any given tumor is likely to harbor few driver events, reducing the likelihood that it harbors a druggable driver event. Recent analyses indicate the existence of around 300 driver genes across cancers[84], yet both functional and genomic evidence suggests that many tumors derive from only a handful of driver alterations. For instance, experiments with normal human lung fibroblasts have demonstrated that transformation can occur with a minimum of five oncogenic drivers[85]. Genomic studies have suggested that the average lung or colorectal tumor may contain only three driver alterations[86], while pan-cancer efforts have identified only four driver mutations in the typical tumor, with a range between one and ten drivers based on tumor type[87]. This small repertoire of initiating alterations limits the probability of finding an actionable mutation in any given tumor. Given these obstacles in drugging many types of driver alterations, there is a need to expand the universe of potential therapeutic targets.

Synthetic lethal targeting of driver genes

Synthetic lethality represents a strategy to broaden the landscape of cancer targets beyond driver genes themselves. Two genes exhibit a synthetic lethal interaction when the disruption of either gene alone has no effects on viability, while the disruption of both genes generates a lethal phenotype[88].

One approach to leveraging synthetic lethal interactions for cancer therapeutics involves identifying genes that are synthetic lethal with altered driver genes. The most clinically advanced example of this approach is the treatment of *BRCA*-mutant tumors

with poly(ADP-ribose) polymerase (PARP) inhibitors. PARPs trigger the repair of single-stranded DNA breaks, which, if not resolved, cause stalled replication forks and potentially double-stranded breaks (DSBs). DSBs are in turn resolved through homology-directed repair in a process requiring *BRCA1* and *BRCA2*. In the setting of *BRCA* mutation, inhibiting PARP leads to an accumulation of unresolved stalled replication forks, causing increased DSB formation, catastrophic DNA damage, and ultimately, apoptosis[88–90]. Additionally, synthetic lethal interactions with activated oncogenes represent an active area of research. For instance, small-molecule and RNAi screens of *KRAS*-mutant and wild-type cell lines have been used to identify putative synthetic lethal interactions with mutant *KRAS*[89,91]. In the case of both BRCA and Ras proteins, targeting the driver mutation directly—by replacing a lost tumor suppressor function or reactivating an inactive GTPase—has proven challenging. Thus, targeting a synthetic lethal partner of driver alterations may provide more tractable routes to therapeutic development.

Targeting non-driver genes

An alternative approach to generating treatments that are cancer-specific is to target non-driver genes that are somatically altered in cancer cells. This includes two classes of genes: genes affected by passenger events, and genes that are collaterally affected by driver events, but whose alteration does not itself confer selective advantage to the cell. Several potential therapeutic approaches have been developed to exploit somatically altered non-driver genes: cancer immunotherapy, targeting genes that have undergone partial copy-number loss, targeting paralogs of inactivated genes, and targeting pathways that are rendered essential in the context of alterations to a non-driver

gene. This thesis explores an additional category: targeting GEMINI genes for which loss of heterozygosity leads to an allele-specific dependency.

Cancer immunotherapy

Harnessing a patient's immune system to treat cancer has been a goal of cancer therapeutic development for centuries[92]. In all cases, cancer immunotherapy requires elements of the immune system to distinguish between cancer cells and normal cells. A major mechanism underlying these distinctions is recognition of the consequences of genetic alterations to non-driver genes.

An immunotherapeutic strategy that has recently been shown to be highly effective in multiple contexts is immune checkpoint blockade[93,94]. The natural immune system is capable of mounting a response against cancer cells, for example through the activation of tumor-specific cytotoxic T cells (CTLs). However, this immune response can diminish or become ineffective due to immunosuppressive signaling (immune checkpoints) or evolution of the tumor cell population away from antigens recognized by CTLs (immunoediting)[95]. Immunotherapies targeting immune checkpoint components can block immunosuppressive signaling, restoring the ability of CTLs to target cancer cells. Such immune checkpoint blockade can occur at the level of interactions between antigen-presenting cells and CTLs (e.g., the anti-CTLA-4 antibody ipilimumab) or between CTLs and tumor cells (e.g., the anti-PD-1 antibody pembrolizumab, the anti-PD-L1 antibody atezolizumab)[96]. Checkpoint inhibitors have been approved for use in several cancer types including metastatic melanoma and non-small-cell lung cancer, and clinical trials are underway for additional indications[96].

Another group of immunotherapeutic approaches relies on adoptive T-cell transfer, which involves the removal and subsequent reinfusion of patient T cells after ex vivo intervention[95]. For instance, in tumor infiltrating lymphocyte (TIL) therapy, T cells that have infiltrated a resected tumor are cultured externally to identify those that can kill tumor cells. These TILs are then expanded and reinfused into the patient, where they then hopefully target residual disease. TIL therapy has been successfully implemented in some tumor types, primarily metastatic melanoma and other cancers with high mutational burden[95].

In both immune checkpoint blockade and adoptive T-cell transfer, genetic alterations to non-driver genes have been suggested to underlie immunorecognition of cancer cells. In the setting of checkpoint blockade, Timothy Chan and others have determined that response was associated with mutational burden[97–103], including (and primarily) mutations in non-driver genes[104,105]. The proposed mechanism was that these mutations generate modified proteins which, when displayed by major histocompatibility complexes (MHCs) on the surface of tumor cells, represent neoantigens that are recognized by CTLs or helper T-cells. Subsequent studies have refined these predictive models by explicitly accounting for which mutations generate neoantigens, how neoantigens vary by MHC type, and even which neoantigens are expressed in a given cancer[104–108].

Similarly, cancer vaccines can rely on neoantigens. A number of vaccines have been developed using tumor tissue or cancer cell line material combined with various adjuvants, presumably relying on neoantigens expressed by the cancerous cells to

generate selective immunorecognition[109]. Indeed, recent work has explicitly accounted for neoantigens in patients' tumors to engineer patient-specific vaccines[110].

Collateral lethalities

Our laboratory and others have explored 'collateral lethalities'[14], which derive from either driver or passenger genetic alterations acting on non-driver genes. In most cases, these collateral lethalities are due to wide-ranging effects of SCNAs in cancer. As discussed above, SCNAs are often large and abundant: nearly 90% of cancers exhibit some degree of aneuploidy[44], and the underlying whole-chromosome or arm-level SCNAs can encompass hundreds of genes in a single event. SCNAs can alter non-driver genes through two scenarios: as part of a driver event or as part of a passenger event. For instance, ~35% of cancers exhibit arm-level deletion of 17p, presumably as a driver event targeting *TP53*[39,44]. This single deletion event alone results in the collateral deletion of nearly 500 neighboring genes, generating broad regions of gene copy-loss[39]. Passenger events, which exert neutral or negative effects on tumor fitness, can also occur frequently and affect large numbers of genes. For example, chromosome 2q represents one of the less frequently deleted chromosome arms (lost in ~16% of cancers, or over 277,600 patients per year in the US[111]), yet this loss encompasses over 1000 genes[39].

Collateral lethalities encompass four classes of vulnerabilities: redundant paralog dependencies; redundant non-homolog dependencies; drug-induced haplolethality, including CYCLOPS genes; and LOH of essential genes.[14]

Paralog dependencies may arise when a gene with a crucial function is homozygously deleted, but a closely related paralog compensates for its activity[112]. For example, *ENO1*, which encodes the glycolytic enzyme enolase 1, resides on chromosome

1p and is frequently deleted in glioblastoma multiforme (GBM). GBM cells express only two enolase paralogs, *ENO1* and *ENO2*; knockdown or inhibition of *ENO2* suppresses growth of *ENO1*-deleted GBM cells but not *ENO1*-wild type GBM cells[112]. Additional validated paralog dependency pairs include *MAGOHB* dependency with *MAGOH* deletion[113], *ARID1B* dependency with *ARID1A* deletion[114], *SMARCA2* dependency with *SMARCA4* deletion[115], *ME3* dependency with *ME2* deletion[116], and *UBC* dependency with *UBB* deletion[117]; dependencies involving other paralog pairs have also been predicted[45,113,117,118].

The homozygous deletion of a gene with critical functions in cellular homeostasis may generate a dependency on a non-homologous gene in a compensatory or related pathway. For instance, the gene encoding methylthioadenosine phosphorylase (*MTAP*) lies in the region that sustains the most homozygous deletions across human tumors, 9p21 (which includes the *CDKN2A/B* tumor suppressor locus)[14]. Because of its high rate of homozygous deletion, many efforts have been dedicated to identifying collateral lethalities resulting from *MTAP* loss[14,119–121]. MTAP is involved in the salvage of methionine and adenine; *MTAP*-deleted cells are sensitive to inhibitors of purine biosynthesis *in vitro*, although *in vivo*, purines in the circulation/microenvironment abrogate this effect[122]. In 2016, investigators at Agios Therapeutics discovered that homozygous *MTAP* deletion sensitizes cells to depletion of arginine methyltransferase (PRMT5); PRMT5 upstream enzyme methionine adenosyltransferase II alpha (MAT2A); and PRMT5 complex members, including RIOK1[123], providing proof-of-concept of this class of collateral lethalities targeting non-paralogous genes.

Homozygous deletion of 6-phosphogluconate dehydrogenase (*PGD*) may represent another example of this class of collateral lethality. Experiments in yeast have indicated that deletion of *PGD*, which codes for a critical enzyme in the oxidative pentose phosphate shunt, causes a dependency on members of the non-oxidative pentose phosphate shunt[124]. This potential vulnerability remains to be validated in cancer cells. However, *PGD*-deleted cells have been shown to be glycolysis-deficient, generating a dependency on the mitochondrial oxidative phosphorylation (OXPHOS) pathway[125]. In addition, cancer cells with inactivating mutations in fumarate hydratase (*FH*), a driver of hereditary leiomyomatosis renal cell carcinoma (HLRCC), are uniquely sensitized to *PGD* inhibition, providing an example of a synthetic lethal interaction with a driver alteration[126].

Heterozygous deletion of essential genes may also generate cancer dependencies. Tumor cells that have undergone single-copy loss of a gene must rely solely on the remaining copy to provide any necessary gene products. Deletion of the second gene copy may not harm cells if the gene is unnecessary for survival; however, loss of the second copy of an essential gene would not be tolerated. Our laboratory has shown that hemizygous loss of essential genes sensitizes cancer cells to further knockdown of those same genes, whereas cells exhibiting no copy loss are unaffected by knockdown. This dependency class is termed CYCLOPS (Copy number alterations Yielding Cancer Liabilities Owing to Partial losS) genes[43]. For example, cancer cells with copy loss of the CYCLOPS gene *PSMC2*, which codes for an essential member of the 19S proteosome, exhibit growth arrest and apoptosis upon *PSMC2* knockdown by RNAi in vitro and in vivo. This result contrasts with cancer cells without *PSMC2* copy loss, which exhibit no proliferation or survival defects upon *PSMC2* knockdown[43]. Similarly, knockdown of the

CYCLOPS gene *SF3B1*, which codes for an essential subunit of the U2 snRNP splicing complex, kills *SF3B1* copy-loss but not *SF3B1* copy-neutral cells[45]. The therapeutic windows generated by gene copy-loss of *PSMC2* and *SF3B1* were both mediated by altered stoichiometry of essential macromolecular complexes, demonstrating a mechanism by which aneuploidy may generate targetable differences in tumor cells.

Additional CYCLOPS genes have been experimentally or therapeutically validated. The first, *POLR2A*, codes for the catalytic subunit of the RNA polymerase II complex. *POLR2A* resides on chromosome 17p and is frequently co-deleted with *TP53*, leading to *POLR2A* hemizygosity in over 50% of colorectal cancers. *POLR2A* copy-loss cells were shown to be uniquely sensitized to *POLR2A* knockdown by shRNAs and POLR2A inhibition by an α-amanitin antibody-drug conjugate[117,127]. A second CYCLOPs gene product, casein kinase 1α (CK1α), has also been verified as the downstream target of a pre-existing therapeutic, lenalidomide. Lenalidomide has been recognized an effective treatment for 5q-deleted myelodysplastic syndrome (MDS) since the early 2000s[128,129], but the mechanism of action remained unclear. In 2015, Benjamin L. Ebert and colleagues determined that lenalidomide acts by inducing $CRL4^{CRBN}$-mediated ubiquitination and subsequent degradation of CK1α. This protein is the gene product of *CSK1A1*, which resides on chromosome 5q. As a result of *CSK1A1* copy-loss in 5q-deleted MDS patients, CK1α levels in hematopoietic stem cells are decreased, leaving these cells uniquely vulnerable to the CK1α degradation triggered by lenalidomide. This finding indicates that the concept of targeting CYCLOPS genes is therapeutically viable.

Large-scale analyses have identified additional candidate CYCLOPS genes[43,45], including those for which decreased expression rather than gene copy-loss confers sensitivity to genetic inhibition[117].

LOH of essential genes

This thesis investigates a fourth class of collateral lethalities, which relies not on relative differences in gene copy number or expression between tumor and normal cells but rather on absolute differences in allelic constitution. When a cancer cell undergoes LOH of a gene, it then relies on the gene products encoded by a single allele of that gene, in contrast to normal cells, which retain both alleles. In the case of a cell-essential gene, further loss or inhibition specifically of the allele retained in the tumor should not be tolerated, whereas normal cells will be able to survive relying solely on the remaining allele[130]. This concept is known as 'variagenic targeting'[131] or 'allele-specific inhibition'[130] and was first proposed by David Housman and colleagues in 1999.

Variagenic targeting has been performed with a degree of success in two essential genes with SNPs. The first gene, *RPA1* (also known as RPA70), is a member of the single-strand binding (SSB) protein complex called replication protein A (RPA)[131]. This complex is required for DNA replication and some DNA repair processes, and inactivation of any of the three RPA subunits is lethal in yeast. To identify candidates for ASI, Basilion et al. genotyped *RPA1* in lymphoblast cell lines from 36 individuals and identified five common SNPs, one of which (*RPA1*^{T1674C}) was chosen for follow-up experiments. Phosphorothioate oligodeoxynucleotides (PS-ODNs) targeting either allele of *RPA1* were administered to cell lines containing only one or the other allele. These reagents knocked down *RPA1* mRNA and decreased cell survival in an allele-selective manner, albeit with a small

therapeutic window[131]. This small therapeutic window and those observed in later studies may have been due to inefficient delivery of PS-ODNs to cells; nonoptimal reagent length, chemistry, or binding site relative to the variant; or limitations on allele-specific binding or inhibition by reagents inherent in the genetic context of the variant itself[131].

In 2000, ten Asbroek et al. characterized a SNP in *POLR2A*, which codes for the core of RNA polymerase II, as a target for ASI. As with *RPA1*, PS-ODNs targeting this SNP (*POLR2A*^{C2673T}) achieved allele-selective knockdown, although not at a level or for a duration necessary to affect cell viability[132]. A later study by Fluiter and colleagues in 2002 tested the same ODNs in vivo using tumor xenografts of cell lines homozygous for either allele. Administering the ODNs continually using an osmotic mini-pump for 10 days led to allele-specific differences in tumor volume, again with a small therapeutic window[133]. In 2009, similar results were obtained in tumor xenograft models using siRNA, with cells of the target genotype exhibiting ~50% less tumor growth than those of the 'mismatched' genotype[134].

Previous studies investigating variagenic targeting also identified and validated a limited number of candidate variants[130,131]. Sequencing of 16 essential genes in 36 lymphoblast cell lines identified 35 common polymorphisms. Of these, 22 targeted by ODNs exhibited mRNA knockdown, with 19 exhibiting at least a degree of allele-selective knockdown[131]. A second study identified 47 additional variants in 12 new genes[130]. However, the high frequency of genetic variation in the human genome[135] combined with the large number of predicted essential genes[136] suggests that many additional targets for ASI exist in the cancer genome.

Human genetic variation

Human populations exhibit an abundance and diversity of genetic variation among individuals. These genetic variants generally originate with de novo mutations in a single individual, with a typical individual's genome containing ~40 new mutations[137]. Most new variants remain rare in a population across generations: deleterious variants are selected against, and even neutral or beneficial variants are likely to remain rare or be lost due to factors such as genetic drift[137]. Accordingly, while human exomes contain approximately one variant for every 4 bp within the exome intervals[138], greater than 99% of these variants have a minor allele frequency of less than 1% and are therefore considered rare[135]. In addition, greater than 50% of variants in exomes were detected only once in a dataset of over 60,000 people. Common variants (minor allele frequency >1%) still abound, however—the genome is estimated to contain one common variant per ~500 bp[137]. These variants are not distributed evenly throughout the genome; rather, the type and location polymorphisms vary based on factors such as mutational class and selection pressures[135].

In addition to being labelled as either common or rare, variants can also be classified as either single-nucleotide polymorphisms (SNPs) or structural variants. SNPs represent the most abundant class of genetic variation. Structural variants comprise all variants that are not single-nucleotide variants, including insertion-deletion variants (indels), block substitutions, inversions, and copy-number variants (CNVs; these are germline variants, distinct from somatic CNAs observed in cancer cells). Structural variants represent a much smaller fraction of total variants but affect a larger fraction of basepairs; indels are the most common structural variant sub-class[139].

Genome-scale determination of genetic variants has only become possible within the last 20 years. Beginning in 1998, the dbSNP database served as a publicly available

repository of genetic variants identified and contributed by the scientific community at large[140], including groups such as The SNP Consortium[141]. The completion of the Human Genome Project (HGP) represented progress towards high-throughput identification of variants, yet the two reference genomes published in 2001 by the Human Genome Sequencing Consortium and Celera Genomics were haploid and did not annotate polymorphisms[139]. In 2007 and 2008, four diploid whole genomes were published—those of J. Craig Venter[142], James D. Watson[143], and two anonymous individuals (one Han Chinese[144] and one Nigerian[145]). These studies expanded the understanding of the classes and evolutionary histories of genetic variants[139]. The 1000 Genomes Project followed in 2008, vastly expanding the number and diversity of individuals whose genetic variation had been profiled[146,147]; at its completion in 2015, this effort had profiled variants across over 2500 individuals (belying the name) from 26 human popuations using whole-genome and targeted-exome sequencing and SNP microarrays[148]. The Exome Sequencing Project (ESP) also made significant contributions, characterizing over 1 million SNPs using whole exomes from over 6500 European and African Americans[149].

The most recent studies of human genetic variation have once again dramatically expanded the number of individuals and variants profiled. Using whole-exome sequencing data from over 60,000 individuals, the Exome Aggregation Consortium (ExAC) indentified over 7 million variants[135]; most recently, the Genome Aggregation Database (gnomAD) identified nearly 15 million variants and 230 million variants using over 125,000 whole exomes and 15,000 whole genomes, respectively[138].

Cell-essential genes

Cell-essential genes are genes that are necessary for viability and successful reproduction of a single cell (in contrast to organism-essential genes, which are required for the survival and successful reproduction of a whole organism)[150]. Three main avenues have been pursued to define essential genes in humans: large-scale germline sequencing of human cohorts; experimental manipulation of non-human model organisms; and forward and reverse genetic approaches in human cancer cell lines[90].

Human sequencing studies rely on rates of gene inactivation to define the 'dispensible genome' (genes whose homozygous inactivation is compatible with life and is thus observed in human populations) and the 'essential genome' (genes that are rarely or never inactivated in human populations). However, because loss-of-function variants are rare in the human germline, current studies are mostly limited to identifying haploinsufficient genes rather than genes for which biallelic inactivation is not tolerated[151]. This approach to identifying essential genes is further complicated by the fact that absence of inactivating variants in a cohort does not indicate they do not exist in the population as a whole; for instance, individuals heterozygous for loss-of-function variants may present with severe phenotypes, leading to their exclusion from studies such as ExAC[151]. Last of all, essential genes identified by large-scale sequencing studies likely include those required for the development or survival of an entire organism but that are non-essential in single cells[151].

Using model organisms allows investigators to experimentally demonstrate rather than statistically infer essentiality in a way not possible with human studies. Methods of identifying essential genes in model organisms have progressed from random chemical- and insertional-mutagenesis screens to targeted single knockouts to targeted genetic

screens using RNA interference (RNAi) and CRISPR-Cas9 systems[150]. The use of model organisms allows investigators to identify genes that are essential in the context of biallelic rather than monoallelic inactivation, which human studies have not yet been able to capture due to insufficient power[151]. Yet model systems also exhibit limitations in determining essential genes, including the evolutionary distance between human and non-human organisms and (in the case of multicellular organisms) the issue of separating organism-essential from cell-essential genes.

Human cell lines studies of cell-essential genes retain the benefits of experimental manipulation found in other model systems while eliminating the drawbacks of evolutionary distance and whole-organism development. RNAi, CRISPR knockout (KO), and CRISPR-interference (CRISPRi) screens have all been used to determine essential genes in cell line models. CRISPR-based systems generated less data variation, a higher fraction of functional reagents, and fewer off-target effects than shRNA-based systems[152]. In addition, CRISPR KO screens identify a greater number of essential genes[153], perhaps because the incomplete suppression provided by RNAi screens may not always decrease expression below minimal thresholds required for survival[43,45]. However, combining results from RNAi and CRISPR KO screens may improve overall essential gene detection[153].

Essentialomes generated using KO screens of human cell lines may vary based on the environmental context (e.g., cell culture conditions) or genetic background (e.g., oncogenic drivers, degree of aneuploidy) of the lines being probed[150]. One can also imagine a class of context-dependent essential genes required for the growth and survival of specific cellular lineages (e.g., breast tissue) or developmental state (e.g., stem cells).

Across the tree of life, essential genesets are enriched for genes involved in basic cellular processes including DNA, RNA, and protein biogenesis[150]. For example, analyses of the evolutionarily divergent yeast species *S. cerevisiae* and *S. pombe* indicate essentialomes enriched for genes involved in DNA, RNA, lipid, and protein production; transcriptional initiation; and ribosome formation[90].

DNA primase

DNA primase is an essential enzyme complex responsible for synthesizing the RNA primers required on both the leading and lagging strands of the DNA replication fork. These primers are necessary because in order to initiate DNA synthesis, DNA polymerases require the presence of a 3'-hydroxyl group on a nucleotide already base-paired to the template strand[154]. DNA primase provides this 3'-hydroxyl group by synthesizing a short RNA oligomer that is complementary to the template strand and 6–10 nucleotides in length[155]. After the primer is synthesized, it is transferred to DNA polymerase α (Pol α), which initiates DNA synthesis[156].

Eukaryotic DNA primase exists in complex with Pol α, forming the Pol α–primase complex. Human DNA primase contains two subunits: a 49-kDa, catalytic subunit (p49, encoded by *PRIM1*) and a 58-kDa, regulatory subunit (p58, encoded by *PRIM2*). The human DNA Pol α complex also consists of two subunits: the 180-kDa, catalytic Pol α and a 70-kDa, regulatory subunit known as polymerase B[155].

The functions of DNA primase extend beyond primer synthesis at the DNA replication fork to include roles in DNA repair, cell cycle control, apoptosis, and telomere maintenance[157]. In *S. cerevisiae*, temperature-sensitive mutants of the *PRIM1* and *PRIM2* homologs *pri1* and *pri2* exhibit defective cell growth and DNA synthesis[158], as well as

increased rates of mutation and mitotic recombination[159]. *pri1* mutant cells in particular escape the G1/S checkpoint and fail to slow S-phase progression in response to DNA damage[160]. Likewise, studies in *Xenopus* extracts have suggested that the presence of primed ssDNA may be required for replication checkpoint induction by the ATR-Chk1 pathway[157,161,162]. PRIM1 may also play a role in cell death: developing zebrafish with a missense mutation in the homolog *prim1* exhibited aberrant apoptosis of differentiating retinal neurons mediated by the ATM-Chk2-p53 pathway[163]. Primase is also required for telomere maintenance in yeast, where it binds ssDNA chromosome ends that have been elongated by telomerase and uses them as templates to initiate synthesis of the complementary strand[157,164].

Unsurprisingly, DNA primase represents an essential enzyme complex in eukaryotes. Mutations in *S. cerevisiae* homologs *pri1* and *pri2* cause lethality[165,166]. In addition, *PRIM1* has been identified as an essential gene in high-throughput genetic knockout screens in human cancer cell lines[136].

RNA exosome

The RNA exosome is a 10- or 11-member protein complex responsible for the processing or degradation of a wide variety of RNA species[167,168]. The nine-member, noncatalytic exosome 'core' contains two subcomplexes. The first, the 'hexamer' or 'ring' structure, consists of six RNase PH-like proteins (Rrp41, Rrp42, Rrp43, Rrp45, Rrp46, and Mtr3). The second, the 'cap' structure, sits atop the ring and consists of three proteins with S1 and/or KH RNA-binding domains (Rrp4, Rrp40, and Csl4)[168]. The barrel-shaped exosome core can associate with two ribonucleases: Rrp6 (also known as exosome component 10 or PM/Scl-100) and Dis3 (also known as Rrp44)[167]. Rrp6 binds

on the cap end of the core and contains one active site with distributive 3'-to-5' exoribonuclease activity. Dis3 binds on the ring end of the core and contains two active sites, one with processive 3'-to-5' exoribonuclease activity and one with distributive endoribonuclease activity[168]. Exosome complexes vary in ribonuclease composition, co-factor association, and function by subcellular compartment.

The nuclear RNA exosome is responsible for the degradation of cryptic transcripts. 'Cryptic' transcripts represent highly unstable species that are rapidly degraded by the RNA surveillance system under normal cellular conditions. Unstable transcripts targeted by the exosome include antisense transcripts from bidirectional promoters (known as promoter upstream transcripts, or PROMPTs), bidirectional enhancer transcripts (known as enhancer RNAs, or eRNAs), intergenic or antisense long non-coding RNAs (lncRNAs), and transcripts of repetitive elements such as ribosomal DNA (rDNA) repeats or centromeres[167].

The exosome also plays a role in the quality control and turnover of stable transcripts. For example, the nuclear exosome degrades RNAs that have been improperly processed, including mRNAs with aberrant splicing or 3' end generation, as well as those that have been inappropriately packaged into ribonucleoprotein particles (mRNPs)[167]. The cytoplasmic exosome likewise participates in RNA surveillance, including the translation-dependent nonsense-mediated decay (NMD), non-stop decay (NSD), and no-go decay (NGD) pathways[168,169]. The exosome also functions to degrade transcripts as a part of normal RNA turnover (e.g., of AU-rich element–containing mRNAs) or specific gene regulation programs[167,169].

Along with its roles in RNA surveillance, the exosome is involved in normal biogenesis of non-coding transcripts. For instance, the exosome complex was first discovered in *S. cerevisiae*, where a mutation in the Dis3 homolog gene *rrp4* disrupts normal 3' end processing of the 7S pre-rRNA into the mature 5.8S rRNA[168]. The exosome also participates in the processing of small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs)[167].

The RNA exosome represents a cell- and organism-essential protein complex. All nine exosome core proteins as well as the DIS3 homolog dis3 are essential in *S. cerevisiae*, and yeast cells lacking the Rrp6 homolog rrp6 exhibit growth defects[168,170,171]. Additionally, hypomorphic alleles of exosome core member genes *EXOSC3* and *EXOSC8* (which code for Rrp40 and Rrp43, respectively) cause recessive neurological diseases, including pontocerebellar hypoplasia and spinal muscular atrophy[167,172,173].

Given their essential roles in maintaining cell viability, genes encoding components of DNA primase and the RNA exosome are likely themselves to be cell-essential. As described below, we identified common variants in two of these genes, *PRIM1* and *EXOSC8*, and studied them as candidates for an approach targeting LOH of essential genes.

This page intentionally left blank.

CHAPTER 2: GENOMIC ANALYSIS OF LOH OF ESSENTIAL GENES

This page intentionally left blank.

CHAPTER 2: GENOMIC ANALYSIS OF LOH OF ESSENTIAL GENES

Despite progress in precision cancer drug discovery, few highly selective therapies exist in the clinic. A current paradigm focuses on drugging driver alterations in cancer; however, many driver genes have proven difficult to target therapeutically[85,87,174], and in many cancers no easily targeted drivers exist. Alterations in non-driver genes represent an alternative target class that merits further investigation.

Loss of heterozygosity (LOH) may generate cancer-specific vulnerabilities by eliminating genetic redundancy in cancer cells. LOH occurs when a cancer cell that is originally heterozygous at a locus loses one of its two alleles at that locus, either by simple deletion of one allele (copy-loss LOH), or by deletion of one allele accompanied by duplication of the remaining allele (copy-neutral LOH). In either case, the cancer cell then relies on the gene products encoded by a single allele, in contrast to normal cells, which retain both alleles. When a cancer cell undergoes LOH of an essential gene, further loss or inhibition specifically of the allele retained in the tumor should not be tolerated, whereas normal cells will be able to survive relying solely on the remaining allele[130] (Figure 1A). We term this target class GEMINI vulnerabilities, after the twins from Greek mythology Castor and Pollux, one of whom was mortal and the other immortal.

While previous reports have described individual GEMINI vulnerabilities[131,132], these studies have not systematically evaluated the landscape of potential targets, taking into account genome-scale assessments of gene essentiality, variation in human genomes, and rates of LOH across cancers. Open questions include which essential genes exhibit widespread variation in human populations and frequent LOH in cancers, providing

Figure 1: Rates of LOH and allelic variation in normal and cancer genomes.

a. Schematic indicating how loss of heterozygosity (LOH) of essential genes represents a potentially targetable difference between cancer and normal cells. **b.** Violin plot of minor allele frequency of polymorphisms in essential versus non-essential genes in the ExAC cohort. Intersecting lines represent median values: essential = 0.141, non-essential = 0.146; one-tailed Student's t-test, $^{**}p < 0.01$. **c.** (Left) Overlap between genes with common polymorphisms in the ExAC database (pink circle) and essential genes (blue circle). (Right) Fraction of essential genes with common polymorphisms. **d.** Percent of genome affected by LOH across 9,686 cancers from TCGA. **e.** Stacked histogram representing the number of genes with copy-loss (yellow) or copy-neutral LOH (purple) across 9,686 cancers from TCGA. **f.** Dot plot of the number of essential genes affected by LOH across 33 TCGA tumor types. Tumor types are indicated by TCGA abbreviations (see https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations). Each dot represents an individual sample. Lines indicate median values.

Figure 1 (Continued).

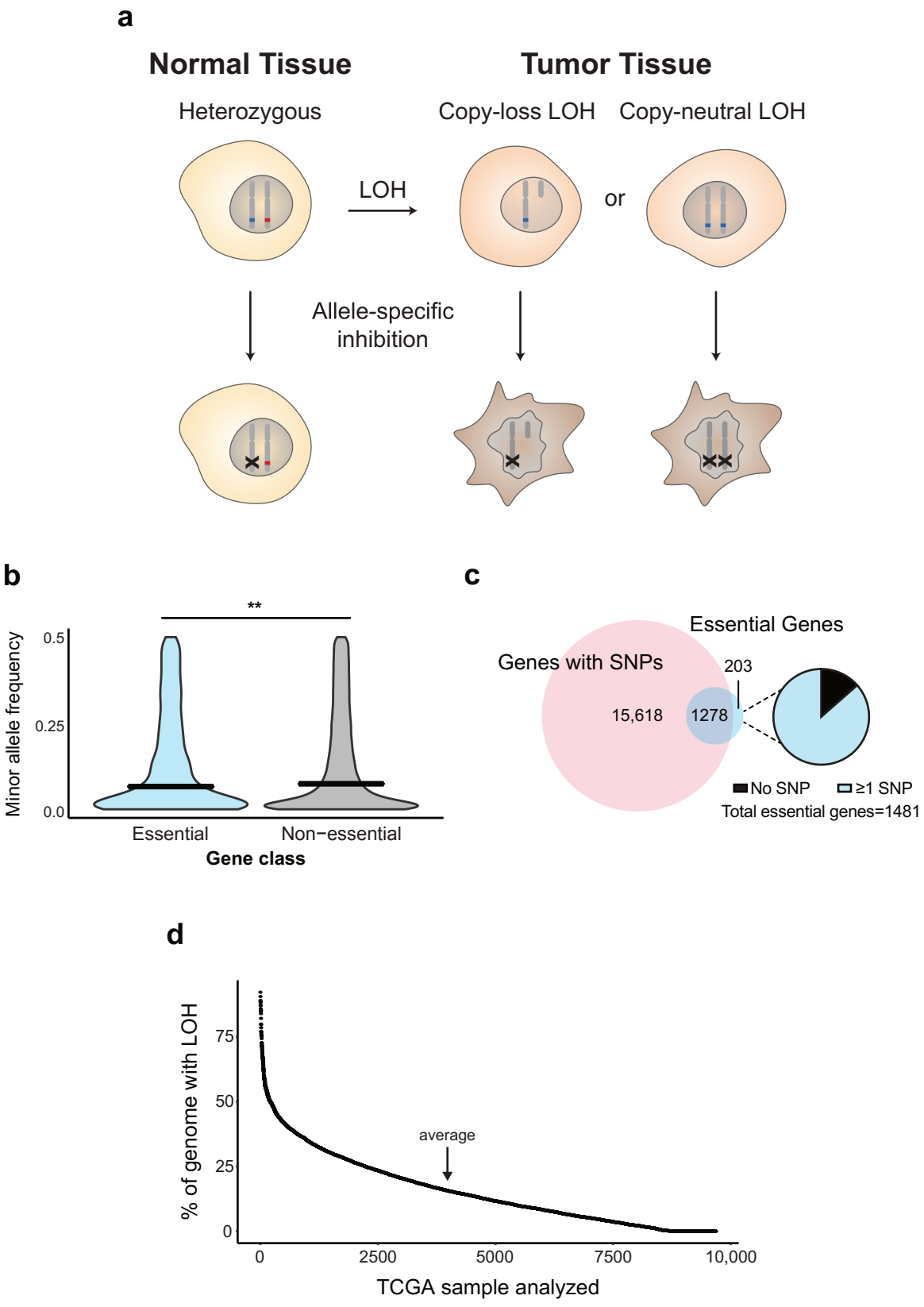
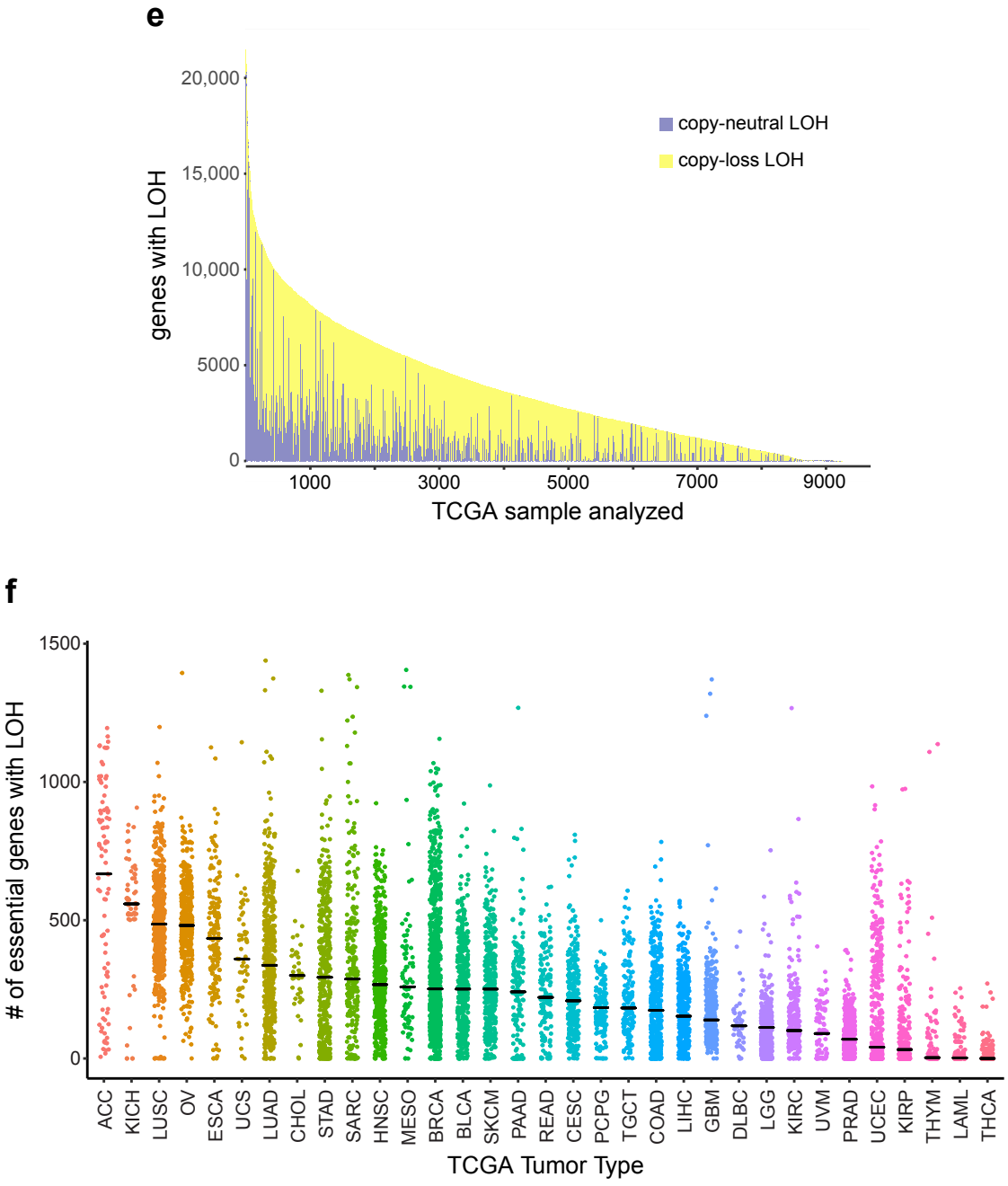


Figure 1 (Continued).



potential GEMINI vulnerabilities, and at what rates these vulnerabilities occur.
To address these questions, we integrated genome-scale copy number, germline allelic variation, and gene essentiality data to identify a list of polymorphisms in cell essential genes that undergo LOH in cancer, serving as a compendium of potential GEMINI targets. These results define for the first time the potential scope of the GEMINI class of vulnerabilities.

Results

To identify potential targets for our approach, we first characterized the landscape of cell-essential genes. We integrated genome-wide gene essentiality data from loss-of-function genetic screens and CCLE cell lines to conservatively estimate 1481 cell-essential genes (Supplementary Table 1; Methods). This list is enriched for genes involved in essential cellular processes including rRNA processing, mRNA splicing, and translation (functional enrichment analysis performed with DAVID[175,176], version 6.8; Supplementary Table 2).

We then assessed germline heterozygosity resulting from normal human genetic variation in coding regions and 5' and 3' untranslated regions (UTRs) using allele frequencies across 60,706 individuals in the Exome Aggregation Consortium database[135]. Variants at 90,409 loci were observed to be present among at least 1% of alleles. As expected, polymorphisms in essential genes are slightly less common than those in non-essential genes (median minor allele frequency: essential = 0.141, non-essential = 0.146; $p = 0.005$, one-tailed Student's t-test; Figure 1B). However, essential genes still contain an abundance of genetic variation: 86% (1278/1481) harbor at least one common germline variant (Figure 1C), with 49% (730/1481) harboring at least one missense

variant. The median essential gene contains 3 germline polymorphisms. The median polymorphism in an essential gene is heterozygous in 13.9% of individuals (Supplementary Figure 1A).

We were interested in how much of this heterozygosity in essential genes is lost in cancer. Loss of heterozygosity (LOH) in cancer frequently results from copy number alterations (CNAs) that can alter dozens to thousands of genes in cancer genomes[39,40]. Most LOH is due to strict copy-loss (copy-loss LOH), where allelic loss occurs in the context of a decrease in gene copy number. However, copy-neutral LOH is also frequently observed, whereby an allele is lost but the number of gene copies remains the same due to a duplication event, or in some cases, even increases. LOH has been frequently described[40,54], but to our knowledge there has not yet been a systematic analysis of the frequency of LOH events across cancer types.

We therefore analyzed copy number and LOH calls from 9,686 patient samples across 33 TCGA tumor types (Methods). On average and across all cancers, 16% of genes undergo LOH (Figure 1D). Genome-wide LOH rates vary widely by tumor type, however, ranging from a median of 45% in adenoid cystic carcinoma to 0.01% in thyroid carcinoma (Supplementary Figure 1B). Approximately 28% of genes undergoing LOH undergo copy-neutral LOH (Figure 1E), and on average across all cancers, 4.4% of all genes undergo copy-neutral LOH.

Rates of LOH are no lower for cell-essential genes relative to the rest of the genome (essential: 16.4%, non-essential: 15.6%; p=1, one-tailed Student's t-test; Supplementary Figure 1C), suggesting that LOH of essential genes does not impose

negative selection pressure. As a result, tumors harbored an average of 189 essential genes with LOH (Figure 1F).

We hypothesized that the widespread nature of LOH of essential genes could represent a new opportunity to target essential genes that are heterozygous in normal tissue but undergo LOH in cancer. Among individuals with heterozygous SNPs within an essential gene, cancer cells with LOH of that gene would rely solely on the gene product encoded by one allele, in contrast to somatic cells, which would retain both alleles. We therefore hypothesized that allele-specific inactivation of the allele that had been retained in the cancer would selectively kill the cancer cells (Figure 1A).

Our analysis identified 5664 polymorphisms in 1278 cell-essential genes, representing a compendium of potential GEMINI vulnerabilities (Supplementary Table 3). These GEMINI genes are enriched for similar pathways as the wider set of essential genes, including rRNA processing, mRNA splicing, and translation (functional enrichment analysis performed with DAVID[175,176], version 6.8; Supplementary Table 4).

Discussion

Our characterization of GEMINI vulnerabilities focuses on those in cell-essential genes. Analyses of *in vivo* human genetic population data or knockout mouse data can provide useful estimates of organism-essential genes. However, not all genes required for the development and reproduction of an entire organism may be essential for the growth and reproduction of an individual tumor cell (e.g., a gene required only in a specific tissue during a specific stage of development)[150,151]. We thus relied on knockout-screen, copy-number, and gene-expression data from over 500 cell lines to derive a conservative list of core cell-essential genes. Because these genes should be essential in individual

cells across a wide range of tissues, a therapeutic targeting one of them could be effective in treating multiple tumor types (given an appropriate delivery mechanism).

Future efforts to catalog GEMINI vulnerabilities should expand to include potential lineage-specific and non-coding dependencies. Along with genes essential across all cells, GEMINI vulnerabilities may also reside in genes essential in specific tissues. Including such lineage-specific essential genes may increase the number of potential candidates. Additionally, targeting a lineage-specific GEMINI gene could decrease therapeutic toxicity by limiting side effects of non-allele specific inhibition to a single tissue type. The identification of such lineage-specific essential genes may already be possible for some tumor types (e.g., breast) using the large amount of publicly available CRISPR knockout screen data[177]. Continuing efforts to catalogue dependencies across a large number of tumors from a diversity of tumor types may aid in these efforts as well[178]. However, CRISPR-based determinations of gene essentiality bear the caveat that essentiality may vary according to environmental context: some genes essential for growth of cell lines *in vitro* may not prove essential for tumor cells *in vivo*, and vice versa[150]. Nonetheless, CRISPR screens may provide a valuable tool for probing the essentiality of the non-coding genome, which may represent an untapped source of GEMINI vulnerabilities[151].

Methods

Variant lists

A list of 228,440 potentially targetable variants was downloaded from the Exome Aggregation Consortium (ExAC) database (exac.broadinstitute.org)[135]. Potentially targetable variants were defined as those in the following classes: 3_prime_UTR_variant,

5_prime_UTR_variant, frameshift_variant, inframe_deletion, inframe_insertion, initiator_codon_variant, missense_variant, splice_acceptor_variant, splice_donor_variant, splice_region_variant, stop_gained, stop_lost, stop_retained_variant, synonymous_variant. These variants were filtered to include only PASSing, common variants (global minor allele frequency ≥ 0.01) in genes for which copy number calls were available through the NCI Genomic Data Commons (see below for further details of copy number analyses).

All variant classes were included in the analysis of potential target SNPs for reversible genetic therapeutic approaches. All variant classes except 3_prime_UTR_variant and 5_prime_UTR_variant were included in the determination of variants generating or disrupting an *S. pyogenes* PAM site.

Genomic analyses of copy number and LOH from TCGA

Patient-derived genome-wide copy number and LOH data were downloaded from the TCGA Pan-Can project via the NCI Genomic Data Commons (https://gdc.cancer.gov/about-data/publications/pancan-aneuploidy) first reported in [44]. For copy number, gene-level log2 relative data were calculated by GISTIC 2.0, referenced in the output file 'all_data_by_genes_whitelisted.tsv'. Copy-loss was defined as log2 relative values ≤ -0.1 and copy-neutral was defined as > -0.1.

For LOH calls, we used TCGA analyses[44]. Briefly, SNP array and exome sequencing data from both tumor samples and paired normal DNA were used as inputs to ABSOLUTE[179], which calculated absolute allelic copy numbers genome-wide for each tumor. These absolute allelic copy numbers took into account the purity and ploidy of each tumor, as determined by ABSOLUTE. Autosomal regions for which the absolute

copy number of one allele was zero were considered to have undergone LOH. These ABSOLUTE calls are in the file 'TCGA_mastercalls.abs_segtabs.fixed.txt' (https://gdc.cancer.gov/about-data/publications/pancan-aneuploidy). These calls were transformed into per-gene calls for all subsequent analyses.

Essential gene list

Candidate essential genes were nominated using data from three genome-scale loss of function screens of haploid human cell lines (KBM7 with CRISPR-Cas9 gene inactivation or mutagenized with gene trapping[136], and pluripotent stem cells with CRISPR-Cas9 gene inactivation[180]). Briefly, all genes that passed a threshold of <10% FDR for a given cell line were included as a candidate essential gene. FDR corrected p-values from the original publications were used for both CRISPR screens; FDR q-values for the KBM7 gene trap scores were calculated using a binomial model (representing equal probability of gene trap inserting in a sense versus anti-sense orientation) and correction for multiple hypotheses using Benjamini and Hochberg. This initial candidate list contained 3431 genes, with 633 scoring as essential in all three screens.

These candidate essential genes were then filtered using CCLE gene copy-number and RNA expression data to determine if loss of function genetic alterations were observed in human cell lines. Genes that met any of the following criteria were excluded: homozygously deleted in >2 cell lines (log2 copy-number < -5); very low RNA expression (< 0.5 RPKM) in >5 cell lines; or homozygously deleted in 1 cell line that also has low RNA expression (<1.0 RPKM). This analysis reduced the list to 2566 candidate essential genes.

Genes were then filtered based on mean CERES score from CRISPR knockout screens of 517 cell lines (https://depmap.org/portal/download; derived from the file 'gene_effects.csv')[26]. Genes with CERES scores < -0.4 were excluded, yielding a list of 1499 essential genes. To account for instances in which CCLE copy-number and/or RNA expression data were not available for a particular gene, genes were rescued from the CCLE filter if they scored as essential in two of the three haploid cell line screens and had mean a CERES score < -0.4. This rescue yielded 17 genes, bringing the total number of candidate essential genes to 1516. This list was further filtered to remove genes classified as Tier 1 tumor suppressor genes in the COSMIC Cancer Gene Census (https://cancer.sanger.ac.uk/census/)[181], yielding a final list of 1482 essential genes. (One essential gene in this list, AK6, was not characterized in TCGA copy-number and LOH data and so was excluded from further analyses.)

This page intentionally left blank.

CHAPTER 3: PROOF-OF-CONCEPT VALIDATION OF TWO GEMINI VULNERABILITIES

This page intentionally left blank.

CHAPTER 3: PROOF-OF-CONCEPT VALIDATION OF TWO GEMINI VULNERABILITIES

While two potential GEMINI vulnerabilities have previously been probed experimentally, GEMINI vulnerabilities have never been validated in isogenic systems to confirm specificity. Among the 5664 GEMINI variants we identified in Chapter 1, 1688 lead to missense changes in amino acid composition of an essential protein, raising the possibility that they could be distinguished by molecules that interact with the protein directly. We focused on two of these missense SNPs for further functional analysis. Specifically, we performed proof-of-principle validation of GEMINI vulnerabilities in the genes *PRIM1* and *EXOSC8* using allele-specific CRISPR in both patient-derived and isogenic models. These results rigorously validate the GEMINI class of vulnerabilities as a whole and the GEMINI genes *PRIM1* and *EXOSC8* in particular.

Results

Validation of PRIM1^{rs2277339} *as a GEMINI vulnerability*

Variants residing in putative CRISPR protospacer adjacent motif (PAM) sites have previously been shown to enable allele-specific gene disruption[182–184]. For nuclease activity, *S. pyogenes* Cas9 requires a PAM site of the canonical motif 5'-NGG-3' downstream of the 20-nucleotide target site; deviations from this motif abrogate Cas9-mediated target cleavage[185,186]. Therefore, we hypothesized that in the case in which one allele of a SNP generates a novel PAM site, Cas9 would be able to disrupt the 'CRISPR-sensitive' (S), G allele that maintains the PAM sequence while leaving the other, 'CRISPR-resistant' (R) allele intact (Figure 2A).

Figure 2: Validation of *PRIM1*^{rs2277339} as a GEMINI vulnerability.

a. Schematic indicating allele-specific CRISPR approach. 'Preexisting genome' represents individuals heterozygous for a germline SNP in a *S. pyogenes* Cas9 protospacer adjacent motif (PAM) site. A 'G' allele (blue) in the PAM retains Cas9 activity at the target site, making this allele CRISPR-sensitive (S). An allele other than 'G,' represented by 'X' (red) abrogates Cas9 activity at the target site, making this allele CRISPR-resistant (R). Expression of an allele-specific (AS) CRISPR sgRNA targeting the polymorphic PAM site leads to specific inactivation of the S allele. **b.** Schematic of *PRIM1* SNP rs2277339 locus showing target sites for positive control, non-allele specific (NA) sgRNA and experimental, allele-specific (AS) sgRNA. Alleles appear in bold. **c.** Crystal structure of *PRIM1* gene product[284] shows the amino acid encoded by rs2277339 (teal) lies on the surface of the primase catalytic subunit (grey) near a potentially small-molecule accessible location. **d.** Immunoblot of PRIM1 protein levels in indicated patient-derived cell lines expressing LacZ, PRIM1 NA, or PRIM1 AS sgRNA. **e.** Representative growth curves of indicated patient-derived cell lines expressing LacZ (black), PRIM1 NA (red), or PRIM1 AS (blue) sgRNA, as measured by CellTiter-Glo luminescence, relative to day of assay plating. n = 5 technical replicates; error bars represent s.d. See Supplementary Figure 2 for additional biological replicates. **f.** Representative growth curves of indicated isogenic cell lines expressing LacZ (black), PRIM1 NA (red), or PRIM1 AS (blue) sgRNA, as measured by CellTiter-Glo luminescence, relative to day of assay plating. n = 5 technical replicates; error bars represent s.d. See Supplementary Figure 2 for additional biological replicates.

Figure 2 (Continued). g. Disruption of *PRIM1* in isogenic hemizygous *PRIM1* resistant ($PRIM1^{R}$) or *PRIM1* sensitive ($PRIM1^{S}$) cells expressing PRIM1 NA or AS sgRNA. Non-inactivated alleles (representing unaltered alleles or alleles with in-frame insertions or deletions, black) and alleles with frameshift alterations (yellow) were assessed by deep sequencing of *PRIM1* four days post-infection with sgRNA. **h.** Ratio of unaltered resistant to unaltered sensitive alleles of $PRIM1^{rs2277339}$ in *PRIM1* heterozygous cells ($PRIM1^{R/S}$) at day 4 and day 18 post-infection with NA or AS sgRNA as assessed by deep sequencing. Dashed line indicates expected ratio of unaltered R: S alleles if an sgRNA targets either allele with equal frequency. Chi-square with Yates correction, ***$p<0.0001$.

Figure 2 (Continued).

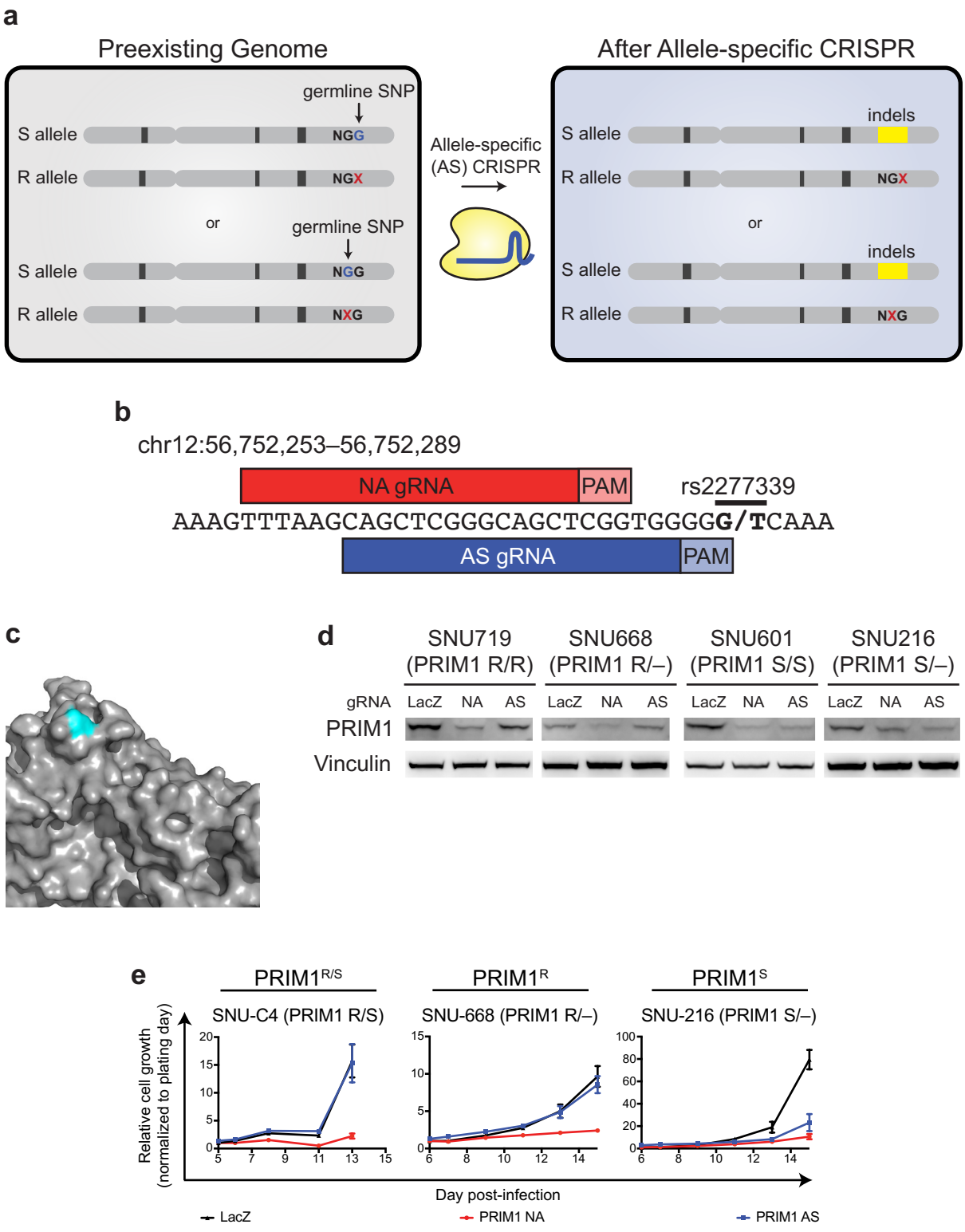
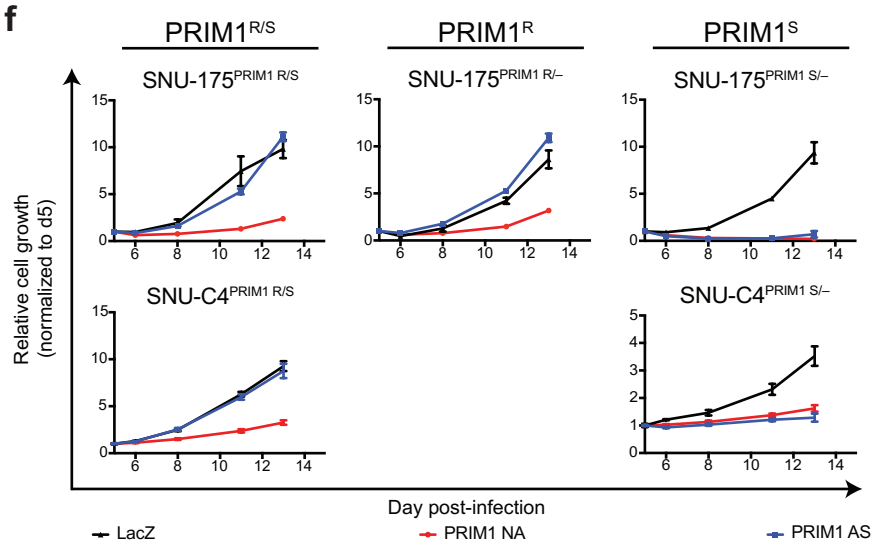
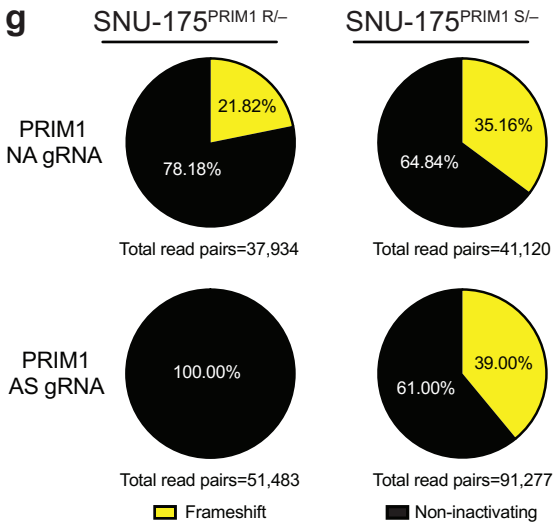
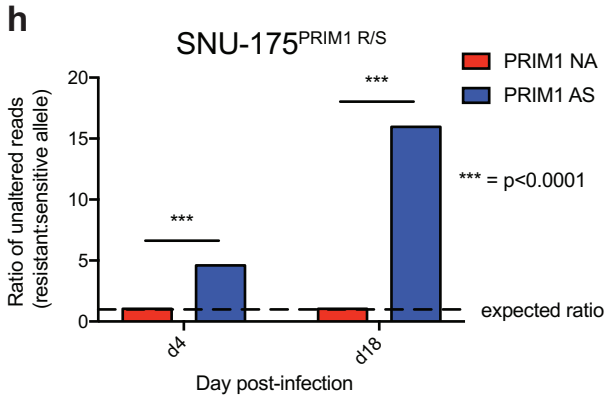


Figure 2 (Continued).







We identified such a SNP in the essential gene *PRIM1* as a promising candidate for proof-of-principle validation. *PRIM1* encodes the catalytic subunit of DNA primase and has previously been determined to be an essential gene[155,158,165]. It contains two common SNPs, of which one (rs2277339) leads to a change in the amino acid sequence: a T to G substitution resulting in conversion of an aspartate on the protein surface to an alanine (Figure 2B and 2C; Supplementary Figure 2A). The minor allele is common (minor allele frequency = 0.177), leading to heterozygosity at this locus in 29% of individuals represented in the ExAC database. This locus also undergoes frequent LOH. Across the 33 cancer types profiled, LOH was observed at the rs2277339 locus in 9% of cancers, including 21% of lung adenocarcinomas, 18% of ovarian cancers, and 17% of pancreatic cancers (Supplementary Figure 2B).

PRIM1^{rs2277339} lies in a polymorphic PAM site—the 'CRISPR-sensitive,' G allele generates a canonical *S. pyogenes* Cas9 PAM site, while the 'CRISPR-resistant,' T allele disrupts the NGG PAM motif. We tested allele-specific *PRIM1* disruption using an allele-specific (AS) CRISPR single guide RNA (sgRNA) designed to target only the G allele at rs2277339, encoding the alanine version of the protein (Figure 2B). In the context of CRISPR experiments, because the G allele should be *sensitive* to allele-specific disruption, we use an 'S' to designate cells with this allele and an 'R' to designate cells with the other, *resistant* allele: for example, PRIM1$^{S/-}$ and PRIM1$^{R/S}$ genotypes reflect cells with one copy of the sensitive G allele and cells with one copy of each allele, respectively.

We transduced four patient-derived cancer cell lines that naturally exhibit either allele with AS sgRNA and verified that AS sgRNA disrupts *PRIM1* in PRIM1^{S} genetic

contexts (Figure 2D). PRIM1$^{S/-}$ and PRIM1$^{S/S}$ cells expressing AS sgRNA show decreased proliferation relative to LacZ-targeting control, whereas cells retaining the resistant allele (PRIM1$^{R/-}$, PRIM1$^{R/R}$, or PRIM1$^{R/S}$) show no such defects (Figure 2E, Supplementary Figure 2C–F).

The specificity of the AS sgRNA for PRIM1^{S} cell lines was not due to a lack of Cas9 activity or *PRIM1* essentiality in the PRIM1^{R} cell lines. We confirmed this finding by transducing four cell lines with a non-allele specific (NA) *PRIM1*-targeting sgRNA. We successfully ablated PRIM1 expression in all contexts (Figure 2D), and cells expressing *PRIM1*-targeting sgRNA showed dramatic and significant decreases in proliferation relative to *LacZ*-targeting control even in cases where expression of the AS sgRNA did not significantly limit growth ($p<0.01$ in all cases, one-tailed Student's t-test; Figure 2E, Supplementary Figure 2C–F).

We further tested isogenic cell lines harboring either allele. Using SNU-175 and SNU-C4 cells, which are heterozygous for *PRIM1*^{rs2277339}, as a base, we transiently transfected a vector expressing Cas9 and two sgRNAs that flank the *PRIM1* gene. We then screened single cell clones for *PRIM1* deletion by PCR. Among deletion-positive clones, we identified heterozygous (PRIM1$^{R/S}$), hemizygous sensitive (PRIM1^{S}), and hemizygous resistant (PRIM1^{R}) lines (Supplementary Figure 3A–E). Using these isogenic cells, we confirmed PRIM1$^{S/-}$ cells expressing AS sgRNA show decreased proliferation relative to LacZ-targeting control, whereas cells retaining the resistant allele (PRIM1$^{R/-}$ or PRIM1$^{R/S}$) show no such defects (Figure 2F, Supplementary Figure 2G–K).

Within these isogenic lines, we also confirmed that AS CRISPR disrupts *PRIM1* in a *PRIM1*^{rs2277339}-dependent manner using deep sequencing. Four days post-infection

with the NA sgRNA, isogenic PRIM1^{S} and PRIM1^{R} cells showed comparable fractions of disrupted alleles (Figure 2G), suggesting both lines exhibit similar levels of Cas9 activity. However, while PRIM1^{S} cells expressing AS sgRNA showed nearly 40% disrupted alleles, resistant cells under the same condition showed 0 disrupted alleles ($p<0.0001$, Chi-square with Yates correction; Figure 2G). This result confirms that AS PRIM1 sgRNA targets PRIM1 in a SNP-specific manner.

We also verified that allele-specific inactivation of essential genes is possible in a heterozygous context. Gene-disrupting indels introduced by the error-prone non-homologous end joining (NHEJ) repair pathway make distinguishing the original genotype (S or R) of an edited allele challenging. Therefore, we compared the number of unaltered reads of each allele in heterozygous cells expressing either NA or AS sgRNA. If the sgRNA disrupts *PRIM1* in a non-allele specific fashion, we would expect a ratio of 1 between unaltered reads of each allele. We infected a Cas9-stable PRIM1$^{R/S}$ line with NA or AS sgRNA as described above and sequenced the target loci 4 and 18 days post-infection. As expected, we saw no substantial difference in the number of unaltered reads between the two alleles in PRIM1$^{R/S}$ cells expressing NA sgRNA (Figure 2H). In contrast, PRIM1$^{R/S}$ cells expressing AS sgRNA showed significantly more unaltered reads from the resistant allele compared to the sensitive allele, with a ratio that increased over time ($p<0.0001$, Chi-square with Yates correction). We conclude that the AS sgRNA disrupts *PRIM1* in a SNP-specific manner even in a heterozygous context.

Validation of EXOSC8^{rs117135638} *as a GEMINI vulnerability*

We also performed proof-of-principle validation for another candidate SNP in the essential gene *EXOSC8*. *EXOSC8* codes for Rrp43, a component of the RNA exosome.

The RNA exosome is an essential multi-protein complex involved in RNA degradation and processing, including processing of pre-rRNA[167,170,187]. Two common SNPs have been described within *EXOSC8*, one of which (rs117135638) represents a C to A change in DNA sequence; this SNP leads to a proline to histidine substitution on the interface between Rrp43 and exosome complex member Mtr3 (Figure 3A and 3B; Supplementary Figure 4A). This candidate SNP is heterozygous in 2% of individuals and undergoes LOH in 29% of cancers, including 72% of lung squamous cell carcinomas, 62% of ovarian cancers, 46% of lung adenocarcinomas, and 40% of breast cancers (Supplementary Figure 4B).

We first tested allele-specific *EXOSC8* disruption using allele-specific (AS) RNA sgRNAs designed to target only the C allele at rs117135638, encoding the proline version of the protein. We designated cells as EXOSC8^{S} (for 'sensitive') if they contained this allele, and as EXOSC8^{R} (for 'resistant') if they contained the A allele. We transduced four patient-derived cancer cell lines with AS sgRNA and verified that AS sgRNA disrupts EXOSC8 protein expression in EXOSC8^{S} genetic contexts (Figure 3C). EXOSC8$^{S/S}$ and EXOSC8$^{S/–}$ cells expressing AS sgRNA showed decreased proliferation relative to LacZ-targeting control, whereas cells retaining the resistant allele (EXOSC8$^{R/–}$ or EXOSC8$^{R/R}$) showed no such defects (Figure 3D, Supplementary Figure 4C–F).

We next verified that the specificity of the AS sgRNA for EXOSC8^{S} cell lines was not due to a lack of Cas9 activity or *EXOSC8* essentiality in the EXOSC8^{R} cell lines. We transduced all four cell lines with a non-allele specific (NA) targeting sgRNA, which successfully ablated EXOSC8 expression in all contexts (Figure 3C). Cells expressing *EXOSC8*-targeting sgRNA showed significant decreases in proliferation relative to

Figure 3: Validation of *EXOSC8*^{rs117135638} as a GEMINI vulnerability.

a. Schematic of *EXOSC8* SNP rs117135638 locus showing target sites for positive control, non-allele specific (NA) sgRNA and experimental, allele-specific (AS) sgRNA. Alleles appear in bold. **b.** Crystal structure of *EXOSC8* gene product, Rrp43^{285} (grey) shows the amino acid encoded by rs117135638 (teal) lies on the surface of the Rrp43 protein near the interface with exosome complex subunit Mtr3 (pink). **c.** Immunoblot of EXOSC8 protein levels in indicated patient-derived and isogenic cell lines expressing LacZ, EXOSC8 NA, or EXOSC8 AS sgRNA. **d.** Representative growth curves of indicated patient-derived and isogenic cell lines expressing LacZ (black), EXOSC8 NA (red), or EXOSC8 AS (blue) sgRNA, as measured by CellTiter-Glo luminescence, relative to day of assay plating. n = 5 technical replicates; error bars represent s.d. See Supplementary Figure 4 for additional biological replicates. **e.** Immunoblot of EXOSC8 protein levels in indicated isogenic cell lines expressing LacZ, EXOSC8 NA, or EXOSC8 AS sgRNA. **f.** Representative growth curves of indicated isogenic cell lines expressing LacZ (black), EXOSC8 NA (red), or EXOSC8 AS (blue) sgRNA, as measured by CellTiter-Glo luminescence, relative to day of assay plating. n = 5 technical replicates; error bars represent s.d. See Supplementary Figure 4 for additional biological replicates.

Figure 3 (Continued). g. Disruption of *EXOSC8* in patient-derived *EXOSC8* resistant ($EXOSC8^{R}$) or *EXOSC8* sensitive ($EXOSC8^{S}$) cells expressing EXOSC8 non-allele specific (NA) positive control sgRNA or allele-specific (AS) experimental sgRNA. Non-inactivated alleles (representing unaltered alleles or alleles with in-frame insertions or deletions, black) and alleles with frameshift alterations (red) were assessed by deep sequencing of *EXOSC8* four days post-infection with sgRNA.

Figure 3 (Continued).

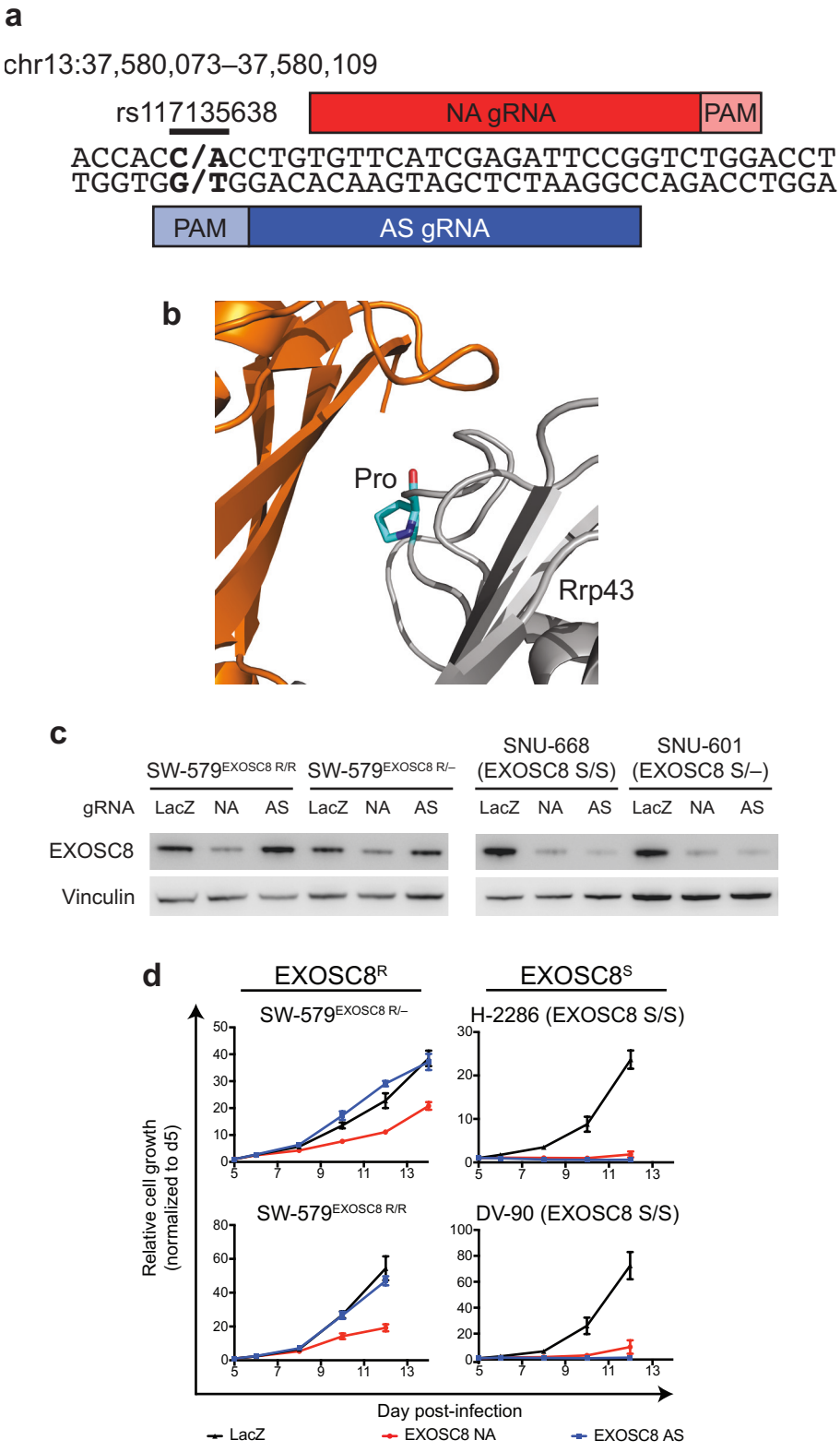
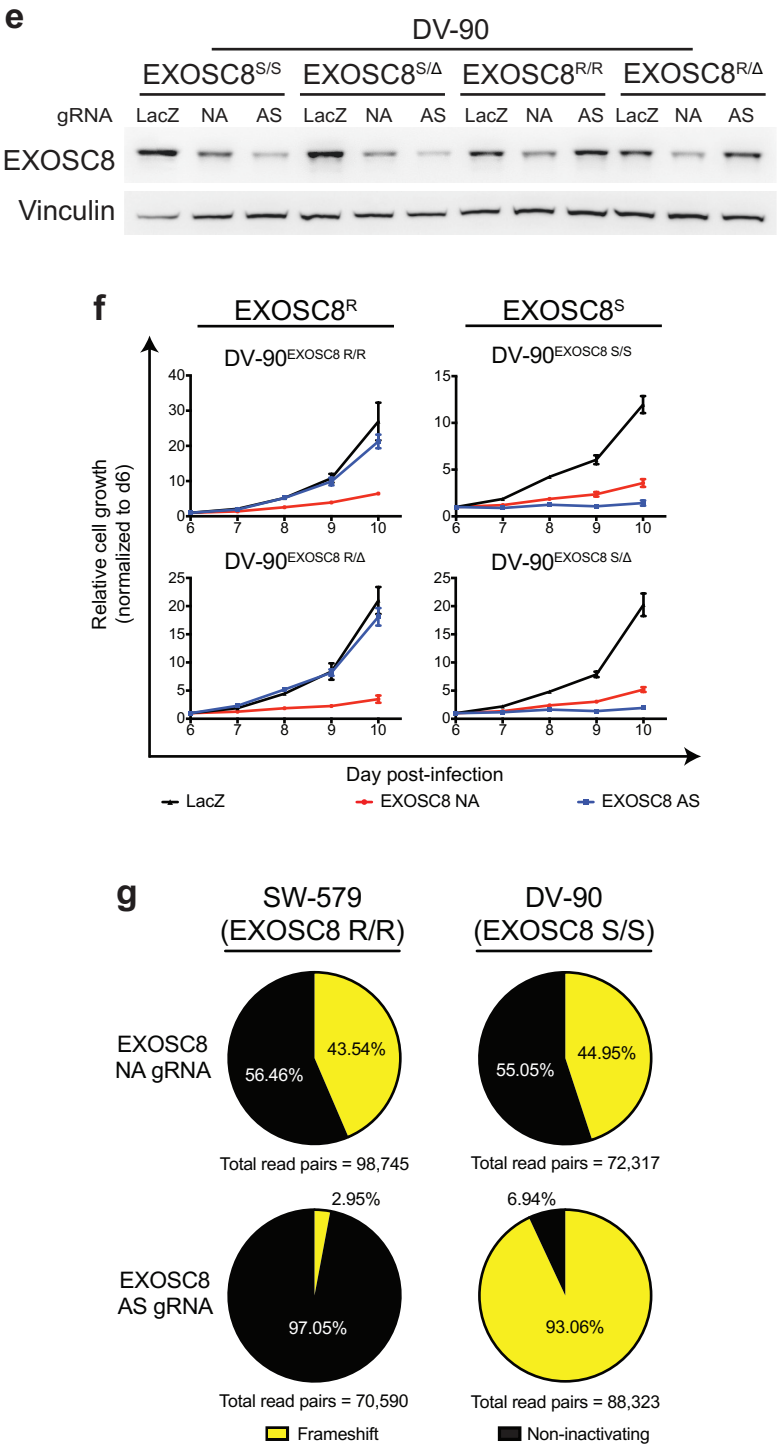


Figure 3 (Continued).



LacZ-targeting control ($p<0.01$ in all cases, one-tailed Student's t-test), confirming that this gene is cell-essential (Figure 3D, Supplementary Figure 4C-F).

We also determined that both copy-loss and copy-neutral LOH of *EXOSC8* represents a vulnerability in an isogenic context. We generated diploid and single-copy knockout isogenic cells representing $EXOSC8^{S}$ and $EXOSC8^{R}$ genotypes by Cas9-mediated homology-directed repair (HDR) editing (Methods; Supplementary Figure 3F–I), and then infected these isogenic Cas9-stable lines with constructs expressing EXOSC8 NA or AS sgRNA. As expected, EXOSC8 NA sgRNA ablated EXOSC8 expression in all contexts, while AS sgRNA ablated EXOSC8 expression only in $EXOSC8^{S}$ cells (Figure 3E). $EXOSC8^{S/S}$ and $EXOSC8^{S/\Delta}$ cells expressing AS sgRNA showed decreased proliferation relative to *LacZ*-targeting control, whereas cells retaining the resistant allele ($EXOSC8^{R/R}$ and $EXOSC8^{R/\Delta}$) showed no such defects (Figure 3F, Supplementary Figure 4G–J).

We also confirmed that AS CRISPR disrupts *EXOSC8* in a SNP-dependent manner using deep sequencing. Four days post-infection with the NA sgRNA, SW-579 (EXOSC8 R/R) and DV-90 (EXOSC8 S/S) cells showed approximately equal fractions of disrupted alleles (Figure 3G). However, while DV-90 sensitive cells expressing AS sgRNA showed upwards of 90% disrupted alleles, SW-579 resistant cells under the same condition showed fewer than 3% disrupted alleles, a significant difference ($p<0.0001$, Chi-square with Yates correction). This result confirms that EXOSC8 AS sgRNA targets *EXOSC8* in a SNP-specific manner.

Discussion

While we here exploited SNPs residing in polymorphic PAM sites to functionally validate *PRIM1* and *EXOSC8* as GEMINI genes, validating the full range of GEMINI vulnerabilities will require additional experimental strategies. 77% of GEMINI variants do not reside in differentially targetable *S. pyogenes* PAM sites (Chapter 4), but *S. pyogenes* Cas9 enzymes with altered[188] or expanded[189,190] PAM specificities or CRISPR effectors from other species[188,191–193] may increase the number of variants that are easily targetable. Variants in the seed region of sgRNAs have also been shown to disrupt normal Cas9 DNA cleavage[186], allowing for allele-specific gene editing[194,195]. Thus, some GEMINI vulnerabilities residing in sgRNA target sites may prove amenable to functional validation through allele-specific gene knockout. Studies attempting to leverage SNP-specific CRISPR disruption should bear in mind that nuclease activity and frameshift frequency can vary according to the nucleotide sequence and chromatin state of the target site[196–199]. Variants not targetable by allele-specific Cas9 disruption due to low sgRNA activity or incompatible sequence context could be experimentally validated through an allele-specific degron approach[200].

Another hurdle to validating the full range of GEMINI vulnerabilities is developing workable cell line models. For more common SNPs (such as *PRIM1*^{2277339}), it may be easy to locate patient-derived cell lines modeling both germline heterozygosity and somatic LOH of the desired allele. For less common variants, appropriate cell lines may be more difficult to obtain (e.g., CCLE contains no lines heterozygous for *EXOSC8*^{rs117135638}). In such cases, cell line models may need to be engineered using gene editing approaches. HDR-mediated editing using CRISPR effectors has become easier as guidelines for the use of this technology have been developed (e.g., recommended design

of repair templates)[201]. Additionally, some GEMINI variants may be amenable to engineering using a base-editor approach[202,203].

Validation or treatment of GEMINI vulnerabilities could benefit from the underlying structure of human genetic diversity and CNAs. Multiple SNPs in the same haplotype block could be tested using a large-scale, allele-specific CRISPR screen against patient-derived cell lines of the appropriate genotypes. Identification of lines with LOH of a targeted haplotype is likely due to the enrichment among deletion events of those affecting whole chromosomes or chromosome arms[43]. Additionally, isogenic models of arm-level deletions have been successfully generated[44,204]; these models or others produced using similar techniques could be used to assess multiple GEMINI variants in a high-throughput manner. Such a multiplexed approach to targeting GEMINI vulnerabilities could have therapeutic applications as well. For instance, it may be possible to develop allele-specific genetic inhibitors targeting multiple variants in a haplotype block that has undergone LOH. Such a combination strategy could enhance the effectiveness of a GEMINI-targeting therapy while decreasing the likelihood of resistance emerging.

Methods

Cell line identification and cell culture

Human cancer cell lines of the appropriate genotypes for *PRIM1* and *EXOSC8* were identified using whole exome sequencing and absolute gene copy number data from the Cancer Cell Line Encyclopedia (https://portals.broadinstitute.org/ccle)[205]. All lines were genotyped for the SNP of interest using Sanger sequencing. Cell lines were maintained in RPMI-1640 supplemented with 10% fetal bovine serum and 1% penicillin.

Lines were not assessed for contamination with mycoplasma. No commonly misidentified cell lines defined by the International Cell Line Authentication Committee have been used in these studies.

Plasmids

lentiCas9-Blast (Addgene plasmid # 52962) and lentiGuide-Puro (Addgene plasmid # 52963) were gifts from Feng Zhang[206]. A Cas9 construct co-expressing GFP and two sgRNAs was a gift from Peter Choi[45]. pLKO.1–TRC cloning vector was a gift from David Root (Addgene plasmid # 10878)[207].

CRISPR sgRNAs

To identify target sites for CRISPR-Cas9-mediated knockout, the genetic sequences of *PRIM1* and *EXOSC8* were obtained from the UCSC genome browser (http://genome.ucsc.edu) using the human assembly GRC38/Hg38 (December 2013). The 20 nucleotides upstream of the polymporphic PAM site containing the SNP for each gene constitutes the AS sgRNA for that gene. All other sgRNAs were designed using the CRISPR sgRNA design tool from the Zhang lab (http://crispr.mit.edu). sgRNAs were cloned into the appropriate vector as described previously[206,208]. The sgRNA sequences were as follows:

LacZ: GTTCGCATTATCCGAACCAT

PRIM1 AS: CAGCTCGGGCAGCTCGGTGG

PRIM1 NA: CGCTGGCTCAACTACGGTGG

EXOSC8 AS: CGGAATCTCGATGAACACAG

EXOSC8 NA: ACCGGAATCTCGATGAACAC

Cell growth assays

Cells were plated in opaque 96-well plates at 500 or 1000 cells per well (Corning) on the indicated day post-lentiviral infection. Cell number was inferred by ATP-dependent luminescence using CellTiter-Glo (Promega) reagent and normalized to the relative luminescence on the day of plating.

Generation of PRIM1-loss and EXOSC8-loss cell lines

A Cas9 construct co-expressing GFP and two sgRNAs with target sites flanking *PRIM1* was used to delete a 20.6kb region encoding PRIM1. Cell lines heterozygous for *PRIM1*^{rs2277339} (SNU-C4, SNU-175, and TYK-nu) were transfected with this construct using LipoD293 transfection reagent (SignaGen), and single GFP+ cells were sorted by FACS and plated at low density for single cell cloning or single-cell sorted into 96-well tissue culture plates containing a 50:50 mix of conditioned and fresh RPMI-1640 media, 20% serum, 1% penicillin-streptomycin, and 10μm ROCK inhibitor Y-27632. Clones were expanded and validated by PCR to harbor the 20.6kb deletion encoding *PRIM1*, and the retained allele was genotyped by Sanger sequencing. These clones were designated SNU-175$^{PRIM1\ S/–}$, SNU-175$^{PRIM1\ R/–}$, SNU-C4$^{PRIM1\ S/–}$ for subsequent experiments. Other clones were determined by PCR and Sanger sequencing to retain both *PRIM1* alleles and not to harbor this deletion and were designated as control cell lines for subsequent experiments (SNU-175$^{PRIM1\ R/S}$ and SNU-C4$^{PRIM1\ R/S}$). The same procedure was employed using a cell line diploid for the EXOSC8^{R} SNP (SW-579) to generate EXOSC8$^{R/–}$ cell lines harboring a 7.1kb deletion and EXOSC8$^{R/R}$ control lines.

The sgRNA sequences were as follows:

PRIM1 upstream: GCGCGGAACTCGCCACGGTA

PRIM1 downstream: CAGAGCTCCTCAAACCATTG

EXOSC8 upstream: GGTTTCTCGGCCGAGCGCCG

EXOSC8 downstream: TGTACCCATCTACTTAAGTT

Primers used to verify gene deletion by PCR were as follows:

PRIM1 deletion genotyping F: ACTGTATGCACCACCACACC

PRIM1 deletion genotyping R: AGTTCACGTGGAGCATCCTT

EXOSC8 deletion genotyping F: TTTGGGGCATACTCATGCTT

EXOSC8 deletion genotyping R: TCCACCTCCAATTATTTGTTCC

Generation of EXOSC8 isogenic cell lines

Cas9 RNPs and a ssODN repair template were used to edit the EXOSC8^{S} SNP to the EXOSC8^{R} SNP. S. pyogenes Cas9-NLS (Synthego) and an sgRNA (sequence: ACCGGAATCTCGATGAACAC) targeting the *EXOSC8* SNP region (Synthego) were complexed as described previously[209]. DV-90 cells (EXOSC8$^{S/S}$) were nucleofected with resulting RNPs, a 50:50 mix of EXOSC8^{S} and EXOSC8^{R} ssODN (IDT), and a GFP-expressing plasmid (pMAX-GFP) (Lonza). ssODN repair templates contained a synonymous mutation introducing a novel *Mnl1* restriction site for downstream genotyping as well as a silent blocking mutation to prevent repeated Cas9 cleavage. Single GFP+ cells were single-cell sorted by FACS into 96-well tissue culture plates containing a 50:50 mix of conditioned and fresh RPMI-1640 media, 20% serum, 1% penicillin-streptomycin, and 10μm ROCK inhibitor Y-27632. Clones were expanded and evaluated for HDR-mediated editing by PCR and restriction digest, and positive clones were genotyped by next-generation sequencing (MGH DNA Core).

CRISPR variant sequencing

Cellular pellets were collected from Cas9-stable cells 4 or 18 days post-infection with lentiGuide-Puro virus encoding the indicated sgRNA. Genomic DNA was isolated using a DNAMini kit (Qiagen), and the target region for each gene was amplified by PCR (EMD Millipore). Amplicons were submitted to NGS CRISPR sequencing by the MGH DNA Core. Frameshift and non-inactivating alleles (nonaltered or in-frame indels) were determined manually using the CRISPR variant output file. PCR primer sequences were as follows:

PRIM1 MGH F: GCACAGAAGGCGCTTCATA

PRIM1 MGH R: CGCCAATTCCTGTGGTAATC

EXOSC8 F: AGCTGCAGAGTGTTTCTTTCA

EXOSC8 R: AGAGCAAAGTAAATGAAAAGCCCAA

Western blotting

Cells were washed in ice-cold PBS and lysed in 1x RIPA buffer (10 mM Tris-Cl Ph 8.0, 1 mM EDTA, 1% Triton X-100, 0.1% SDS and 140 mM NaCl) supplemented with 1x protease and phosphatase inhibitor cocktail (PI-290, Boston Bioproducts). Lysates were sonicated in a bioruptor (Diagenode) for 5 min (medium intensity) and cleared by centrifugation at 15,000 x g for 15 min at 4°C. Proteins were electrophoresed on polyacrylamide gradient gels (Life Technologies) and detected by chemiluminescence (Bio-rad).

Antibodies used were as follows:

EXOSC8: Proteintech #11979-1-AP

PRIM1: Cell Signaling Technology #4725

Vinculin: Sigma #V9131

This page intentionally left blank.

CHAPTER 4: POTENTIAL APPROACHES TO TARGETING GEMINI VULNERABILITIES

This page intentionally left blank.

CHAPTER 4: POTENTIAL APPROACHES TO TARGETING GEMINI VULNERABILITIES

The major challenge to exploiting GEMINI vulnerabilities is identifying means to target them in humans. Different GEMINI vulnerabilities may require different therapeutic approaches to exploit them, due to the location of the variant within each gene and its effects on the amino acid composition of the protein. These differences have not been explored. Three therapeutic approaches that may be contemplated are DNA-targeting CRISPR effectors (e.g., Cas9), RNA-targeting approaches (e.g., RNAi), and allele-specific small molecules. We characterized each GEMINI vulnerability according to criteria that would indicate its amenability to targeting by each of these approaches. We also defined the potential scope of patients that could benefit from therapeutic approaches targeting each GEMINI vulnerability.

Results

For each GEMINI variant, we calculated the number of new patients per year in the US that exhibit LOH of the hypothetical 'targetable' allele (Methods). Across the 33 tumor types we profiled, the median GEMINI vulnerability could be targetable in 17,747 patients per year (Figure 4A). *PRIM*^{rs2277339} and *EXOSC8*^{rs117135638} could be targetable in a theoretical 22,470 and 5,307 patients per year, respectively (Supplementary Figure 2A and 4A).

We next sought to prioritize the compendium of GEMINI vulnerabilities identified in Chapter 1 by potential druggability by a variety of therapeutic approaches. First, we analyzed the list of GEMINI vulnerabilities to identify polymorphisms whose targeting on the DNA level may enable allele-selective gene disruption by a

Figure 4: Potential approaches to targeting GEMINI vulnerabilities.

a. Number of GEMINI variants (vertical axis) plotted against the number of patients per year in the US whose tumors might respond to therapeutics targeting those variants (i.e., have lost the resistant allele from a heterozygous germline; horizontal axis). Bin width = 1000 patients. **b.** Growth of heterozygous and hemizygous cells expressing positive control, non-allele specific *PRIM1*-targeting shRNAs versus *PRIM1* mRNA expression. Cell growth measured by CellTiter-Glo luminescence relative to day 2 post-infection and shGFP (n = 5 technical replicates). *PRIM1* mRNA expression assessed by qRT-PCR (n = 3 technical replicates). Dashed grey line indicates *PRIM1* expression threshold below which substantial decreases in cell viability are observed. **c.** Summary table representing challenges to developing allele-specific small molecules that target GEMINI vulnerabilities and associated analyses to prioritize targets. **d.** Crystal structure of *ACTR8* gene product[217] shows the amino acid encoded by rs3733082 (magenta) in the mouth of the ATP-binding pocket of Arp8 (grey). ATP shown in green. **e.** Crystal structure of the *DDX19A* gene product[286] shows the amino acid encoded by rs1055783 (yellow) near the ATP-binding pocket of DDX19A (grey). ADP shown in green.

Figure 4 (Continued).

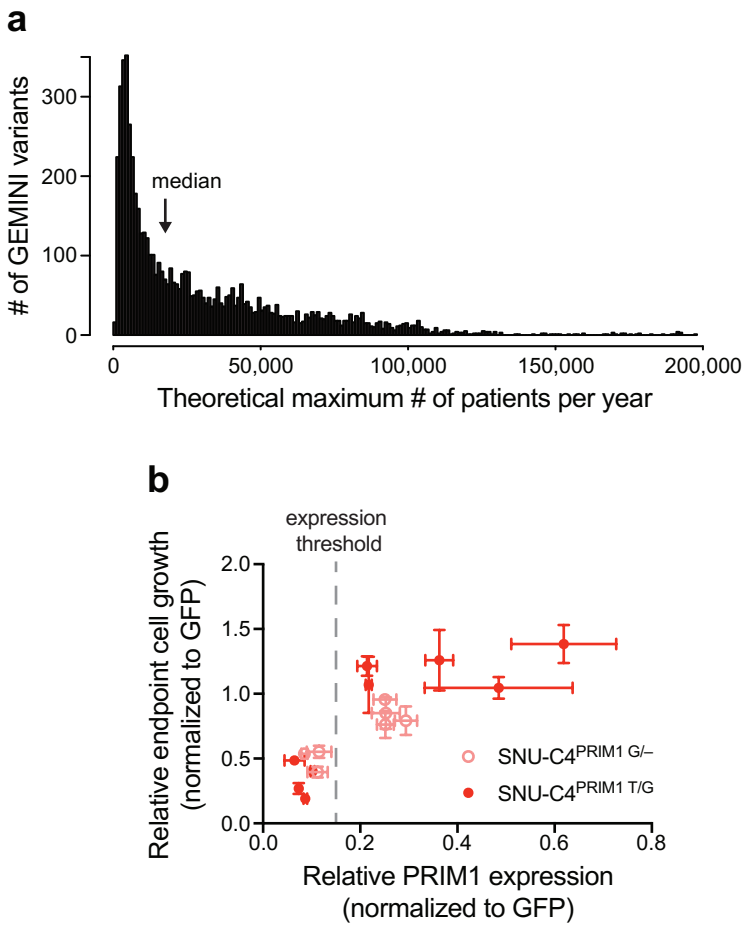


Figure 4 (Continued).

c

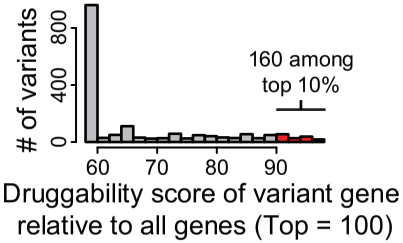
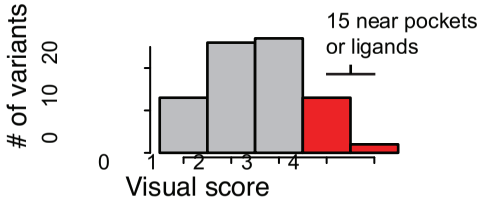
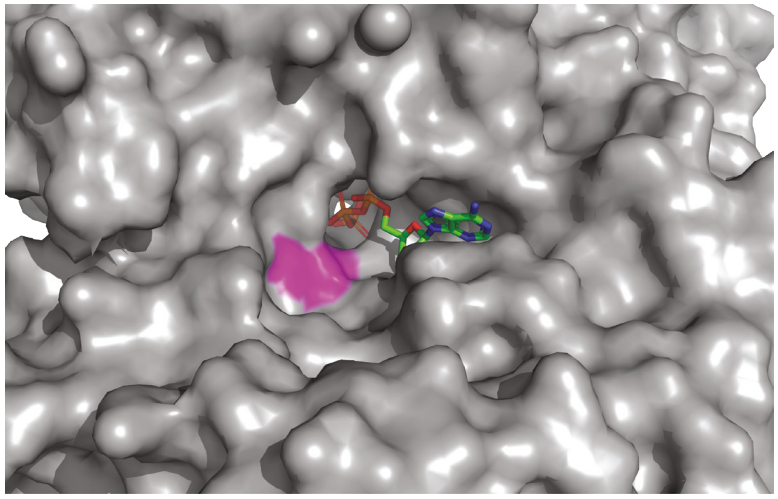
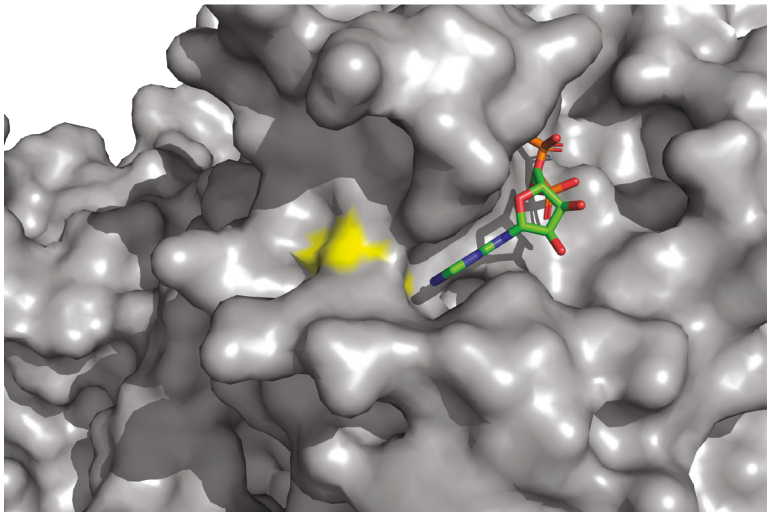
Challenge	Approach	Analysis
Is the protein amenable to inhibition?	canSAR structure-based druggability	212 variants are in proteins with ligand-bound structures 674 are in proteins predicted to be druggable/tractable
	canSAR ligand-based druggability	
	Enzyme Commission annotation	441 variants are in enzymes
Does the polymorphism reside in a region amenable to small-molecule binding?	p-blast and visual score of 81 variants for which homologous X-ray structures exist	
Is the variant amenable to allele-specific small-molecule binding?	Residue charge alteration	584 variants alter residue charge
	Potential for covalent inhibition	95 variants introduce a cysteine

Figure 4 (Continued).

d



e



CRISPR-based approach. For this analysis, we included both the canonical *S. pyogenes* PAM, NGG, as well as the weaker, non-canonical PAM, NAG[186,210]. Of the 4648 GEMINI vulnerabilities in open reading frames, 23% (1088/4648, or 19% of all GEMIMI vulnerabilities) generate a PAM site in one allele but not the other, suggesting the potential for allele-specific knockout (Supplementary Table 3).

In theory, every GEMINI variant could be the target of allele-specific RNAi reagents. However, it is possible that, for some GEMINI variants, RNAi reagents would be unable to suppress expression sufficiently to reduce cell viability, or that sufficient allelic specificity might not be achieved. For example, we tested the hypothesis that *PRIM1*^{rs2277339} may be targetable in an allele-specific manner using RNAi. For these experiments, we refer to the *PRIM1* alleles by their identifying nucleotide; for example, heterozygous cells are referred to as PRIM1$^{T/G}$, and cells hemizygous for the minor allele are referred to as PRIM1$^{G/–}$. We first sought to determine the level of PRIM1 knockdown required to substantially decrease cell proliferation. Accordingly, we infected hemizygous and heterozygous isogenic cells with non-allele specific *PRIM1*-targeting shRNAs and assessed *PRIM1* expression and cell growth. We observed that substantial decreases in cell growth were possible under conditions of robust *PRIM1* knockdown (>85%) (Figure 4B).

We then asked whether allele-specific shRNAs targeting the *PRIM1*^{rs2277339} locus could decrease growth in cells representing the fully matched genotype. PRIM1$^{T/–}$ and PRIM1$^{G/–}$ cells were infected with constructs encoding fully complementary shRNAs tiling across the SNP and assessed for growth. Only one shRNA, shG7 (targeting the minor, G allele at position 7) significantly reduced cell growth relative to GFP-targeting

control (Supplementary Figure 5A–B). We then selected the four *PRIM1* SNP-targeting shRNAs that yielded the lowest average cell growth relative to GFP-targeting control and assessed their ability to decrease cell growth in an allele-specific manner. Heterozygous cells ($PRIM1^{T/G)}$ and hemizygous cells of the targeted genotype ($PRIM1^{T/–}$ or $PRIM1^{G/–}$) were infected with constructs encoding the appropriate shRNA. No putative allele-specific shRNAs were found to significantly decrease cell growth in hemizygous cells of the targeted genotype relative to heterozygous cells (Supplementary Figure 5C–D). We conclude that $PRIM1^{rs2277339}$ may not represent an optimal candidate for allele-specific shRNA-mediated inhibition.

Given the large number of additional GEMINI variants that may be suitable for RNAi-mediated targeting, we sought to prioritize GEMINI genes that may be amenable to allele-specific inhibition using mRNA-targeting approaches. RNAi-mediated knockdown of some essential genes may be more effective at inducing cell death than others, based on differential expression thresholds required for cell survival[43,45,117]. We hypothesized that GEMINI genes representing strong dependencies in shRNA screens would be most amenable to potential targeting using an RNAi-based therapeutic. We therefore analyzed shRNA data representing 17,212 genes in 712 cell lines[211], including 1183 GEMINI genes, and looked for genes whose suppression led to at least a moderately strong response in most of the cell lines (median DEMETER2 score < -0.5; Methods). Approximately 35% of GEMINI genes (413/1183), including PRIM1 (median DEMETER2 score = -0.52), fit this category, representing 35% (1804/5196) of GEMINI vulnerabilities (Supplementary Table 3). In comparison, only 3.6% of all genes profiled (623/17,212) passed this dependency threshold, indicating a significant enrichment for

GEMINI genes ($p < 0.0001$, binomial proportion test). However, this level of dependency was not observed for all GEMINI or essential genes. For example, the median DEMETER2 score for *EXOSC8* was only -0.14, despite our and others' extensive data showing its essentiality in multiple cell types[170]. These results raise the possibility that RNAi-based approaches may not be able to exploit many GEMINI vulnerabilities.

Both CRISPR- and RNAi-based therapeutic approaches suffer from difficulties in effectively delivering reagents to all cancer cells in an animal. Small molecule–based approaches can overcome such delivery issues, but substantial obstacles exist to developing allele-specific small molecules that target GEMINI vulnerabilities. These challenges include identifying GEMINI genes that are amenable to small-molecule inhibition, determining which GEMINI variants lie near potentially druggable pockets, and predicting which GEMINI variants are most likely to facilitate allele-specific drug binding. We therefore analyzed 1749 protein-altering GEMINI vulnerabilities (missense, insertion, and deletion variants) to prioritize candidates for potential allele-specific drug development (Supplementary Table 5).

To identify GEMINI genes that may be amenable to small-molecule inhibition, we annotated those containing protein-altering alleles using the canSAR Protein Annotation Tool[212,213] (cansar.icr.ac.uk; Methods). This tool uses publicly available structural and chemical information to generate structural and ligand-based druggability scores. While these scores do not necessarily reflect potential for allele-specific small molecule inhibition based on the GEMINI variant of interest, they may nonetheless allow prioritization of targets based on general druggability. This analysis found that of the 1734 protein-altering variants in genes assessed by canSAR, 12% (212) reside in proteins

with a small molecule ligand–bound structure (Figure 4C). Additional assessments of potential small-molecule binding sites on structures with and without existing ligands indicated that 39% of protein-altering variants (674) lie in proteins with molecular structures that are predicted to be druggable (drug-like compound modulates activity in vivo) or tractable (tool compound modulates activity in vitro) (Figure 4C). Furthermore, 160 GEMINI variants reside in proteins in the top 90th percentile of ligand-based druggability as assessed by the physiochemical properties of small molecules tested against the protein or its homologs (Figure 4C). We also found that 25% of protein-altering GEMINI variants (441/1734) reside in enzymes as defined by their annotation with an Enzyme Commission (EC) number[214] (Figure 4C).

To assess which variants may reside in protein regions amenable to small molecule binding, we performed a p-blast of the 1749 protein-altering variants against protein sequences for molecular structures found in the Protein Data Bank[215] (rcsb.org; Methods). This analysis identified 153 variants characterized in a homologous structure. We then visually scored 81 missense and indel variants in X-ray crystal structures for their proximity to solvent-exposed pockets or known small-molecule binding sites using a scale of 0 to 4 (Methods). Of the variants analyzed, 15 were near a potential binding pocket on the surface of the protein (score = 3), with two of these pockets containing a small-molecule ligand (score = 4) (Figure 4C).

The first variant near a small-molecule ligand lies in the gene *ACTR8*. This gene codes for actin-related protein 8 (Arp8), a member of the INO80 chromatin remodeling complex with roles in DNA repair, homologous recombination, and mitotic chromosome alignment[216]. The *ACTR8* variant of interest (rs3733082) is heterozygous in 15% of

people and undergoes LOH in 29% of cancers (Supplementary Table 3). *ACTR8*^{rs3733082} leads to a threonine to isoleucine substitution at the mouth of the ATP-binding pocket (Figure 4D), which exhibits basal ATPase activity[217] and may play a role in regulating Arp8 binding to DNA[218]. While *ACTR8* exhibits strong evidence of essentiality in CRISPR knockout screens (Supplementary Table 1), previous functional evidence in human cell lines and yeast suggests that it may not be cell-essential. Thus, additional studies into the essentiality of *ACTR8* and the role of its ATP-binding pocket must be performed prior to moving forward with this candidate. Even if the ATPase activity of Arp8 is not required for its function, the polymorphic residue encoded by rs3733082 could represent a target for a proteolysis-targeting chimera (PROTAC) approach.

The second promising variant identified by the visual screen lies in the gene *DDX19A*, which codes for DEAD-box helicase 19A (a homolog of yeast protein Dbp5)[219,220]. This essential ATP-dependent enzyme associates with the nuclear pore and is required for the remodeling of ribonucleoprotein complexes (RNPs) during mRNA nuclear export[221]. The *DDX19A* variant of interest (rs1055783) is heterozygous in 2% of people and undergoes LOH in 24% of cancers. Additionally, this variant generates a proline to leucine substitution near the ATP-binding pocket of Ddx19 (Figure 4E).

We also assessed protein-altering GEMINI variants to prioritize those that may be most amenable for allele-specific small-molecule inhibition. For this analysis, we scored variants that altered the number or sign of residue charges. For example, a variant that changes the charge of a residue from neutral to negative or that adds an additional negative charge through an inserted residue would qualify as a charge-altering variant. Of the 1749 protein-altering variants, 584 induced a change in residue charge (Figure 4C).

We further hypothesized that variants introducing a cysteine residue could provide additional allele selectivity by enabling the potential development of a covalent inhibitor. Among the missense and indel GEMINI variants, 95 generate a cysteine in one allele.

We then integrated each of these analyses to characterize the potential druggability landscape of these protein-altering GEMINI vulnerabilities. Every variant was given a score from 0 to 7 based on the number of analyses in which it scored among the top candidates. One variant, *TGS1*^{rs7823773}, earned a score of 6, including in the visual scoring and cysteine categories. Nine additional variants earned a score of 5 (Supplementary Table 5). These may be among the highest-priority candidates for further exploration.

Discussion

Like any target class, we expect resistance mechanisms to arise in response to targeting GEMINI vulnerabilities. Base pair substitutions that replace the targeted allele with an alternative are likely to occur in one in every 10^8–10^9 cells, given observed mutation rates per cell division[222]. Additional alterations affecting nearby nucleic or amino acids could interfere with genetic (e.g., CRISPR and RNAi) or protein-targeting approaches. Additionally, high-level amplification or overexpression of the allele of the GEMINI gene retained in the tumor could overcome the effects of an allele-specific inhibitor. It is also possible that alternative pathways exist for some GEMINI genes, whereby alterations of other genes compensate for inhibition of a GEMINI gene. However, our list of GEMINI genes is highly enriched for components of universal cellular processes, such as DNA, RNA, and protein biogenesis, for which no alternative pathways exist to compensate their loss[223]. Nonetheless, this potential mode of resistance

could be addressed up-front by performing suppressor screens for candidate genes and prioritizing those least likely to escape inhibition through alterations in other genes[150].

Biomarkers for detection of patients who may benefit from a GEMINI approach are relatively straightforward: one would select patients who are heterozygous for the targeted allele, and for whom the tumor is found to have lost the alternative allele. One consideration is tumor heterogeneity; if the LOH event is present in only part of the tumor, resistance would be expected to arise quickly. However, in prior analyses[222,224–226], a majority of somatic copy-number alterations appeared to be clonal, although the fraction of clonal events can be lower in some loci in some tissues[227]. One approach to minimize clonal variation in LOH is to prioritize GEMINI genes that lie on chromosomes or chromosome arms that are characteristically lost early in oncogenesis (e.g., 3p in renal clear cell carcinoma)[112].

It may also be necessary to consider genomic imprinting of GEMINI genes when prioritizing candidates for potential therapeutic development. Genomic imprinting refers to the epigenetic process of limiting certain genes to monoallelic expression based on allelic parent-of-origin[228]. In theory, a gene imprinted organism-wide or in a tissue-specific manner could also be a GEMINI gene. Tumors undergoing LOH of this GEMINI gene would express only one allele, and therefore be dependent upon it. However, many normal cells may also only express this allele, if the other one happened to be imprinted. Any allele-specific therapy against the non-imprinted allele would then target both tumor and normal cells. Such a situation would complicate the identification of eligible patients by requiring a determination of the imprinting status of the GEMINI gene in normal tissue. Fortunately, this scenario is unlikely to apply to many GEMINI genes. While the

complete list of organism-wide or tissue-specific imprinted genes has yet to be completed, imprinting is thought to affect only a few hundred genes in the human genome and has only been verified in approximately 50 genes[229,230].

Another mechanism of monoallelic expression known as autosomal monoallelic expression (MAE) may have a larger impact on selection of GEMINI vulnerabilities for therapeutic follow-up. Like genomic imprinting, autosomal MAE is mitotically stable; however, patterns of autosomal MAE are not determined by allelic parent-of-origin, but rather are 'clone-specific'[231]. Assessments of MAE across multiple tissue types in human and mouse have indicated that 10–30% of genes may be subject to MAE in some contexts[231–234]. Future efforts to prioritize GEMINI vulnerabilities for potential drug development should consider MAE for reasons similar to those described above for genomic imprinting.

While we show that cells heterozygous for the targeted SNPs of *PRIM1* show no substantial proliferation defects upon ablation of the targeted allele (Figure 3D, 3E), systemic knockout of one allele of an essential gene across all cells in a patient is not likely to be a tractable therapeutic strategy. Thus, potential allele-specific gene editing approaches to leverage GEMINI vulnerabilities in the clinic would rely on a highly cancer cell–specific delivery system to avoid knockout of the targeted allele in normal tissue. While much work remains to achieve the necessary targeting specificity, advances in nanoparticle delivery systems present the possibility of targeting Cas9 DNA, mRNA, or protein in a tumor-specific manner[235–238]. Additionally, *S. pyogenes* Cas9 enzymes with altered[188] or expanded[189,190] PAM specificities or CRISPR effectors from other species[188,191–193] have broadened the total number of targetable loci in the genome and,

thereby, the number of targetable SNPs. However, any CRISPR-based therapeutic strategy will need to consider possible immune responses to Cas9 or other CRISPR effectors[239–241].

GEMINI vulnerabilities could be also targetable by reversible genetic approaches. Early studies of the GEMINI genes *POLR2A* and *RPA1* achieved allele-specific growth suppression of cancer cells using antisense oligonucleotide (ASO) [131,133,242] and RNAi[134] reagents. Since these studies, progress in ASO- and siRNA-based drug development has made oligonucleotide (ON)-based therapies clinically actionable[14]. Advancements in ON chemistry have improved reagent stability (e.g., resistance to nucleases) and on-target affinity while decreasing potential toxicity due to off-target gene knockdown or immune activation[243,244]. Accordingly, ON-based therapies have been approved by the FDA for several indications[245,246], and multiple clinical trials of ON drugs in the treatment of cancer are ongoing[247–249]. Despite these advances, however, delivery of an ON to every cell in a tumor remains challenging. Obstacles to efficient and targeted delivery include factors related to nucleic acid chemistry (e.g., cell membrane permeability) and the nature of tumor microvasculature[243]. Furthermore, systemic delivery of ONs leads to accumulation in the liver, kidneys, and other potentially non-target tissues, while truly tumor- or tissue-specific targeting strategies remain elusive[243]. Nonetheless, novel delivery mechanisms such as antibodies[250]; nucleic acid aptamers[251]; and lipid-, polymer- and inorganic-based nanoparticles[243,252] present opportunities for continued progress in the targeted delivery of ON therapies.

Peptide nucleic acids, or PNAs, represent another potential genetic approach for targeting GEMINI vulnerabilities. PNAs mimic the structure of and ability to duplex with

naturally occurring nucleic acids, enabling them to block transcription or translation of target genes[253–255]. However, the unnatural, uncharged backbone of PNAs makes them less susceptible to degradation[256] and increases the stability of PNA–nucleic acid duplexes relative to duplexes of naturally occurring nucleic acids. PNA–DNA duplexes are also more sensitive to mismatches than are DNA–DNA duplexes[257], suggesting the possibility of their use in allele-specific therapeutics. Such PNA therapeutics would require specialized delivery systems due to the poor cell permeability of PNAs; however, PNA reagents have been successfully delivered in vivo using several technologies[258], including cell-penetrating peptides (CPPs), lipid-based nanoparticles, and the anthrax protective antigen (PA) system[259].

Finally, the RNA-targeting CRISPR effector Cas13 has shown the ability to knock down target genes[260] and decrease proliferation of cancer cells[261] in an allele-specific manner. Unlike Cas9, the Cas13 enzyme from *L. wadei* previously used for allele-specific RNA cleavage does not require a downstream PAM-like motif[260], potentially expanding the number of targetable sites beyond those tractable with DNA-targeting CRISPR effectors. Like Cas9-based modalities, however, CRISPR-Cas13 and other reversible genetic approaches to exploit GEMINI vulnerabilities would require the development of novel delivery systems.

Allele-specific small molecule inhibitors present another attractive possibility for drugging GEMINI vulnerabilities. Allele-specific therapeutics in clinical use include rationally designed drugs (e.g., mutant EGFR inhibitors[262]) as well as those whose genotype-specific effects were identified by pharmacogenomic studies (e.g., warfarin and VKORC1[263]). However, GEMINI vulnerabilities present an additional challenge for

allele-specific inhibitor development: most variants in cell-essential genes do not reside in or near an active site (Figure 5C) or other functionally critical protein region whose binding by a small molecule could inhibit protein function. This challenge may be addressed through alternative small-molecule approaches, such as allosteric inhibitors[264] or proteolysis-targeting chimera (PROTAC)-mediated degradation. PROTACs function by binding both the target protein and E3 ubiquitin ligase, leading to the ubiquitination and subsequent proteasomal degradation of the target protein[265]. This function allows for the selective depletion of target proteins that may not be susceptible to traditional modes of pharmacological intervention, e.g., enzymatic inhibition. SNPs for which one allele is a cysteine could be prioritized for a PROTAC approach because of the possibility of engineering a covalent inhibitor[266]. Additional residues—primarily lysine, but also serine, threonine, tyrosine, methionine, glutamate, and aspartate—have been explored as targets for covalent inhibitors[266]; these residues could also be considered in prioritizing GEMINI vulnerabilities as the ability to target them progresses. Chemoproteomic approaches using reactive fragment–based screening may facilitate discovery of tool compounds for this kind of allele-specific, covalent small-molecule drug development program[267,268].

While we have rigorously validated *PRIM1* and *EXOSC8* as genetic dependencies in cancer, further work is necessary to explore potential therapeutic modalities for targeting them. As an enzyme, PRIM1 represents a potential target for small-molecule drug development. However, while several potential inhibitors of human DNA primase have been proposed, none have yet advanced to clinical stages of development[269]. Efforts to discover additional putative primase inhibitors may benefit from approaches used to target bacterial and viral primases. For instance, a combined fragment-based/virtual

screening approach has been used to identify novel inhibitors of the T7 bacteriophage primase. Such an approach may aid PRIM1 lead compound discovery by eliminating the challenging task of initial primase functional screening for large numbers of compounds[270]. We note that the *PRIM1* variant validated in this thesis, rs2277339, may not be near enough to the active site of PRIM1 to be targetable by a nucleotide-competitive primase inhibitor. Therefore, allosteric or PROTAC approaches should be considered in future efforts to develop an allele-specific small-molecule inhibitor targeting this variant.

The residue coded for by *EXOSC8*^{rs117135638} lies on the interface of *EXOSC8* gene product Rrp43 and exosome complex member Mtr3, raising the possibility of developing an allele-specific inhibitor of exosome complex formation. Pharmacologic inhibition of protein-protein binding has generally proven challenging due to the large, often flat, surfaces involved in protein complex formation[271]. However, disruption of several clinically relevant protein-protein interactions has been achieved previously, as illustrated by small-molecule or peptidomimetic inhibitors of p53-Mdm2 binding[272–274], Notch complex assembly[275], and herpesvirus DNA polymerase complex assembly[276]. As with *PRIM1*, any such approach to targeting LOH of *EXOSC8* must also overcome the substantial hurdle of achieving allele specificity.

Methods

shRNA sequences

pLKO.1 GFP shRNA (target sequence: GCAAGCTGACCCTGAAGTTCAT) was a gift from David Sabatini (Addgene plasmid # 30323)[277]. Lentiviral expression constructs for non-allele specific shRNA-mediated suppression of *PRIM1* were obtained

through the Broad Institute of MIT and Harvard Genomic Perturbation Platform (https://portals.broadinstitute.org/gpp/public/). The names, clone IDs, and target sequences used in our studies are as follows:

shPRIM1 (TRCN0000275194): AGCATCGTCTCTGGGTATATT

TRCN0000151860: CCGAGCTGCTTAAACTTTATT

TRCN0000275194: AGCATCGTCTCTGGGTATATT

TRCN0000275195: GATTGATATAGGCGCAGTATA

TRCN0000275196: CCGAGCTGCTTAAACTTTATT

The following allele-specific shRNA sequences were cloned into the vector pLKO.1 as described previously[207]:

PRIM1^{rs2277339} major-allele (T) targeting:

sh3T: TCAATGGAGACGTTTGACC

sh4T: CAATGGAGACGTTTGACCC

sh5T: AATGGAGACGTTTGACCCC

sh6T: ATGGAGACGTTTGACCCCA

sh7T: TGGAGACGTTTGACCCCAC

sh8T: GGAGACGTTTGACCCCACC

sh9T: GAGACGTTTGACCCCACCG

sh10T: AGACGTTTGACCCCACCGA

sh11T: GACGTTTGACCCCACCGAG

sh16T: TTGACCCCACCGAGCTGCC

PRIM1^{rs2277339} minor-allele (G) targeting:

sh3G: TCAATGGAGACGTTTGCCC

sh4G: CAATGGAGACGTTTGCCCC

sh5G: AATGGAGACGTTTGCCCCC

sh6G: ATGGAGACGTTTGCCCCCA

sh7G: TGGAGACGTTTGCCCCCAC

sh8G: GGAGACGTTTGCCCCCACC

sh9G: GAGACGTTTGCCCCCACCG

sh10G: AGACGTTTGCCCCCACCGA

sh11G : GACGTTTGCCCCCACCGAG

sh16G : TTGCCCCCACCGAGCTGCC

Quantitative and reverse transcription PCR

RNA was extracted using the RNeasy extraction kit (Qiagen) and subjected to on-column DNase treatment. cDNA was synthesized with the Superscript II Reverse Transcriptase kit (Life Technologies) with no reverse transcriptase samples serving as negative controls. Gene expression was quantified by Power Sybr Green Master Mix (Applied Biosystems). *PRIM1* expression values were normalized to vinculin (*VCL*) and the fold change calculated by the DDCt method. Primers used in our studies are as follows:

PRIM1-F: GCTCAACTACGGTGGAGTGAT

PRIM1-R: GGTTGTTGAAGGATTGGTAGCG

VCL-F: CGCTGAGGTGGGTATAGGTG

VCL-R: TTGGATGGCATTAACAGCAG.

Calculations of theoretical patient numbers in the US

To determine number of patients that could benefit from a therapeutic approach targeting each GEMINI vulnerability, we used the following formula:

$\kappa \times \chi \times \lambda \times 0.5 = \Pi$

in which

$\kappa =$ # new pan-cancer cases per year in the US (1,735,350)

$\chi =$ rate of heterozygosity of GEMINI variant

$\lambda =$ pan-cancer rate of LOH of GEMINI gene

$0.5 =$ fraction of patients with LOH that undergo loss of theoretical 'targetable' allele, assuming that the allele lost during LOH is random

$\Pi =$ theoretical number of new patients per year.

Estimate of new pan-cancer cases per year derived from SEER Cancer Statistics Review[111]. Rate of heterozygosity estimated using 2pq from Hardy-Weinberg equation[278,279].

DEMETER2 analyses

For a detailed description of the screening and analysis methodology used to generate DEMETER2 scores, please see [211]. Briefly, DEMETER2 generates an absolute dependency score for each gene suppressed in each cell line. A score of 0 signifies no dependency and a score of 1 signifies a strong dependency as estimated by scaling the effect to a panel of known pan-essential genes. DEMETER2 scores were obtained from the Cancer Depencency Map Portal (https://depmap.org/portal/download/) using the file 'D2_combined_gene_dep_scores.csv'. We classified genes that exhibited a median DEMETER2 score of ≤ -0.5 across all cell lines as moderately strong dependencies.

canSAR protein annotation

The canSAR protein annotation tool (cansar.icr.ac.uk) was run on a list of 741 unique genes containing 1749 insertion, deletion, and missense variants. Structures with >90% sequence homology were included in structural druggability and chemical matter analyses.

Determination of structures corresponding to variants

To determine which variants were present in PDB, DNA sequences (30mer) encapsulating 1749 insertion, deletion, and missense variants were translated in all 6 frames using the Bio.Seq Python module. Output was blasted using the Bio.Blast Python module against the PDB database with E-value thresholds of 0.001 or less, resulting in hits for 267 variants. We manually curated these structures to verify the presence of the variant within the PDB file and eliminated structures for which correspondence between the PDB protein sequence, ExAC amino acid prediction, and UCSC Genome Browser amino acid sequence was inconclusive. This curation yielded 153 protein-altering variants in proteins with homologous molecular structures.

Visual scoring was performed on 81 protein-altering variants that lie in X-ray crystal structures. Variants were scored using the following scale: 0 = no clear pockets on the protein surface, 1 = SNP far from pocket on protein surface, 2 = SNP near pocket on protein surface, 3 = SNP in pocket on protein surface, 4 = SNP near pocket containing small molecule.

This page intentionally left blank.

CHAPTER 5: DISCUSSION

This page intentionally left blank.

CHAPTER 5: DISCUSSION

Leveraging synthetic lethal interactions in cancer cells represents a promising avenue to targeting genomic differences between tumor and normal tissue. Synthetic lethality between genes occurs when singly inactivating one gene or the other maintains viability, but inactivating both genes simultaneously causes lethality[280]. Over the past 20 years, many efforts have been directed toward discovering synthetic lethal interactions with genetic driver alterations of oncogenes and tumor suppressor genes[91,281]. However, the number of genetically activated oncogenes and inactivated tumor suppressor genes in any given tumor is limited and, in many cancer types, is vastly outnumbered by genetically altered non-driver genes (e.g., due to passenger events). Therefore, identifying synthetic lethalities with genetic alterations affecting non-driver genes (also termed 'collateral lethalities'[14]) would increase the scope of potential therapeutic approaches. While individual GEMINI genes have been described previously[130], our work integrated genome-wide assessments of gene essentiality, genetic variation, and LOH to generate the first systematic analysis of this target class.

GEMINI vulnerabilities represent one of four classes of collateral lethalities. In addition to GEMINI vulnerabilities, deletion of paralogs can result in dependency on the remaining paralog; loss or gain of function of a non-driver pathway can lead to dependencies on alternative non-driver pathways[14]; and hemizygous loss of essential genes can result in dependency on the remaining copy (CYCLOPS)[43,45]. Prior analyses have indicated CYCLOPS genes to be the most frequent class of these synthetic lethal interactions[45,117], but we find that GEMINI vulnerabilities provide similar numbers of potential targets. In comparison, fewer paralog dependencies have been described[112–114,116,117]. Larger numbers of paralog vulnerabilities have been predicted[118], but it is unclear whether these predictions represent viable candidates[14]. The 1278 GEMINI genes

that we identified also exceed the 299 known driver genes[84], many of which are proposed therapeutic targets.

In comparison to CYCLOPS, targeting GEMINI vulnerabilities has two distinct advantages. First, whereas CYCLOPS genes must lie in regions of copy loss, GEMINI genes encompass genes that undergo both copy-loss and copy-neutral LOH. Second, while CYCLOPS vulnerabilities rely on relative differences between tumor and normal cells (differential expression of target genes), GEMINI vulnerabilities exploit absolute differences (the presence or absence of the allele that has undergone LOH). Thus, the possibility of allele-specific targeting presented by GEMINI genes may widen prospective therapeutic windows. In 269 cases, GEMINI vulnerabilities we detected reside in CYCLOPS genes[45] (see Supplementary Table 3). If GEMINI and CYCLOPS vulnerabilities are additive, targeting these genes might offer an even wider therapeutic window in cancers where CYCLOPS-GEMINI genes suffer LOH due to copy loss. However, individual GEMINI alterations may be less common among patients than individual CYCLOPs alterations due to the requirement that the germline genome be heterozygous at the GEMINI locus.

Strategies targeting LOH of non–cell-essential genes have also been explored. The gene N-acetyltransferase 2 (*NAT2*) encodes an enzyme (NAT2) involved in drug detoxification. *NAT2* exhibits several high-frequency polymorphisms that affect NAT2 kinetics, with some alleles generating ‘rapid’ or ‘slow’ N-acetylator phenotypes in the homozygous context, or ‘intermediate’ phenotypes in the heterozygous context[282]. Additionally, *NAT2* undergoes frequent LOH in colorectal cancers and other malignancies[283]. Ivaylo Stoimenov, Verónica Rendo, and colleagues have shown that the drug 6-(4-aminophenyl)-N-(3, 4, 5-trimethoxyphenyl)pyrazin-2-amine (also known as APA) selectively slows growth of cells with

slow NAT2 acetylator phenotypes, potentially by decreasing aurora kinase A (AURKA) activity and disrupting mitotic signaling (although APA has also been described as a BRAF inhibitor). They also found that patient response to APA depended on tumor *NAT2* acetylator status[283]. This work demonstrates the existence of a small-molecule inhibitor whose mechanism of action is predicated on an LOH event. Rendo and colleagues have also identified other missense and loss-of-function variants that may alter enzyme function and generate a therapeutic window upon loss of one allele.

In this thesis, we attempted to lay the foundation for future efforts to target GEMINI vulnerabilities. We first characterized the genomic landscape of LOH of essential genes and performed carefully controlled proof-of-concept validation of two GEMINI vulnerabilities. We also tested the possibility of targeting one of these vulnerabilities, *PRIM1*^{rs2277339}, using allele-specific RNAi and determined that not all GEMINI vulnerabilities may be ideal candidates for such an approach due to expression threshold effects or variant sequence context. Finally, we analyzed protein-altering GEMINI variants for their potential as allele-specific drug targets. These studies have improved our understanding of the difficulties of selecting GEMINI targets for future development and of designing potential allele-specific therapeutics. The development of a tractable therapeutic that targets any single GEMINI gene in an allele-specific manner is a substantial challenge. However, the sheer number of potential candidates suggests that some of these GEMINI vulnerabilities may represent viable targets.

References

1. Nichols, C. A. *et al.* Loss of heterozygosity of essential genes represents a widespread class of potential cancer vulnerabilities. *bioRxiv* **29 January,** 1–52 (2019).

2. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70 (2000).

3. Hajdu, S. I. & Vadmal, M. A Note From History : Landmarks in History of Cancer , Part 6. (2013). doi:10.1002/cncr.28319

4. Papac, R. J. Origins of Cancer Therapy. **74,** 391–398 (2002).

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

6. Shih, C. & Weinberg, R. A. Isolation of a Transforming Sequence from a Human Bladder Carcinoma Cell Line. **29,** (1982).

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

8. Friend, S. H. *et al.* A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323,** 643–646 (1986).

9. Wheeler, D. A. & Wang, L. From human genome to cancer genome : The first decade. 1054–1062 (2013). doi:10.1101/gr.157602.113.1054

10. Merlo, L. M. F., Pepper, J. W., Reid, B. J. & Maley, C. C. Cancer as an evolutionary and ecological process. *Nat. Rev. Cancer* **6,** 924–935 (2006).

11. Marusyk, A. & Kornelia, P. Tumor heterogeneity: causes and consequences Andriy. *Biochim. Biophys. Acta* **1805,** 1–28 (2011).

12. Lipinski, K. A. *et al.* Cancer Evolution and the Limits of Predictability in Precision Cancer Medicine. *Trends in Cancer* **2,** 49–63 (2016).

13. Zack, T. I. Exploring Cancer's Fractured Genomic Landscape: Searching for Cancer Drivers and Vulnerabilities in Somatic Copy Number Alterations. (2014).

14. Muller, F. L., Aquilanti, E. A. & DePinho, R. A. Collateral Lethality: A New Therapeutic Strategy in Oncology. *Trends in Cancer* **1,** 161–173 (2015).

15. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168,** 613–628 (2017).

16. Johnson, B. E. *et al.* Mutational Analysis Reveals the Origin and Therapy-Driven Evolution of Recurrent Glioma. *Science (80-.).* **343,** 189–194 (2014).

17. Piotrowska, Z. *et al.* Heterogeneity underlies the emergence of EGFRT790 wild-type clones following treatment of T790M-positive cancers with a third-generation EGFR inhibitor. *Cancer Discov.* **5,** 713–723 (2015).

18. Bettegowda, C. *et al.* Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Sci. Transl. Med.* **6,** 224ra24-224ra24 (2014).

19. Zhao, Y. & Adjei, A. A. Targeting Oncogenic Drivers. **41,** 1–14 (2014).

20. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nat. Med.* **10,** 789–799 (2004).

21. Bos, J. L. *et al.* Prevelence of ras gene mutations in human colorectal cancer. *Nature* **329,** 855–857 (1987).

22. Forrester, K., Allmoguera, C., Perucho, M., Han, K. & Grizzle, W. E. Detection of high incidence of K-ras oncogenes during human colon tumorigenesis. *Nature* **327,** 298–303 (1987).

23. Lynch, T. J. *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).

24. Paez, J. G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science (80-.).* **304,** 1497–500 (2004).

25. Olivier, M., Hollstein, M. & Hainaut, P. TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. *Cold Spring Harb. Perspect. Biol.* **2,** 1–17 (2010).

26. Giannakis, M. *et al.* RNF43 is frequently mutated in colorectal and endometrial cancers. *Nat. Genet.* **46,** 1264–1266 (2014).

27. Maruvka, Y. E. *et al.* Analysis of somatic microsatellite indels identifies driver events in human tumors. *Nat. Biotechnol.* **35,** 951–959 (2017).

28. Dego, L. The History of Acute Promyelocytic Leukaemia. *Br. J. Haematol.* **122,** 539–553 (2003).

29. Ren, R. Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. *Nat. Rev. Cancer* **5,** 172–183 (2005).

30. Adjei, A. A. *et al.* ALK inhibitors in the treatment of advanced NSCLC. *Cancer Treat. Rev.* **40,** 300–306 (2013).

31. Bandopadhayay, P. *et al.* MYB-QKI rearrangements in angiocentric glioma drive tumorigenicity through a tripartite mechanism. *Nat. Genet.* **48,** 273–282 (2016).

32. Epstein, M. A., Achong, B. G. & Barr, Y. M. Virus Particles in Cultured Lymphoblasts

from Burkitt's Lymphoma. *Lancet* **1,** 702–703 (1964).

33. Durst, M., Gissmann, L., Ikenberg, H. & Hausen, H. Zur. A papillomavirus, DNA-from a cervical carcinoma and-its-prevalence in cancer biopsy samples from different geographic regions (human papillomaviruses/low-stringency hybridization/molecular cloning/genital tumors). *Proc. NatL Acad. Sci. USA* **80,** 3812–3815 (1983).

34. Cavenee, W. K. *et al.* Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* **305,** 779–784 (1983).

35. Sansal, I. & Sellers, W. R. The Biology and Clinical Relevance of the PTEN Tumor Suppressor Pathway. **22,** 2954–2963 (2019).

36. Cao, Y., Bryan, T. M. & Reddel, R. R. Increased copy number of the TERT and TERC telomerase subunit genes in cancer cells. *Cancer Sci.* **99,** 1092–1099 (2008).

37. Brison, O. Gene amplification and tumor progression. *BBA - Rev. Cancer* **1155,** 25–41 (1993).

38. Mitelman, F. Recurrent chromosome aberrations in cancer. *Mutat. Res. - Rev. Mutat. Res.* **462,** 247–253 (2000).

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

40. Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45,** 1134–1140 (2013).

41. Baudis, M. Genomic imbalances in 5918 malignant epithelial tumors: An explorative meta-analysis of chromosomal CGH data. *BMC Cancer* **7,** 1–15 (2007).

42. Cowey, C. L. & Rathmell, W. K. VHL gene mutations in renal cell carcinoma: Role as a biomarker of disease outcome and drug efficacy. *Curr. Oncol. Rep.* **11,** 94–101 (2009).

43. Nijhawan, D. *et al.* Cancer Vulnerabilities Unveiled by Genomic Loss. *Cell* **150,** 842–854 (2012).

44. Taylor, A. M. *et al.* Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* **33,** 676–689 (2018).

45. Paolella, B. R. *et al.* Copy-number and gene dependency analysis reveals partial copy loss of wild-type SF3B1 as a novel cancer vulnerability. *Elife* **6,** 1–35 (2017).

46. Gönczy, P. Centrosomes and cancer: revisiting a long-standing relationship. *Nat. Rev. Cancer* **15,** 639–652 (2015).

47. Bakhoum, S. F., Thompson, S. L., Manning, A. L. & Compton, D. A. Genome stability is ensured by temporal control of kinetochore – microtubule dynamics. **11,** (2009).

48. Hills, S. A. & Diffley, J. F. X. DNA Replication and Oncogene-Induced Replicative Stress. *CURBIO* **24,** R435–R444 (2014).

49. Yi, K. & Ju, Y. S. Patterns and mechanisms of structural variations in human cancer. *Exp. Mol. Med.* (2018). doi:10.1038/s12276-018-0112-3

50. Maciejowski, J. *et al.* Chromothripsis and Kataegis Induced by Telomere Article Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell* **163,** 1641–1654 (2015).

51. Maciejowski, J. & Lange, T. De. Telomeres in cancer: tumour suppression and genome instability. *Nat Rev Mol Cell Biol* **18,** 175–186 (2017).

52. Crasta, K. *et al.* DNA breaks and chromosome pulverization from errors in mitosis. *Nature* **482,** 53–58 (2012).

53. Zhang, C. *et al.* CHROMOTHRIPSIS FROM DNA DAMAGE IN MICRONUCLEI Cheng-Zhong. *Nature* **522,** 179–184 (2015).

54. Holland, A. J. & Cleveland, D. W. Boveri revisited: Chromosomal instability, aneuploidy and tumorigenesis. *Nat. Rev. Mol. Cell Biol.* **10,** 478–487 (2009).

55. Slamon, D. *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).

56. Slamon, D. J. *et al.* Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science (80-.).* **235,** 177 LP-182 (1987).

57. Maximiano, S., Magalhães, P., Guerreiro, M. P. & Morgado, M. Trastuzumab in the Treatment of Breast Cancer. *BioDrugs* **30,** 75–86 (2016).

58. Wykosky, J., Fenton, T., Furnari, F. & Cavenee, W. K. Therapeutic targeting of epidermal growth factor receptor in human cancer: successes and limitations. *Chin. J. Cancer* **30,** 5–12 (2011).

59. Tang, Y. C., Williams, B. R., Siegel, J. J. & Amon, A. Identification of aneuploidy-selective antiproliferation compounds. *Cell* **144,** 499–512 (2011).

60. Manchado, E. & Malumbres, M. Targeting aneuploidy for cancer therapy. *Cell* **144,** 465–466 (2011).

61. Knudson, A. G. Genetics of human cancer. *J. Cell. Physiol.* **129,** 7–11 (1986).

62. Little, M. & Wells, C. A Clinical Overview of WT 1 Gene Mutations. **225,** (1997).

63. Cawthon, R. M. *et al.* A Major Segment of the Neurofibromatosis Type 1 Gene : cDNA Sequence, Genomic Structure, and Point Mutations. **62,** 193–201 (1990).

64. Su, L.-K. *et al.* Multiple intestinal neoplasia caused by a mutation in the murine homolog

of the APC gene. (1992). doi:10.1126/science.1350108

65. Weith, A. *et al.* Neuroblastoma consensus deletion maps to 1p36.1–2. *Genes, Chromosom. Cancer* **1,** 159–166 (1989).

66. Steck, P. A. *et al.* Functional and molecular analyses of 10q deletions in human gliomas. *Genes, Chromosom. Cancer* **24,** 135–143 (1999).

67. Reifenberger, J. *et al.* Molecular Genetic Analysis of Oligodendroglial Tumors Shows Preferential Allelic Deletions on 19q and 1p. **145,** 1175–1190 (1994).

68. Beroukhim, R. *et al.* Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. *PLoS Comput. Biol.* **2,** 323–332 (2006).

69. Thiagalingam, S. *et al.* Mechanisms underlying losses of heterozygosity in human colorectal cancers. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 2698–702 (2001).

70. Haber, J. E. DNA recombination : the replication connection. **0004,** 271–275 (1999).

71. Meuth, M. The structure of mutation in mammalian cells. **1032,** 1–17 (1990).

72. Stanbridge, E. Human tumor suppressor genes. (1990).

73. Cavenee, W. K. *et al.* Genetic origin of mutations predisposing to retinoblastoma. *Science (80-.).* **228,** 501 LP-503 (1985).

74. Jordan, V. C. Tamoxifen: A Most Unlikely Pioneering Medicine. *Nat. Rev. Drug Discov.* **2,** 205–213 (2003).

75. Yan, L., Rosen, N. & Arteaga, C. Editorial Challenges in Targeted Cancer Drug Development. *Chin. J. Cancer* **30,** 1–4 (2011).

76. Gross, J. M. & Yee, D. How does the estrogen receptor work? *Breast Cancer Res.* **4,** 62–64 (2002).

77. Warrell, R. P., de The, H., Wang, Z.-Y. & Degos, L. Acute Promyelocytic Leukemia. (1993).

78. Mukherjee, S. *The Emperor of All Maladies: A Biography of Cancer*. (Scribner, 2010).

79. Kumar, G. L. & Badve, S. S. Milestones in the Discovery of HER2 Proto-Oncogene and Trastuzumab (Herceptin™). *Connection* 9–14 (2008).

80. Afghahi, A. & Sledge, G. W. Targeted Therapy for Cancer in the Genomic Era. **21,** 294–298 (2015).

81. Pray, L. A. Gleevec: The Breakthrough in Cancer Treatment. *Nat. Educ.* **1,** 37 (2008).

82. Tsimberidou, A. M. Targeted therapy in cancer. *Cancer Chemother. Pharmacol.* **76,**

1113–1132 (2015).

83. Lazo, J. S. & Sharlow, E. R. Drugging Undruggable Molecular Cancer Targets. *Annu. Rev. Pharmacol. Toxicol.* **56,** 23–40 (2016).

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

85. Boehm, J. S., Hession, M. T., Bulmer, S. E. & Hahn, W. C. Transformation of Human and Murine Fibroblasts without Viral Oncoproteins. *Mol. Cell. Biol.* **25,** 6464–6474 (2005).

86. Tomasetti, C., Marchionni, L., Nowak, M. A., Parmigiani, G. & Vogelstein, B. Only three driver gene mutations are required for the development of lung and colorectal cancers. **112,** (2015).

87. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171,** 1029–1041 (2017).

88. de la Cruz, F. F., Gapp, B. V & Nijman, S. M. B. Synthetic Lethal Vulnerabilities of Cancer. *Annu. Rev. Pharmacol. Toxicol.* **55,** 513–531 (2015).

89. Chan, D. A. & Giaccia, A. J. Harnessing synthetic lethal interactions in anticancer drug discovery. (2011). doi:10.1038/nrd3374

90. Zhan, T. & Boutros, M. Towards a compendium of essential genes – From model organisms to synthetic lethality in cancer cells. *Crit. Rev. Biochem. Mol. Biol.* **51,** 74–85 (2015).

91. McLornan, D. P., List, A. & Mufti, G. J. Applying Synthetic Lethality for the Selective Targeting of Cancer. *N. Engl. J. Med.* **371,** 1725–1735 (2014).

92. Decker, W. K. *et al.* Cancer Immunotherapy: Historical Perspective of a Clinical Revolution and Emerging Preclinical Animal Models. *Front. Immunol.* **8,** (2017).

93. Topalian, S. L., Drake, C. G. & Pardoll, D. M. Immune checkpoint blockade: A common denominator approach to cancer therapy. *Cancer Cell* **27,** 450–461 (2015).

94. Zhang, H. & Chen, J. Current status and future directions of cancer immunotherapy. *J. Cancer* **9,** 1773–1781 (2018).

95. Voena, C. & Chiarle, R. Advances in Cancer Immunology and Cancer Immunotherapy. *Discov. Med.* **21,** 125–133 (2016).

96. Buchbinder, E. I. & Desai, A. CTLA-4 and PD-1 Pathways. *Am. J. Clin. Oncol.* **39,** 98–106 (2016).

97. Lu, Y. C. & Robbins, P. F. Targeting neoantigens for cancer immunotherapy. *Int. Immunol.* **28,** 365–370 (2016).

98. Borcherding, N. *et al.* Keeping Tumors in Check: A Mechanistic Review of Clinical Response and Resistance to Immune Checkpoint Blockade in Cancer. *J. Mol. Biol.* **430,** 2014–2029 (2018).

99. Le, D. T. *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science (80-.).* **357,** 409–413 (2017).

100. Rizvi, N. A. *et al.* Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science (80-.).* **348,** 124–128 (2015).

101. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science (80-.).* **351,** 1463–1470 (2016).

102. Allen, E. M. Van *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science (80-.).* **350,** 207–211 (2015).

103. Snyder, A. *et al.* Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *N. Engl. J. Med.* **371,** 2189–2199 (2014).

104. T.N., S. & R.D., S. Neoantigens in cancer immunotherapy. *Science (80-.).* **348,** 69–74 (2015).

105. Heemskerk, B., Kvistborg, P. & Schumacher, T. N. M. The cancer antigenome. *EMBO J.* **32,** 194–203 (2013).

106. Wu, J. *et al.* TSNAdb: A Database for Tumor-specific Neoantigens from Immunogenomics Data Analysis. *Genomics, Proteomics Bioinforma.* **16,** 276–282 (2018).

107. Lee, C. H., Yelensky, R., Jooss, K. & Chan, T. A. Update on Tumor Neoantigens and Their Utility: Why It Is Good to Be Different. *Trends Immunol.* **39,** 536–548 (2018).

108. Editorial. The problem with neoantigen prediction. *Nat. Biotechnol.* **35,** 97–97 (2017).

109. Hu, Z., Ott, P. A. & Wu, C. J. Towards personalized, tumour-specific, therapeutic vaccines for cancer. *Nat. Rev. Immunol.* **18,** 168–182 (2018).

110. Keskin, D. B. *et al.* Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature* **565,** 234–239 (2019).

111. Noone, A. *et al.* SEER Cancer Statistics Review, 1975-2015. *National Cancer Institute* (2017). Available at: https://seer.cancer.gov/csr/1975_2015/,.

112. Muller, F. L. *et al.* Passenger Deletions Generate Therapeutic Vulnerabilities in Cancer. *Nature* **488,** 337–342 (2012).

113. Viswanathan, S. R. *et al.* Genome-scale analysis identifies paralog lethality as a vulnerability of chromosome 1p loss in cancer. *Nat. Genet.* **50,** 937–943 (2018).

114. Helming, K. C. *et al.* ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nat.*

Med. **20,** 251–254 (2014).

115. Hoffman, G. R. *et al.* Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proc. Natl. Acad. Sci.* **111,** 3128–3133 (2014).

116. Dey, P. *et al.* Genomic deletion of malic enzyme 2 confers collateral lethality in pancreatic cancer. *Nature* **542,** 119–123 (2017).

117. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170,** 564–576 (2017).

118. Aksoy, B. A. *et al.* Prediction of individualized therapeutic vulnerabilities in cancer from genomic profiles. *Bioinformatics* **30,** 2051–2059 (2014).

119. Tang, B., Testa, J. R. & Kruger, W. D. Increasing the therapeutic index of 5-fluorouracil and 6-thioguanine by targeting loss of MTAP in tumor cells. *Cancer Biol. Ther.* **13,** 1082–1090 (2012).

120. Bertino, J. R., Waud, W. R., Parker, W. B. & Lubin, M. Targeting tumors that lack methylthioadenosine phosphorylase (MTAP) activity: Current strategies. *Cancer Biol. Ther.* **11,** 627–632 (2011).

121. Li, A. *et al. Status of Methylthioadenosine Phosphorylase and its Impact on Cellular Response to l-Alanosine and Methylmercaptopurine Riboside in Human Soft Tissue Sarcoma Cells*. *Oncology research* **14,** (2004).

122. Ruefli-Brasse, A. *et al.* Methylthioadenosine (MTA) Rescues Methylthioadenosine Phosphorylase (MTAP)-Deficient Tumors from Purine Synthesis Inhibition in Vivo via Non-Autonomous Adenine Supply. *J. Cancer Ther.* **02,** 523–534 (2011).

123. Marjon, K. *et al.* MTAP Deletions in Cancer Create Vulnerability to Targeting of the MAT2A/PRMT5/RIOK1 Axis. *Cell Rep.* **15,** 574–587 (2016).

124. Folger, O. *et al.* Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* **7,** 1–10 (2011).

125. Molina, J. R. *et al.* An inhibitor of oxidative phosphorylation exploits cancer vulnerability. *Nat. Med.* **24,** 1036–1046 (2018).

126. Sun, Y. *et al.* Functional Genomics Reveals Synthetic Lethality between Phosphogluconate Dehydrogenase and Article Functional Genomics Reveals Synthetic Lethality between Phosphogluconate Dehydrogenase and Oxidative Phosphorylation. *Cell Rep.* **26,** 469–482.e5 (2019).

127. Liu, Y. *et al.* TP53 loss creates therapeutic vulnerability in colorectal cancer. (2015). doi:10.1038/nature14418

128. Zeldis, J. B. *et al.* Efficacy of Lenalidomide in Myelodysplastic Syndromes. *N. Engl. J.*

Med. **352,** 549–557 (2005).

129. Schmidt, M. *et al.* Lenalidomide in the Myelodysplastic Syndrome with Chromosome 5q Deletion. *N. Engl. J. Med.* **355,** 1456–1465 (2006).

130. Fluiter, K., Housman, D., Ten Asbroek, a L. M. a & Baas, F. Killing cancer by targeting genes that cancer cells have lost: Allele-specific inhibition, a novel approach to the treatment of genetic disorders. *Cell. Mol. Life Sci.* **60,** 834–843 (2003).

131. Basilion, J. P. *et al.* Selective Killing of Cancer Cells Based on Loss of Heterozygosity and Normal Variation in the Human Genome: A New Paradigm for Anticancer Drug Therapy. *Mol. Pharmacol.* **56,** 359–369 (1999).

132. ten Asbroek, A. L. M. A., Fluiter, K., van Groenigen, M., Nooij, M. & Baas, F. Polymorphisms in the large subunit of human RNA polymerase II as target for allele-specific inhibition. *Nucleic Acids Res.* **28,** 1133–1138 (2000).

133. Fluiter, K. *et al.* Tumor Genotype-specific Growth Inhibition in Vivo by Antisense Oligonucleotides against a Polymorphic Site of the Large Subunit of Human RNA Polymerase II. *Cancer Res.* **62,** 2024–2028 (2002).

134. Mook, O. R. F., Baas, F., de Wissel, M. B. & Fluiter, K. Allele-specific cancer cell killing in vitro and in vivo targeting a single-nucleotide polymorphism in POLR2A. *Cancer Gene Ther.* **16,** 532–538 (2009).

135. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).

136. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. 1–10 (2015). doi:10.1126/science.aac7041

137. Raychaudhuri, S. Primer Mapping Rare and Common Causal Alleles for Complex Human Diseases. *Cell* **147,** 57–69 (2011).

138. Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of- function intolerance across human protein-coding genes. *bioRxiv* (2019).

139. Frazer, K. A., Murray, S. S., Schork, N. J. & Topol, E. J. Human genetic variation and its contribution to complex traits. **10,** 18–20 (2009).

140. Smigielski, E. M., Sirotkin, K., Ward, M. & Sherry, S. T. dbSNP : a database of single nucleotide polymorphisms. **28,** 352–355 (2000).

141. Group, T. I. S. M. W. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409,** 928–933 (2001).

142. Levy, S. *et al.* The Diploid Genome Sequence of an Individual Human. **5,** (2007).

143. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. **452,** 872–877 (2008).

144. Wang, J. *et al.* The diploid genome sequence of an Asian individual. **456,** 60–66 (2008).

145. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–59 (2008).

146. Consortium, T. 1000 G. P. A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).

147. Consortium, T. 1000 G. P. An integrated map of genetic variation. *Nature* **135,** 56–65 (2012).

148. Consortium, G. P. A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

149. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493,** 216–220 (2013).

150. Rancati, G., Moffat, J., Typas, A. & Pavelka, N. Emerging and evolving concepts in gene essentiality. *Nat. Rev. Genet.* **19,** 34–49 (2018).

151. Bartha, I., Di Iulio, J., Venter, J. C. & Telenti, A. Human gene essentiality. *Nat. Rev. Genet.* **19,** 51–62 (2018).

152. Evers, B. *et al.* CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat. Biotechnol.* **34,** 631–633 (2016).

153. Morgens, D. W., Deans, R. M., Li, A. & Bassik, M. C. Systematic comparison of CRISPR-Cas9 and RNAi screens for essential genes. *Nat. Biotechnol.* **34,** 634–636 (2016).

154. JM, B., JL, T. & L, S. Section 27.2, DNA Polymerases Require a Template and a Primer. in *Biochemistry* (W. H. Freeman, 2002).

155. Frick, D. N. & Richardson, C. C. DNA Primases. *Annu. Rev. Biochem.* **70,** 39–80 (2001).

156. Kuchta, R. D. & Stengel, G. Mechanism and evolution of DNA primases. *Biochim. Biophys. Acta - Proteins Proteomics* **1804,** 1180–1189 (2010).

157. Arezi, B. & Kuchta, R. D. Eukaryotic DNA primase. *Trends Biochem. Sci.* **25,** 572–576 (2000).

158. Francesconi, S. *et al.* Mutations in conserved yeast DNA primase domains impair DNA replication in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 3877–3881 (1991).

159. Longhese, M. P., Jovine, L., Plevani, P. & Lucchini, G. Conditional mutations in the yeast DNA primase genes affect different aspects of DNA metabolism and interactions in the

DNA polymerase ??-primase complex. *Genetics* **133,** 183–191 (1993).

160. Marini, F. *et al.* A role for DNA primase in coupling DNA replication to DNA damage response. *EMBO J.* **16,** 639–650 (1997).

161. Michael, W. M., Ott, R., Fanning, E. & Newport, J. Activation of the DNA replication checkpoint through RNA synthesis by primase. *Science* **289,** 2133–2137 (2000).

162. Macdougall, C. A., Byun, T. S., Van, C., Yee, M. & Cimprich, K. A. The structural determinants of checkpoint activation. 898–903 (2007). doi:10.1101/gad.1522607.most

163. Yamaguchi, M. *et al.* Mutation of DNA primase causes extensive apoptosis of retinal neurons through the activation of DNA damage checkpoint and tumor suppressor p53. *Development* **135,** 1247–57 (2008).

164. Diede, S. J. & Gottschling, D. E. Telomerase-Mediated Telomere Addition In Vivo Requires DNA Primase and DNA Polymerases ␣ and ␦. **99,** 723–733 (1999).

165. Lucchini, G., Francesconi, S., Foiani, M., Badaracco, G. & Plevani, P. Yeast DNA polymerase–DNA primase complex: cloning of PRI 1, a single essential gene related to DNA primase activity. *EMBO J.* **6,** 737–742 (1987).

166. Foiani, M., Santocanale, C., Plevani, P. & Lucchini, G. A Single Essential Gene , PRI2 , Encodes the Large Subunit of DNA Primase in Saccharomyces cerevisiae. **9,** 3081–3087 (1989).

167. Kilchert, C., Wittmann, S. & Vasiljeva, L. The regulation and functions of the nuclear RNA exosome complex. *Nat. Rev. Mol. Cell Biol.* **17,** 227–239 (2016).

168. Zinder, J. C. & Lima, C. D. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev.* **31,** 88–100 (2017).

169. Chlebowski, A., Lubas, M., Jensen, T. H. & Dziembowski, A. RNA decay machines: The exosome. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829,** 552–560 (2013).

170. Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollervey, D. The Exosome: A Conserved Eukaryotic RNA Processing Complex Containing Multiple 3'-->5' Exoribonucleases. *Cell* **91,** 457–466 (1997).

171. Allmang, C. *et al.* The yeast exosome and human PM-Scl are related complexes of 3' → 5' exonucleases. *Genes Dev.* **13,** 2148–2158 (1999).

172. Wan, J. *et al.* Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration. *Nat. Genet.* **44,** 704–708 (2012).

173. Boczonadi, V. *et al.* EXOSC8 mutations alter mRNA metabolism and cause hypomyelination with spinal muscular atrophy and cerebellar hypoplasia. *Nat. Commun.*

5, 4287 (2014).

174. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science (80-.).* **339,** 1546–1558 (2013).

175. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37,** 1–13 (2009).

176. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 (2009).

177. Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49,** 1779–1784 (2017).

178. Boehm, J. S. & Golub, T. R. An ecosystem of cancer cell line factories to support a cancer dependency map. *Nat. Rev. Genet.* **16,** 373–374 (2015).

179. Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30,** 413–421 (2012).

180. Yilmaz, A., Peretz, M., Aharony, A., Sagi, I. & Benvenisty, N. Defining essential genes for human pluripotent stem cells by CRISPR–Cas9 screening in haploid cells. *Nat. Cell Biol.* **20,** (2018).

181. Sondka, Z. *et al.* The COSMIC Cancer Gene Census : describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **18,** (2018).

182. Courtney, D. G. *et al.* CRISPR/Cas9 DNA cleavage at SNP-derived PAM enables both in vitro and in vivo KRT12 mutation-specific targeting. *Gene Ther.* **23,** 108–112 (2016).

183. Shin, J. W. *et al.* Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. *Hum. Mol. Genet.* **25,** 4566–4576 (2016).

184. Christie, K. A. *et al.* Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders. *Sci. Rep.* **7,** 1–11 (2017).

185. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155,** 733–740 (2009).

186. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31,** 827–832 (2013).

187. Schmid, M. & Jensen, T. H. The exosome: a multipurpose RNA-decay machine. *Trends Biochem. Sci.* **33,** 501–510 (2008).

188. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* (2015). doi:10.1038/nature14592

189. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556,** 57–63 (2018).

190. Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science (80-.).* **9129,** eaas9129 (2018).

191. Hou, Z. *et al.* Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *PNAS* **110,** 15644–15649 (2013).

192. Kim, E. *et al.* In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat. Commun.* **8,** 1–12 (2017).

193. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163,** 759–771 (2015).

194. Smith, C. *et al.* Efficient and Allele-Specific Genome Editing of Disease Loci in Human iPSCs. *Mol. Ther.* 1–8 (2014). doi:10.1038/mt.2014.226

195. Li, P. *et al.* Allele-specific CRISPR/Cas9 genome editing of the single-base P23H mutation for rhodopsin associated dominant retinitis pigmentosa. *Cris. J.* **1,** 55–64 (2018).

196. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343,** 80–4 (2014).

197. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34,** 184–191 (2016).

198. van Overbeek, M. *et al.* DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol. Cell* **63,** 633–646 (2016).

199. Chakrabarti, A. M., Brownhill, T. H., Monserrat, J. & Poetsch, A. R. Target - specific precision of CRISPR - mediated genome editing. *bioRxiv* (2018).

200. Bondeson, D. P. & Crews, C. M. Targeted Protein Degradation by Small Molecules. *Annu Rev Pharmacol Toxicol* **57,** 107–123 (2017).

201. Bollen, Y., Post, J., Koo, B.-K. & Snippert, H. J. G. How to create state-of-the-art genetic model systems: strategies for optimal CRISPR-mediated genome editing. *Nucleic Acids Res.* **46,** 6435–6454 (2018).

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

203. Gaudelli, N. m. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551,** 464–471 (2017).

204. Cai, Y. *et al.* Loss of Chromosome 8p Governs Tumor Progression and Drug Response by

Altering Lipid Metabolism. *Cancer Cell* **29,** 751–766 (2016).

205. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–307 (2012).

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

207. Moffat, J. *et al.* A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen. *Cell* **124,** 1283–1298 (2006).

208. Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout. *Science* **343,** 84–88 (2014).

209. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34,** 339–344 (2016).

210. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. a. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31,** 233–9 (2013).

211. McFarland, J. M. *et al.* Improved estimation of cancer dependencies from large-scale RNAi screens using model- based normalization and data integration. *bioRxiv* 305656 (2018). doi:10.1101/305656

212. Halling-Brown, M. D., Bulusu, K. C., Patel, M., Tym, J. E. & Al-Lazikani, B. canSAR: an integrated cancer public translational research and drug discovery resource. *Nucleic Acids Res.* **40,** 947–956 (2012).

213. Tym, J. E. *et al.* canSAR: An updated cancer research and drug discovery knowledgebase. *Nucleic Acids Res.* **44,** D938–D943 (2016).

214. *Enzyme Nomenclature 1992*. (Academic Press, 1992).

215. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).

216. Saravanana, M. *et al.* Interactions between the nucleosome histone core and Arp8 in the INO80 chromatin remodeling complex. *Proc. Natl. Acad. Sci.* **109,** 20883–20888 (2012).

217. Gerhold, C. B. *et al.* Structure of Actin-related protein 8 and its contribution to nucleosome binding. *Nucleic Acids Res.* **40,** 11036–11046 (2012).

218. Osakabe, A. *et al.* DNA Binding Properties of the Actin-Related Protein Arp8 and Its Role in DNA Repair. *PLoS One* **9,** e108354 (2014).

219. Schmitt, C. *et al.* Dbp5, a DEAD-box protein required for mRNA export, is recruited to the cytoplasmic fibrils of nuclear pore complex via a conserved interaction with CAN/Nup159p. *EMBO J.* **18,** 4332–4347 (1999).

220. Zhao, J., Jin, S. B., Björkroth, B., Wieslander, L. & Daneholt, B. The mRNA export factor

Dbp5 is associated with Balbiani ring mRNP from gene to cytoplasm. *EMBO J.* **21,** 1177–1187 (2002).

221. Bonnet, A. & Palancade, B. Regulation of mRNA trafficking by nuclear pore complexes. *Genes (Basel).* **5,** 767–791 (2014).

222. Wang, Y. *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512,** 155–160 (2014).

223. Liu, G. *et al.* Gene Essentiality Is a Quantitative Property Linked to Cellular Evolvability. *Cell* **163,** 1388–1399 (2015).

224. Kim, T. *et al.* Subclonal Genomic Architectures of Primary and Metastatic Colorectal Cancer Based on Intratumoral Genetic Heterogeneity. *Clin. Cancer Res.* **21,** 4461–4473 (2015).

225. Gibson, W. J. *et al.* The genomic landscape and evolution of endometrial carcinoma progression and abdominopelvic metastasis. *Nat. Genet.* **48,** 848–855 (2016).

226. Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **376,** 2109–2121 (2017).

227. McGranahan, N. *et al.* Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* **171,** 1259–1271.e11 (2017).

228. Peters, J. The role of genomic imprinting in biology and disease: An expanding view. *Nat. Rev. Genet.* **15,** 517–530 (2014).

229. Wrzeska, M. & Rejduch, B. Genomic imprinting in mammals. *J Appl Genet* **45,** 427–433 (2004).

230. Ishida, M. & Moore, G. E. The role of imprinted genes in humans. *Mol. Aspects Med.* **34,** 826–840 (2013).

231. Savova, V., Vigneau, S. & Gimelbrant, A. A. Autosomal monoallelic expression: Genetics of epigenetic diversity? *Curr. Opin. Genet. Dev.* **23,** 642–648 (2013).

232. Gimelbrant, A., Hutchinson, J. N., Thompson, B. R. & Chess, A. Widespread Monoallelic Expression on Human Autosomes. *Science (80-.).* **318,** 1136–1140 (2007).

233. Vigneau, S., Vinogradova, S., Savova, V. & Gimelbrant, A. High prevalence of clonal monoallelic expression. *Nat. Genet.* **50,** 1198–1199 (2018).

234. Savova, V., Patsenker, J. & Gimelbrant, A. A. dbMAE : the database of autosomal monoallelic. *Nucleic Acids Res.* **44,** 753–756 (2016).

235. Wang, Y. *et al.* Systemic Delivery of Modified mRNA Encoding Herpes Simplex Virus 1 Thymidine Kinase for Targeted Cancer Gene Therapy. *Mol. Ther.* **21,** 358–367 (2013).

236. Wang, M., Alberti, K., Sun, S., Arellano, C. L. & Xu, Q. Combinatorially Designed Lipid-like Nanoparticles for Intracellular Delivery of Cytotoxic Protein for Cancer Therapy. *Angew. Chemie - Int. Ed.* **53,** 2893–2898 (2014).

237. Sun, W. *et al.* Self-Assembled DNA Nanoclews for the Efficient Delivery of CRISPR-Cas9 for Genome Editing. *Angew. Chemie - Int. Ed.* **54,** 12029–12033 (2015).

238. Wang, H.-X. *et al.* Nonviral gene editing via CRISPR/Cas9 delivery by membrane-disruptive and endosomolytic helical polypeptide. *Proc. Natl. Acad. Sci.* **115,** 4903–4908 (2018).

239. Charlesworth, C. T. *et al.* Identification of Pre-Existing Adaptive Immunity to Cas9 Proteins in Humans. *bioRxiv* (2018).

240. Moreno, A. M. *et al.* Exploring protein orthogonality in immune space: a case study with AAV and Cas9 orthologs. *bioRxiv* 245985 (2018). doi:10.1101/245985

241. Ferdosi, S. R. *et al.* Multifunctional CRISPR/Cas9 with engineered immunosilenced human T cell epitopes. *bioRxiv* 360198 (2018). doi:10.1101/360198

242. Fluiter, K. *et al.* In vivo tumor growth inhibition and biodistribution studies of locked nucleic acid (LNA) antisense oligonucleotides. *Nuc* **31,** 953–962 (2003).

243. Moreno, P. M. D. & Pego, A. P. Therapeutic antisense oligonucleotides against cancer: hurdling to the clinic. *Front. Chem.* **2,** 1–7 (2014).

244. Smith, C. I. E. & Zain, R. Therapeutic Oligonucleotides : State of the Art. *Annu. Rev. Pharmacol. Toxicol.* **59,** 1–26 (2018).

245. Stein, C. A. & Castanotto, D. FDA-Approved Oligonucleotide Therapies in 2017. *Mol. Ther.* **25,** 1069–1075 (2017).

246. Hoy, S. M. Patisiran: First Global Approval. *Drugs* (2018). doi:10.1007/s40265-018-0983-6

247. ClinicalTrials.gov. Available at: clinicaltrials.gov. (Accessed: 9th October 2018)

248. Zuckerman, J. E. & Davis, M. E. Clinical experiences with systemically administered siRNA-based therapeutics in cancer. *Nat. Rev. Drug Discov.* **14,** 843–856 (2015).

249. Kaczmarek, J. C., Kowalski, P. S. & Anderson, D. G. Advances in the delivery of RNA therapeutics: From concept to clinical reality. *Genome Med.* **9,** 1–16 (2017).

250. Ngamcherdtrakul, W. *et al.* Current development of targeted oligonucleotide-based cancer therapies : Perspective on HER2-positive breast cancer treatment. 19–29 (2017). doi:10.1016/j.ctrv.2016.02.005.Current

251. Porciani, D. *et al.* Modular cell-internalizing aptamer nanostructure enables targeted

delivery of large functional RNAs in cancer cell lines. *Nat. Commun.* **9,** (2018).

252. Sun, C. Y. *et al.* Tumor Acidity-Sensitive Polymeric Vector for Active Targeted siRNA Delivery. *J. Am. Chem. Soc.* **137,** 15217–15224 (2015).

253. Hanvey, J. C. *et al.* Antisense and antigene properties of peptide nucleic acids. *Science (80-.).* **258,** 1481–1485 (1992).

254. Nielsen, P. E., Egholm, M. & Buchardt, O. Sequence-specific transcription arrest by peptide nucleic acid bound to the DNA template strand. *Gene* **149,** 139–145 (1994).

255. Gambacorti-Passerini, B. C. *et al.* In Vitro Transcription and Translation Inhibition by Anti-Promyelocytic Leukemia (PML)/Retinoic Acid Receptor Alpha and Anti-PML Peptide Nucleic Acid. *Blood* **88,** 1411–1417 (1996).

256. Demidov, V. V. *et al.* Stability of peptide nucleic acids in human serum and cellular extracts. *Biochem. Pharmacol.* **48,** 1310–1313 (1994).

257. Egholm, M. *et al.* PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules. *Nature* **365,** 566–568 (1993).

258. Quijano, E., Bahal, R., Ricciardi, A., Saltzman, W. M. & Glazer, P. M. Therapeutic peptide nucleic acids: Principles, limitations, and opportunities. *Yale J. Biol. Med.* **90,** 583–598 (2017).

259. Wright, D. G., Zhang, Y. & Murphy, J. R. Effective delivery of antisense peptide nucleic acid oligomers into cells by anthrax protective antigen. *Mol. Med.* **376,** 200–205 (2009).

260. Abudayyeh, O. O. *et al.* RNA targeting with CRISPR–Cas13. *Nature* **550,** 280–284 (2017).

261. Zhao, X. *et al.* A CRISPR-Cas13a system for efficient and specific therapeutic targeting of mutant KRAS for pancreatic cancer treatment. *Cancer Lett.* **431,** 171–181 (2018).

262. Sullivan, I. & Planchard, D. Next-Generation EGFR Tyrosine Kinase Inhibitors for Treating EGFR-Mutant Lung Cancer beyond First Line. *Front. Med.* **3,** 1–13 (2017).

263. Rost, S. *et al.* Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature* **427,** 537–541 (2004).

264. Jia, Y. *et al.* Overcoming EGFR(T790M) and EGFR(C797S) resistance with mutant-selective allosteric inhibitors. *Nature* **534,** 129–132 (2016).

265. Toure, M. & Crews, C. M. Small-molecule PROTACS: New approaches to protein degradation. *Angew. Chemie - Int. Ed.* **55,** 1966–1973 (2016).

266. Lonsdale, R. & Ward, R. A. Structure-based design of targeted covalent inhibitors. *Chem. Soc. Rev.* **47,** 3816–3830 (2018).

267. Roberts, A. M. *et al.* Chemoproteomic Screening of Covalent Ligands Reveals UBA5 As a Novel Pancreatic Cancer Target. *ACS Chem. Biol.* **12,** 899–904 (2017).

268. Anderson, K. E., To, M., Olzmann, J. A. & Nomura, D. K. Chemoproteomics-Enabled Covalent Ligand Screening Reveals a Thioredoxin-Caspase 3 Interaction Disruptor That Impairs Breast Cancer Pathogenicity. *ACS Chem. Biol.* **12,** 2522–2528 (2017).

269. Boudet, J., Devillier, J. C., Allain, F. H. T. & Lipps, G. Structures to complement the archaeo-eukaryotic primases catalytic cycle description: What's next? *Comput. Struct. Biotechnol. J.* **13,** 339–351 (2015).

270. Ilic, S. *et al.* Identification of DNA primase inhibitors via a combined fragment-based and virtual screening. *Sci. Rep.* **6,** 1–10 (2016).

271. Bakail, M. & Ochsenbein, F. Targeting protein-protein interactions, a wide open field for drug design. *Comptes Rendus Chim.* **19,** 19–27 (2016).

272. Lao, B. B. *et al.* Rational design of topographical helix mimics as potent inhibitors of protein-protein interactions. *J. Am. Chem. Soc.* **136,** 7877–7888 (2014).

273. Vassilev, L. T. *et al.* In Vivo Activation of the p53 Pathway by Small-MoleculeAntagonists of MDM2. *Science (80-.).* **303,** 844–848 (2004).

274. Fasan, R. *et al.* Using a β-hairpin to mimic an α-helix: Cyclic peptidomimetic inhibitors of the p53-HDM2 protein-protein interaction. *Angew. Chemie - Int. Ed.* **43,** 2109–2112 (2004).

275. Moellering, R. E. *et al.* Direct inhibition of the NOTCH transcription factor complex. *Nature* **462,** 182–188 (2009).

276. Loregian, A., Marsden, H. S. & Palu, G. Protein-protein interactions as targets for antiviral chemotherapy. *Rev. Med. Virol.* **12,** 239–262 (2002).

277. Sancak, Y. *et al.* The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1. *Science (80-.).* **320,** 1496–1501 (2008).

278. Hardy, G. H. Mendelian Proportions in a Mixed Population. *Science* **28,** 49–50 (1908).

279. Weinberg, W. Uber den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkd. Württemberg.* **64,** 368–382 (1908).

280. Nijman, S. M. B. Synthetic lethality: General principles, utility and detection using genetic screens in human cells. *FEBS Lett.* **585,** 1–6 (2011).

281. Hartwell, L. H., Szankasi, P., Roberts, C. J., Murray, A. W. & Stephen H. Friend. Integrating Genetic Approaches into the Discovery of Anticancer Drugs. *Science (80-.).* **278,** 1064–1068 (1997).

282. Hein, D. W. Molecular genetics and function of NAT1 and NAT2: Role in aromatic amine metabolism and carcinogenesis. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **506–507,** 65–77 (2002).

283. Rendo, V. Targeting allelic loss in colorectal cancer. (Uppsala University, 2018).

284. Vaithiyalingam, S. *et al.* Insights into eukaryotic primer synthesis from structures of the p48 subunit of human DNA primase. *J. Mol. Biol.* **426,** 558–569 (2014).

285. Liu, Q., Greimann, J. C. & Lima, C. D. Reconstitution, Activities, and Structure of the Eukaryotic RNA Exosome. *Cell* **127,** 1223–1237 (2006).

286. Napetschnig, J. *et al.* Structural and functional analysis of the interaction between the nucleoporin Nup214 and the DEAD-box helicase Ddx19. *Proc. Natl. Acad. Sci.* **106,** 3089–3094 (2009).

APPENDIX

This page intentionally left blank.

Supplemental Material

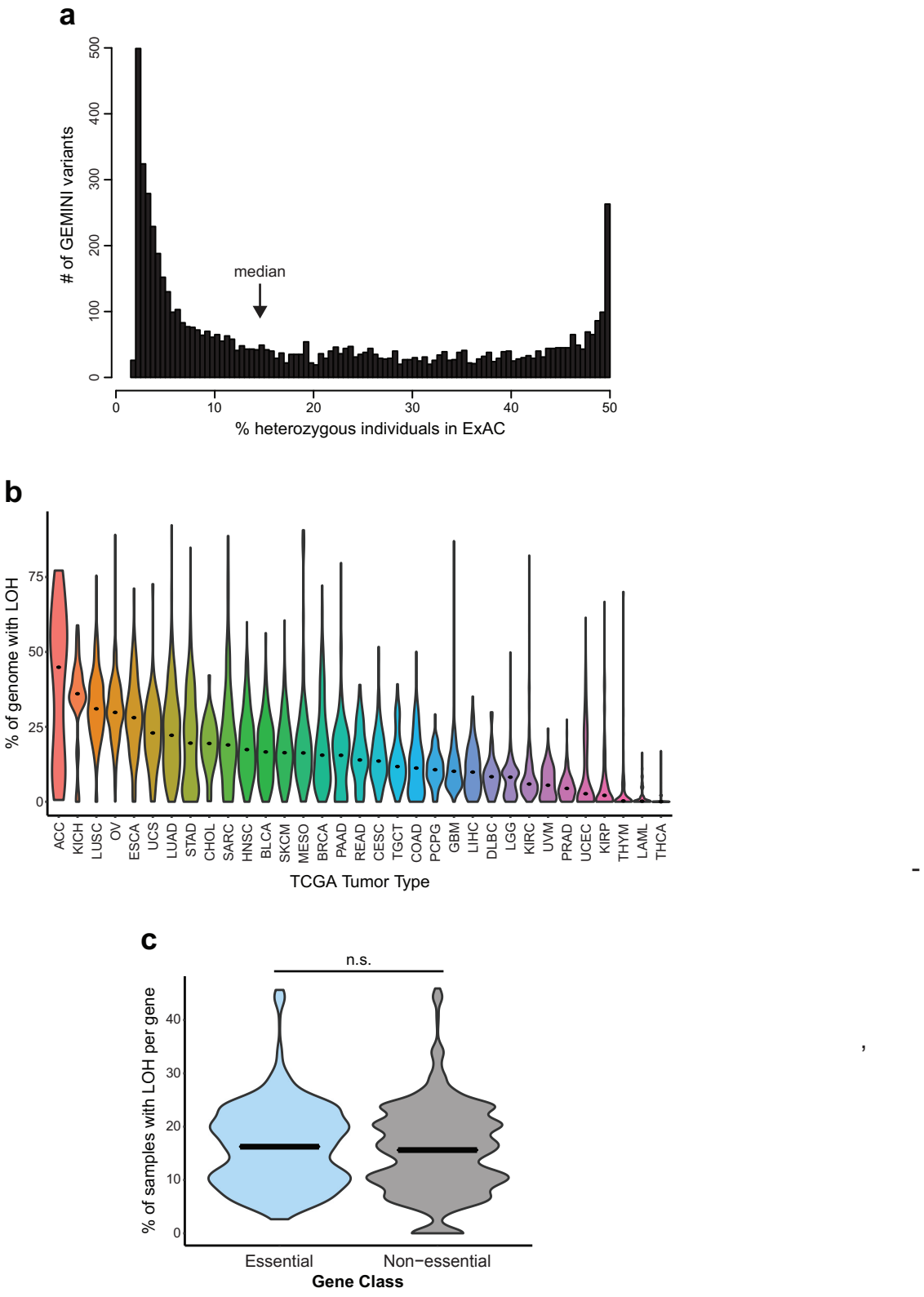
Supplemental Tables 1–5 can be found at the following stable URL:
https://www.biorxiv.org/content/10.1101/534529v2.supplementary-material

This page intentionally left blank.

Supplementary Figure 1.

a. Number of GEMINI variants (vertical axis) plotted against the fraction of individuals that are heterozygous for each variant in the ExAC cohort (horizontal axis). Bin width = 0.5%. **b.** Violin plot of the percent of genome affected by LOH across 33 TCGA tumor types. Tumor types are indicated by TCGA abbreviations (see https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations). Plot width represents relative sample density and dots indicate median values. **c.** Violin plot demonstrating the rate of LOH for essential (blue) and non-essential genes (grey). Essential genes do not have a significantly higher rate of LOH (One-tailed Student's t-test, p=1 [n.s. = not significant]). Intersecting lines indicate median values.

Supplementary Figure 1 (Continued).



Supplementary Figure 2.

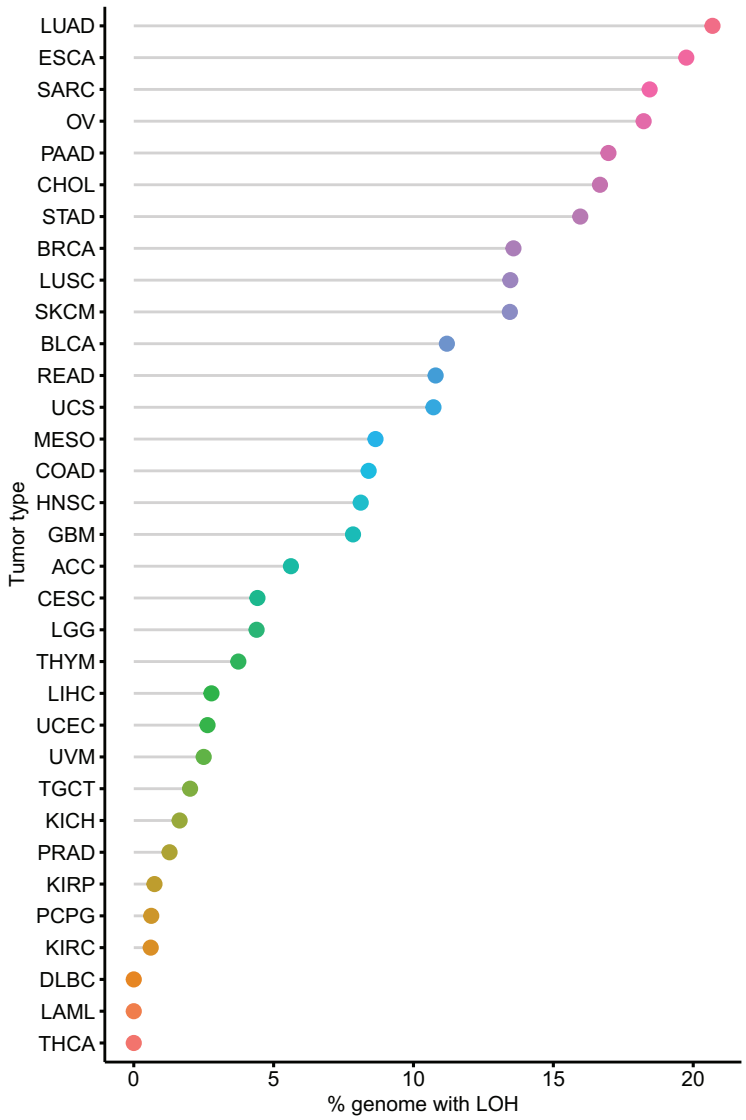
a. Table of *PRIM1*^{rs2277339} SNP statistics, including calculation of theoretical number of treatable patients per year in the US (see Methods). **b.** Cleveland dot plot of the rate of *PRIM1* LOH across 33 TCGA tumor types. Tumor types are indicated by TCGA abbreviations (see https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations). **c.–f.** Growth of indicated patient-derived cell lines expressing LacZ (black), PRIM1 NA (red), or PRIM1 AS (blue) sgRNA as measured by CellTiter-Glo luminescence, relative to day of assay plating. n = 5 technical replicates; error bars represent s.d. **g.–k.** Growth of indicated isogenic cell lines expressing LacZ (black), PRIM1 NA (red), or PRIM1 AS (blue) sgRNA as measured by CellTiter-Glo luminescence, relative to day of assay plating. n = 5 technical replicates; error bars represent s.d.

Supplementary Figure 2 (Continued).

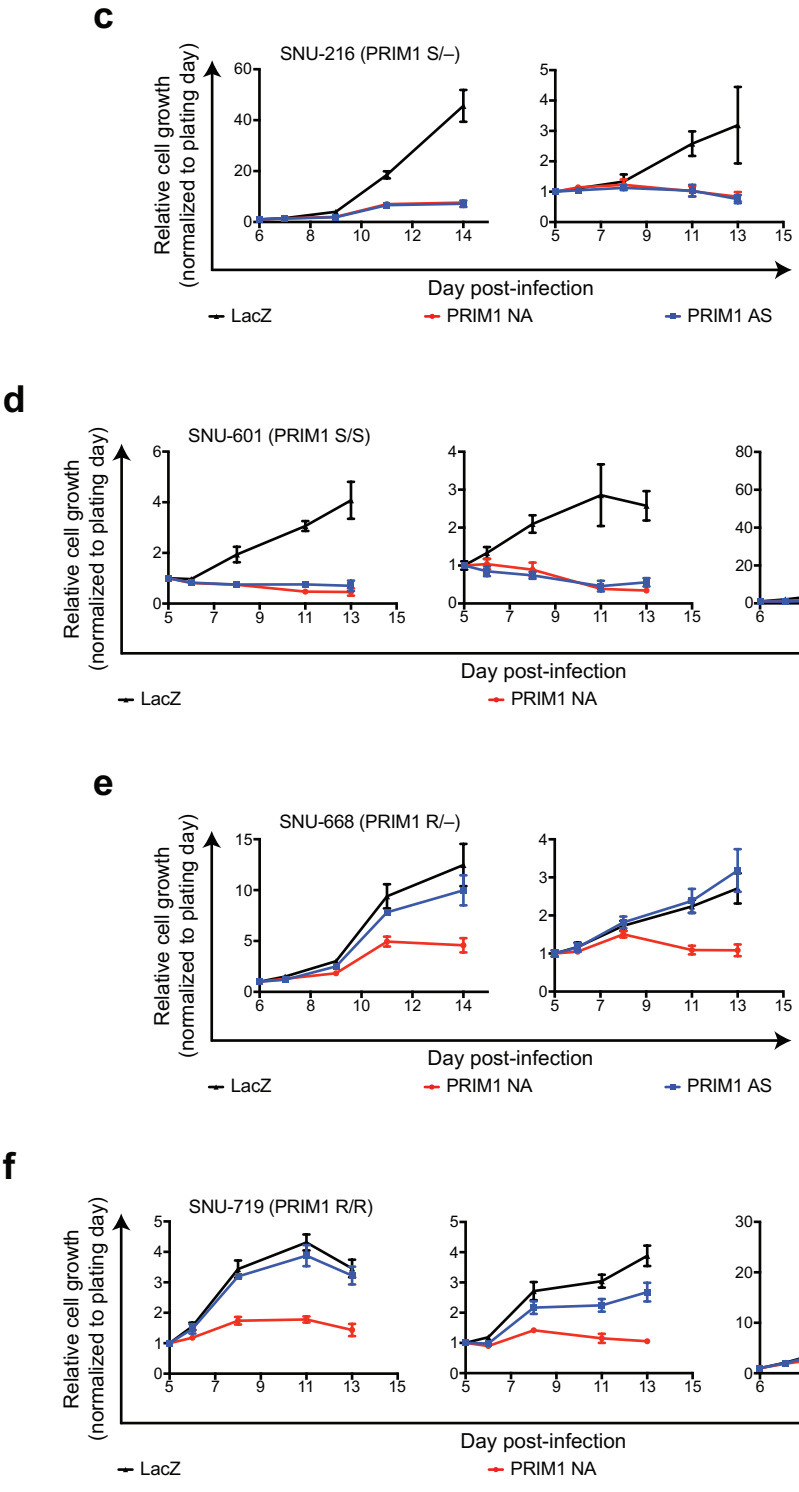
a

PRIM1[rs2277339]	
position	chr12:57146069
variant	T/G
consequence	missense variant (D5A)
minor allele frequency	0.177
predicted frequency of heterozygosity	0.291
pan-cancer LOH frequency	0.089
theoretical US patients per year	22,470

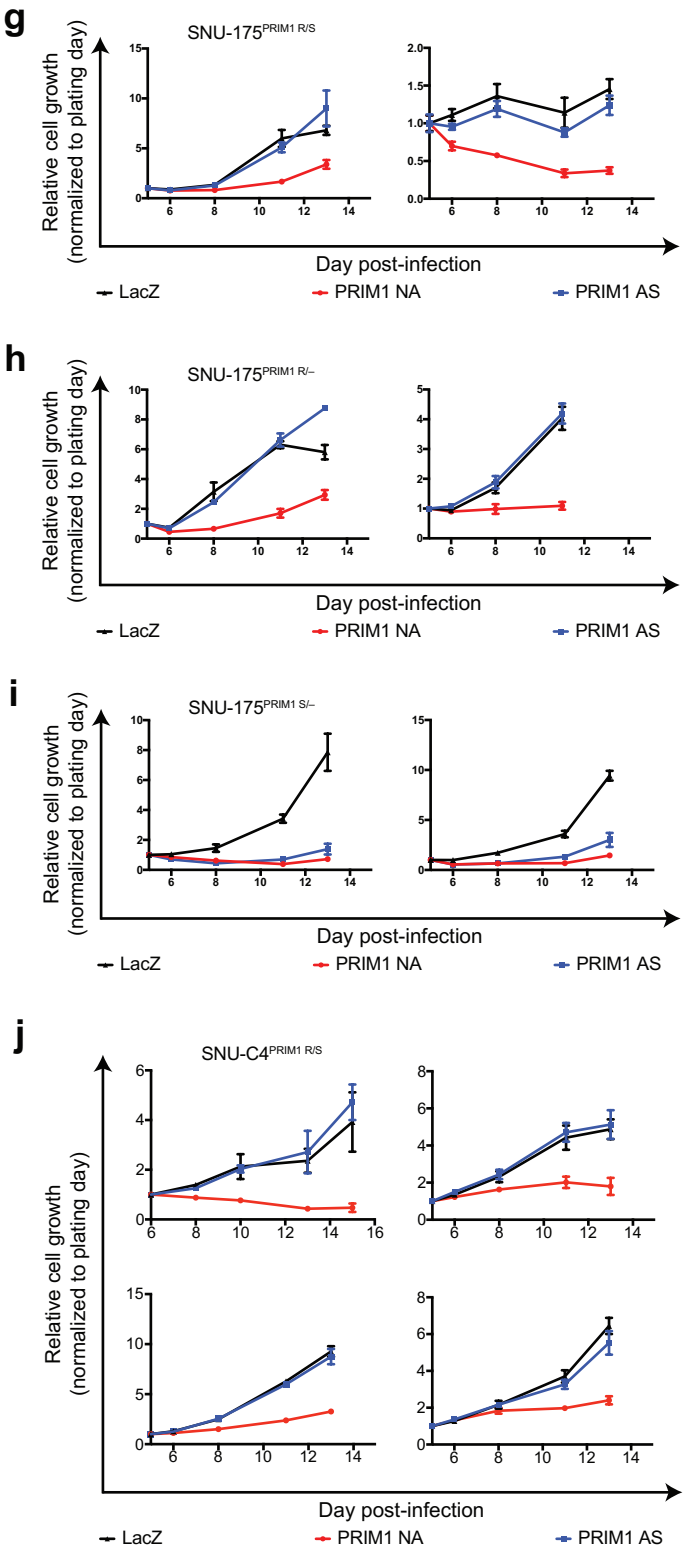
b



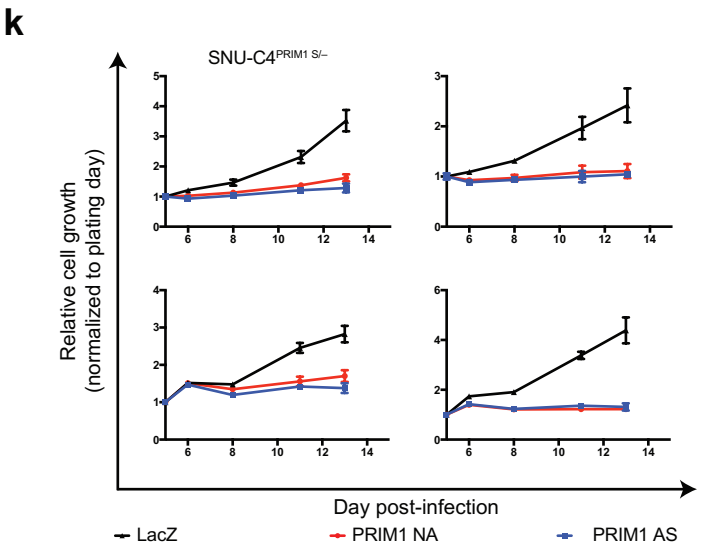
Supplementary Figure 2 (Continued).



Supplementary Figure 2 (Continued).



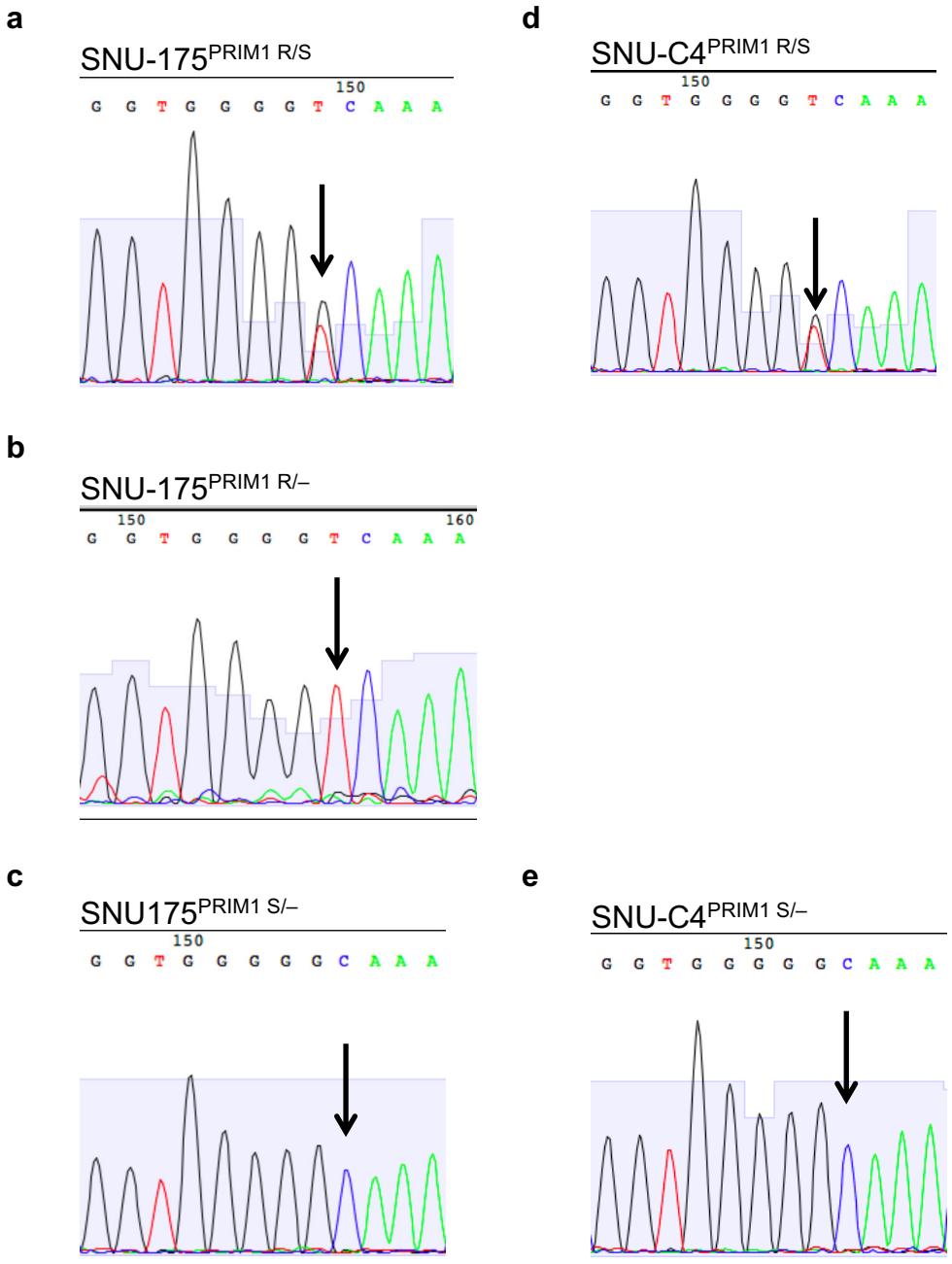
Supplementary Figure 2 (Continued).



Supplementary Figure 3.

a.–e. Genotyping of *PRIM1*^{rs2277339} locus in indicated isogenic cell lines by Sanger sequencing. Sequencing peaks representing SNP loci indicated by black arrow. **f.–i.** Genotyping of *EXOSC8*^{rs117135638} locus in indicated isogenic cell lines by next generation sequencing. Relevant sequence features indicated by font or highlighting color. Off-target reads and species representing less than 0.05% of total read pairs removed for clarity.

Supplementary Figure 3 (Continued).



Supplementary Figure 3 (Continued).

resistant (R) allele
sensitive (S) allele
editing sgRNA blocking mutation
synonymous mutation introducing MnlI cut site
insertion
deletion

f

DV-90$^{EXOSC8\ R/R}$

Reference

```
AGCTGCAGAGTGTTTCTTTCAGTTCCTAATGTGGATCTACCACCCCTGTGTTCATCGAGA
```

Edited R allele, 88,509 read pairs

```
AGCTGCAGAGTGTTTCTTTCAGTTCCTAATGTGGATCTACCTCATCTGTGTtcatcgaga
**************************************** *  **************
```

g

DV-90$^{EXOSC8\ R/\Delta}$

Reference

```
AGCTGCAGAGTGTTTCTTTCAGTTCCTAATGTGGATCTACCACCCCTGTG-TTCATCGAG
```

Edited R allele, 44,454 read pairs

```
AGCTGCAGAGTGTTTCTTTCAGTTCCTAATGTGGATCTACCTCATCTGTG-Ttcatcgag
```

Disrupted S allele, 42,169 read pairs

```
AGCTGCAGAGTGTTTCTTTCAGTTCCTAATGTGGATCTACCACCCCTGTGTTtcatcgag
**************************************** *  ***** *********
```

h

DV-90$^{EXOSC8\ S/S}$

Reference

```
AGCTGCAGAGTGTTTCTTTCAGTTCCTAATGTGGATCTACCACCCCTGTGTTCATCGAGA
```

Edited S allele, 54,799 read pairs

```
AGCTGCAGAGTGTTTCTTTCAGTTCCTAATGTGGATCTACCTCCGCTGTGTtcatcgaga
**************************************** ** ***************
```

i

DV-90$^{EXOSC8\ S/\Delta}$

Reference

```
AGCTGCAGAGTGTTTCTTTCAGTTCCTAATGTGGATCTACCACCCCTGTGTTCATCGAGA
```

Disrupted S allele, 22,831 read pairs

```
AGCTGCAGAGTGTTTCTTTCAGTTCCTAATGTGGATCTACCACCCCTGTG-TCAtcgaga
```

Edited S allele, 23,409 read pairs

```
AGCTGCAGAGTGTTTCTTTCAGTTCCTAATGTGGATCTACCTCCGCTGTGTTCAtcgaga
**************************************** ** ***** *********
```

Supplementary Figure 4.

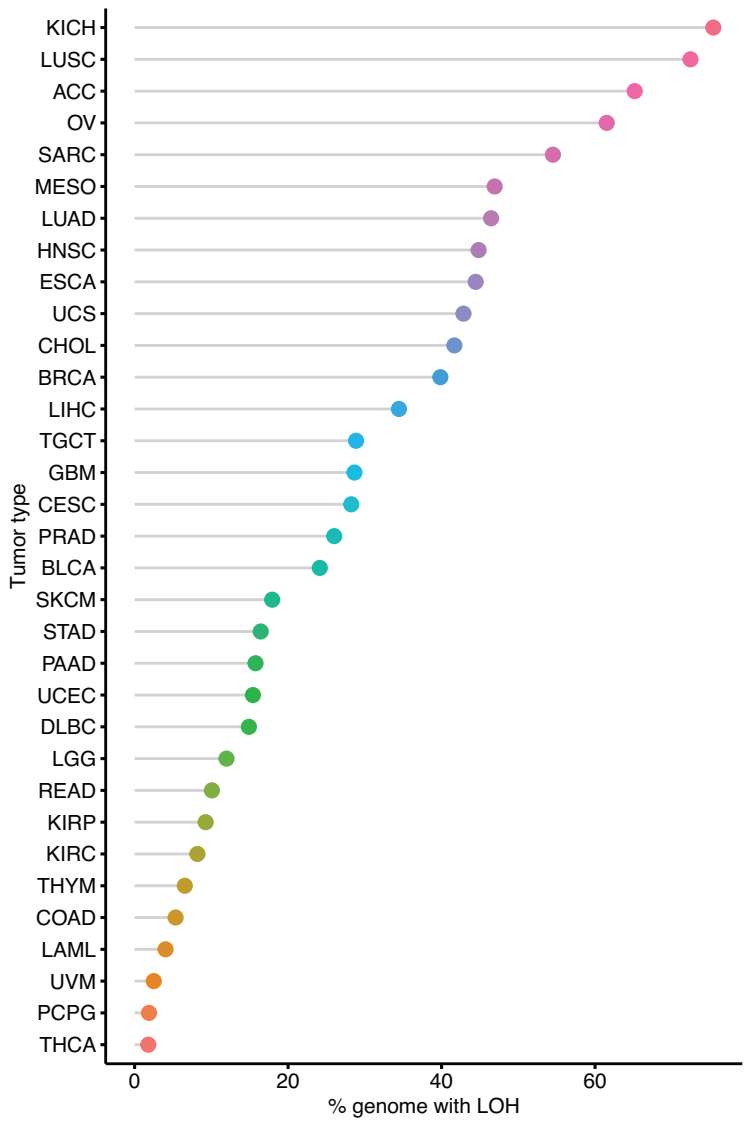
a. Table of *EXOSC8*^{rs117135638} SNP statistics, including calculation of theoretical number of treatable patients per year in the US (see Methods). **b.** Cleveland dot plot of the rate of *EXOSC8* LOH across 33 TCGA tumor types. Tumor types are indicated by TCGA abbreviations (see https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations). **c.–f.** Growth of indicated patient-derived cell lines expressing LacZ (black), EXOSC8 NA (red), or EXOSC8 AS (blue) sgRNA as measured by CellTiter-Glo luminescence, relative to day of assay plating. n = 5 technical replicates; error bars represent s.d. **g.–j.** Growth of indicated isogenic cell lines expressing LacZ (black), EXOSC8 NA (red), or EXOSC8 AS (blue) sgRNA as measured by CellTiter-Glo luminescence, relative to day of assay plating. n = 5 technical replicates; error bars represent s.d.

Supplementary Figure 4 (Continued).

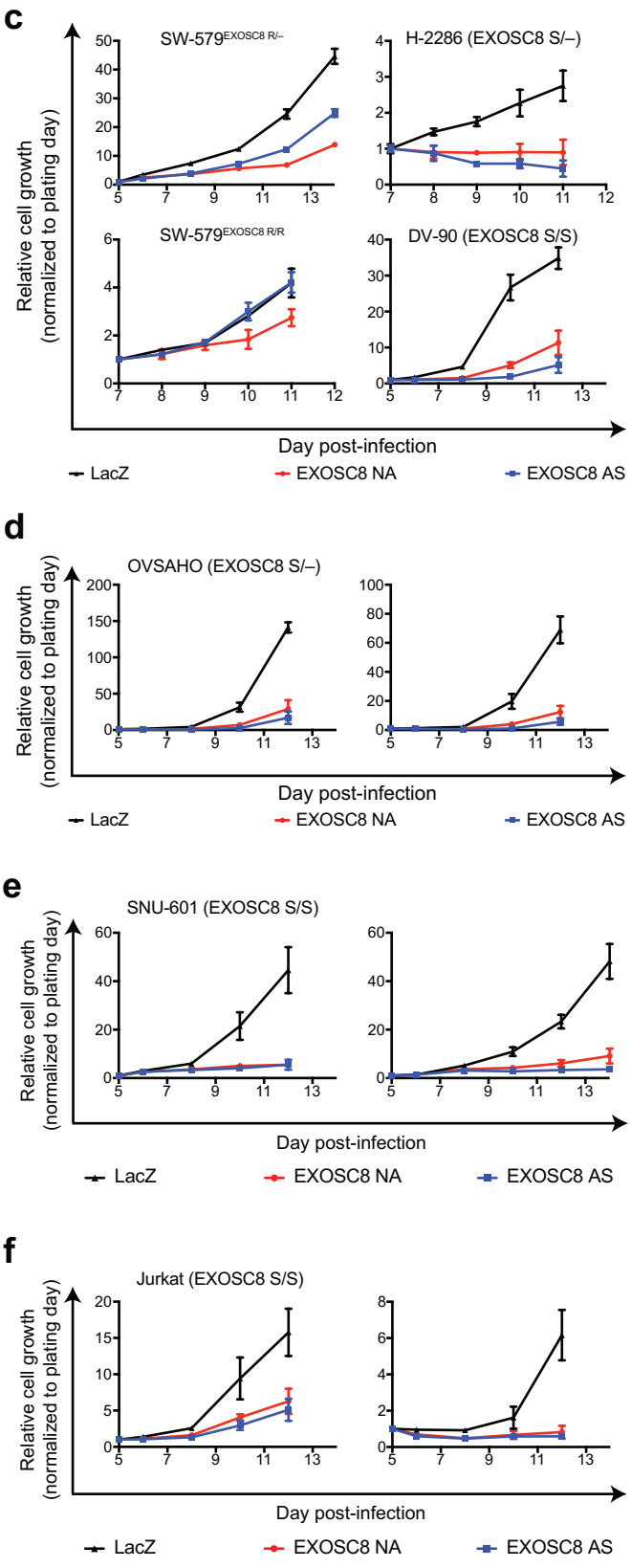
a

EXOSC8 rs117135638	
position	13:37580078
variant	C/A
consequence	missense variant (P87H)
minor allele frequency	0.011
predicted frequency of heterozygosity	0.021
pan-cancer LOH frequency	0.291
theoretical US patients per year	5307

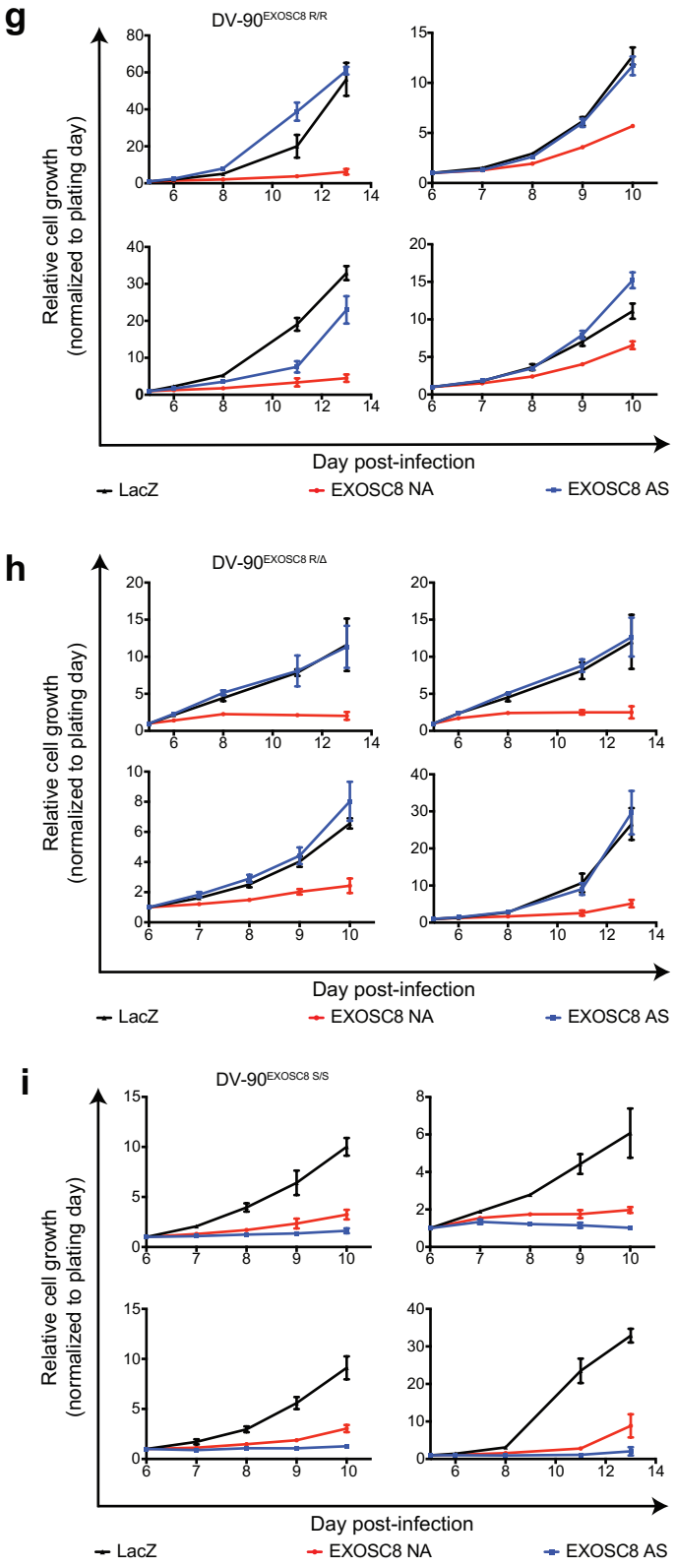
b



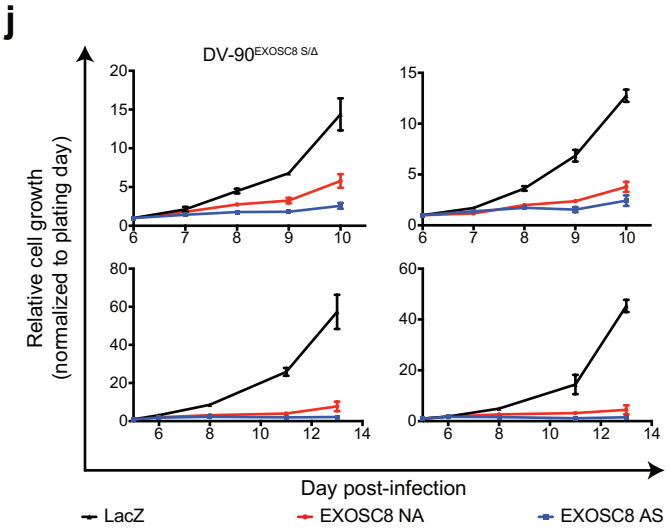
Supplementary Figure 4 (Continued).



Supplementary Figure 4 (Continued).



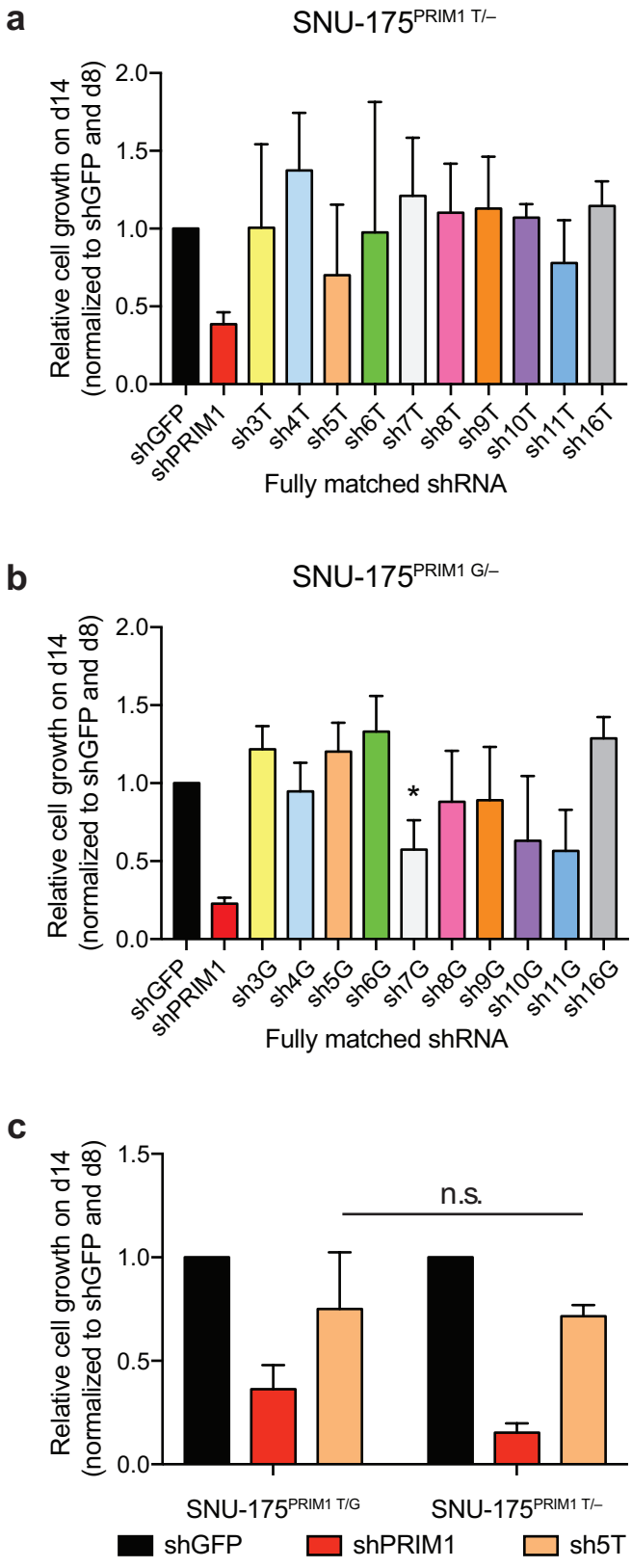
Supplementary Figure 4 (Continued).



Supplementary Figure 5.

a. Growth of cells hemizygous for the major *PRIM1*^{rs2277339} allele (SNU-175$^{PRIM1\ T/-}$) expressing shGFP, positive control shPRIM1, or indicated putative allele-specific shRNAs targeting the T allele of *PRIM1*^{rs2277339}, as measured by CellTiter-Glo luminescence, relative to day of assay plating. n = 3 biological replicates; error bars represent s.d. **b.** Growth of cells hemizygous for the minor *PRIM1*^{rs2277339} allele (SNU-175$^{PRIM1\ G/-}$) expressing shGFP, positive control shPRIM1, or indicated putative allele-specific shRNAs targeting the G allele of *PRIM1*^{rs2277339} as measured by CellTiter-Glo luminescence, relative to day of assay plating. n = 3 biological replicates; error bars represent s.d. One-tailed Student's t-test, *$p < 0.05$. **c.** Growth of heterozygous cells (SNU-175$^{PRIM1\ T/G}$) and cells hemizygous for the major *PRIM1*^{rs2277339} allele (SNU-175$^{PRIM1\ T/-}$) expressing shGFP, positive control shPRIM1, or indicated putative allele-specific shRNAs targeting the T allele of *PRIM1*^{rs2277339} as measured by CellTiter-Glo luminescence, relative to day of assay plating. n = 4 biological replicates; error bars represent s.d. One-tailed Student's t-test, n.s. = not significant. **d.** Growth of heterozygous cells (SNU-175$^{PRIM1\ T/G}$) and cells hemizygous for the minor *PRIM1*^{rs2277339} allele (SNU-175$^{PRIM1\ G/-}$) expressing shGFP, positive control shPRIM1, or indicated putative allele-specific shRNAs targeting the G allele of *PRIM1*^{rs2277339} as measured by CellTiter-Glo luminescence, relative to day of assay plating. n = 3 biological replicates; error bars represent s.d. One-tailed Student's t-test, n.s. = not significant.

Supplementary Figure 5 (Continued).



Supplementary Figure 5 (Continued).