

Loss of heterozygosity of essential genes represents a widespread class of potential cancer vulnerabilities

Caitlin A. Nichols^{1,2,3}, William J. Gibson^{3,4}, Meredith S. Brown^{1,2}, Jack A. Kosmicki^{5,6,7,8}, John P. Busanovich^{1,2}, Hope Wei^{1,2}, Laura M. Urbanski^{1,2}, Naomi Curimjee^{1,2}, Ashton C. Berger^{2,3}, Galen F. Gao³, Andrew D. Cherniack^{2,3}, Sirano Dhe-Paganon^{1,9}, Brenton R. Paoella^{1,3,10}✉ & Rameen Beroukhim^{1,2,3,4,10}✉

Alterations in non-driver genes represent an emerging class of potential therapeutic targets in cancer. Hundreds to thousands of non-driver genes undergo loss of heterozygosity (LOH) events per tumor, generating discrete differences between tumor and normal cells. Here we interrogate 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 identified 5664 variants in 1278 essential genes that undergo LOH in cancer and evaluated the potential for each to be targeted using allele-specific gene-editing, RNAi, or small-molecule approaches. We further show that allele-specific inactivation of either of two essential genes (*PRIM1* and *EXOSC8*) reduces growth of cells harboring that allele, while cells harboring the non-targeted allele remain intact. We conclude that LOH of essential genes represents a rich class of non-driver cancer vulnerabilities.

¹Departments of Cancer Biology, Boston, MA, USA. ²Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA 02215, USA. ³Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. ⁴Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA. ⁵Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA. ⁶Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. ⁷Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA. ⁸Program in Bioinformatics and Integrative Genomics, Harvard University, Cambridge, MA 02138, USA. ⁹Department of Biological Chemistry & Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. ¹⁰These authors jointly supervised this work: Brenton R. Paoella, Rameen Beroukhim. ✉email: paoella@broadinstitute.org; rameen_beroukhim@dfci.harvard.edu

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^{1,2}, 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 allele³ (Fig. 1a). We term this target class GEMINI vulnerabilities, after the twins from Greek mythology Castor and Pollux, one of which was mortal and the other immortal.

While previous reports have described individual GEMINI vulnerabilities^{4,5}, 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 potential GEMINI vulnerabilities, and at what rates these vulnerabilities occur. Moreover, 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. Furthermore, GEMINI vulnerabilities have never been validated in isogenic systems to confirm specificity.

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. We also performed proof-of-principle validation of GEMINI vulnerabilities for two candidate genes in this list, *PRIMI* and *EXOSC8*, using allele-specific CRISPR in both patient-derived and isogenic models. These results rigorously validate the GEMINI class of vulnerabilities and define its potential scope.

Results

Genome-wide identification of GEMINI vulnerabilities. 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 genes that are essential across lineages (Supplementary Data 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^{6,7}, version 6.8; Supplementary Data 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 database⁸. 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; Fig. 1b). However, essential genes still contain an abundance of genetic variation: 86% (1278/1481) harbor at least one common germline variant (Fig. 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 Fig. 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^{9,10}. 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 or in some cases even increases due to a duplication event. LOH has been frequently described^{10,11}, 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 9686 patient samples across 33 TCGA tumor types (Methods)¹². On average and across all cancers, 16% of genes undergo LOH (Fig. 1d). Genome-wide LOH rates vary widely by tumor type, ranging from a median of 45% in adenoid cystic carcinoma to 0.01% in thyroid carcinoma (Supplementary Fig. 1b). Approximately 28% of genes undergoing LOH undergo copy-neutral LOH (Fig. 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 Fig. 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 (Fig. 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 (Fig. 1a).

Our analysis identified 5664 polymorphisms in 1278 cell-essential genes, representing a compendium of potential GEMINI vulnerabilities (Supplementary Data 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^{6,7}, version 6.8; Supplementary Data 4). Among the 5664 GEMINI variants, 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.

Validation of *PRIMI*^{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^{13–15}. 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

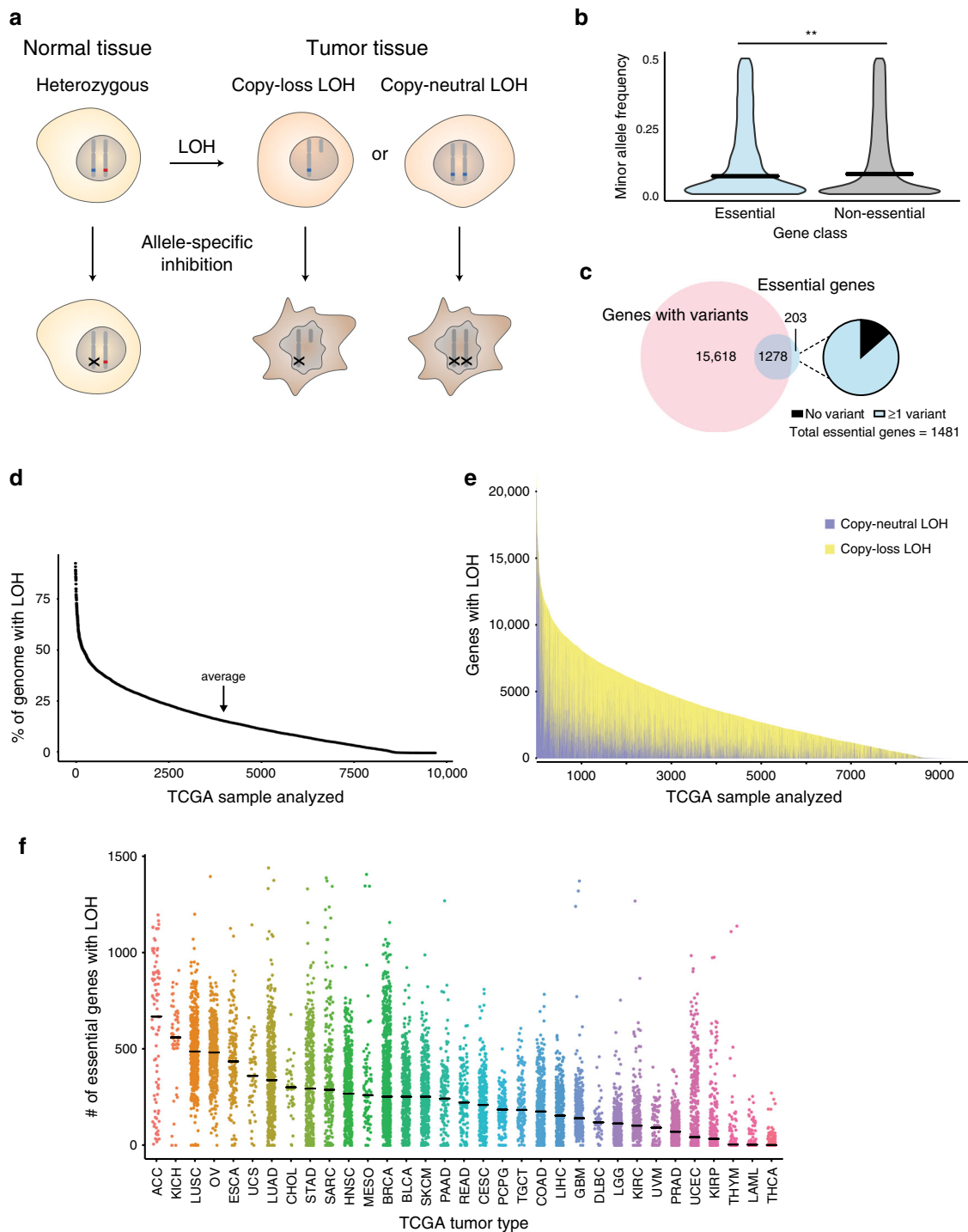
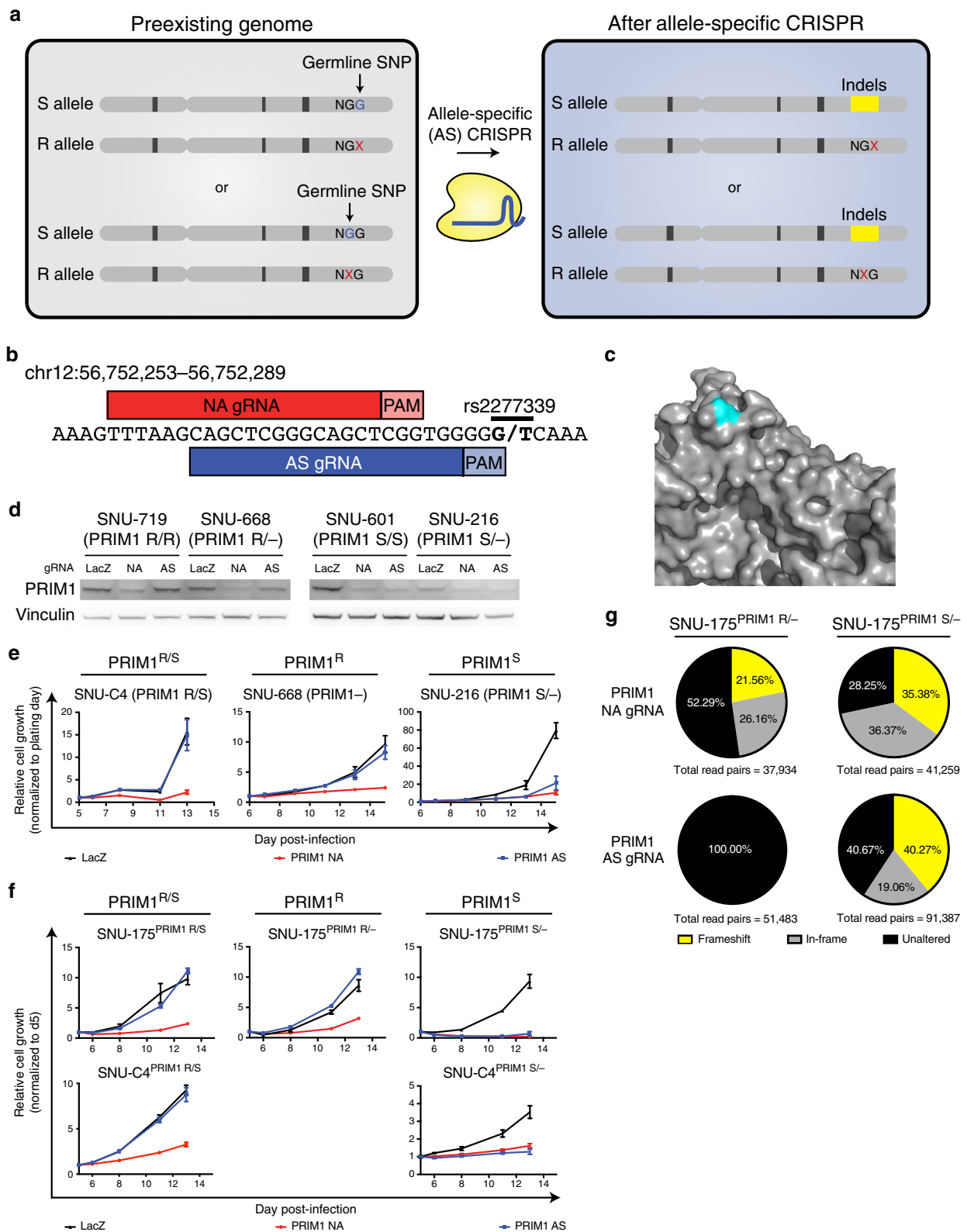


Fig. 1 Genomic rates of LOH and allelic variation in normal and cancer genomes. **a** Schematic indicating how loss of heterozygosity (LOH) of essential genes represents a potentially targetable difference between cancer and normal cells. **b** Violin plot of minor allele frequency of polymorphisms in 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, $**p = 0.005$. **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 9686 cancers from TCGA. **e** Stacked histogram representing the number of genes with copy-loss (yellow) or copy-neutral LOH (purple) across 9686 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.

abrogate Cas9-mediated target cleavage^{16,17}. 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 (Fig. 2a).

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^{18–20}. 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 (Fig. 2b–c, Supplementary Fig. 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 Fig. 2b).

PRIM1^{rs2277339} lies in a polymorphic PAM site—the “CRISPR-sensitive,” G allele generates a canonical *S. pyogenes* Cas9 PAM

Fig. 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 product⁸⁸ shows the amino acid encoded by rs2277339 (teal) lies on the surface of the primase catalytic subunit (gray) 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 ($n = 1$ biological replicate). **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. Data are presented as mean values \pm s.d. See Supplementary Fig. 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. Data are presented as mean values \pm s.d. See Supplementary Fig. 2 for additional biological replicates. **g** Disruption of *PRIM1* in isogenic hemizygous *PRIM1* resistant (*PRIM1*^R) or *PRIM1* sensitive (*PRIM1*^S) cells expressing *PRIM1* NA or AS sgRNA. Unaltered alleles (black), alleles with in-frame insertions or deletions (gray), and alleles with frameshift alterations (yellow) were assessed by deep sequencing of *PRIM1* four days post-infection with sgRNA. Source data for Fig. 2d–g are provided as a Source Data file.

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 (Fig. 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 rs2277339 allele with AS sgRNA and verified that AS sgRNA disrupts *PRIM1* in *PRIM1*^S genetic contexts (Fig. 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 (Fig. 2e, Supplementary Fig. 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 (Fig. 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; Fig. 2e, Supplementary Fig. 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 Fig. 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 (Fig. 2f, Supplementary Fig. 2g–k).

Within these isogenic lines, we also confirmed that AS CRISPR disrupts *PRIM1* in a *PRIM1*^{rs2277339}-dependent manner using deep sequencing. In order to assay *PRIM1* knockout efficiency before cells started to die due to loss of the essential *PRIM1* protein, we performed deep sequencing on DNA collected from cells at the early time point of four days post-infection. At this

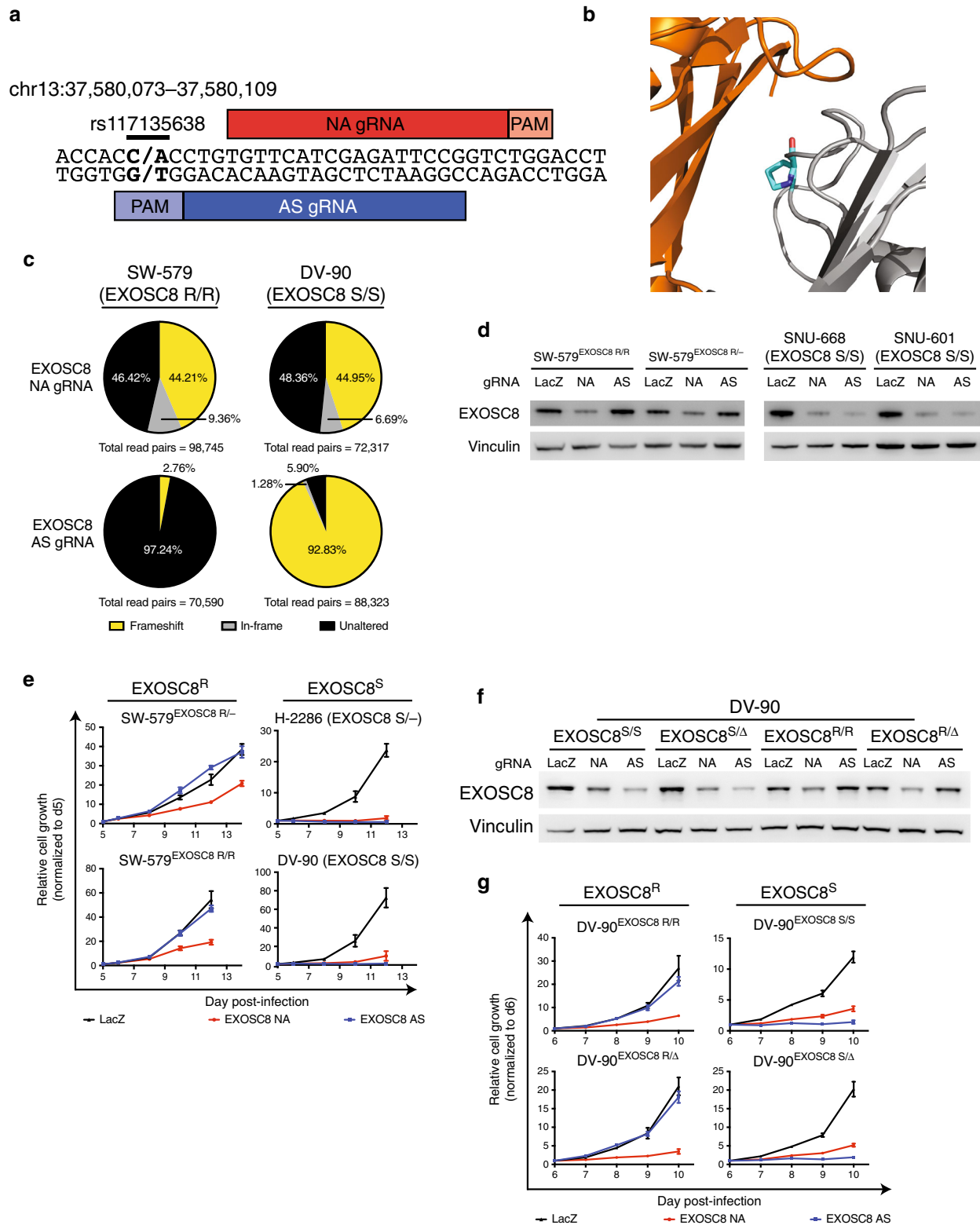
early time point, isogenic *PRIM1*^S and *PRIM1*^R cells infected with the NA sgRNA showed comparable fractions of disrupted alleles (Fig. 2g), suggesting both lines exhibit similar levels of Cas9 activity. However, while *PRIM1*^S cells expressing AS sgRNA showed approximately 40% disrupted alleles, resistant cells under the same condition showed 0 disrupted alleles ($p < 0.0001$, Chi-square with Yates correction; Fig. 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 (Supplementary Fig. 2l, Supplementary Note 1).

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^{21–23}. 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 (Fig. 3a, b, Supplementary Fig. 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 Fig. 4b).

We first tested allele-specific (AS) *EXOSC8* disruption using an sgRNA 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. Using patient-derived cell lines, we verified the allele-specific effects of *EXOSC8* AS sgRNA on both the DNA and the protein level (Fig. 3c, d, Supplementary Note 2). Consistent with these observations, *EXOSC8*^{S/-} and *EXOSC8*^{S/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 (Fig. 3e, Supplementary Fig. 4c–f, Supplementary Note 3).

We next 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 Fig. 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 (Fig. 3f). EXOSC8^{S/S} and EXOSC8^{S/Δ} cells expressing AS sgRNA showed decreased proliferation relative to LacZ-targeting control, whereas cells retaining the resistant allele (EXOSC8^{R/R} and EXOSC8^{R/Δ}) showed no such defects (Fig. 3g, Supplementary Fig. 4g–j).

Potential approaches to targeting GEMINI vulnerabilities. We were interested in understanding the potential scope of patients that could benefit from therapeutic approaches targeting GEMINI vulnerabilities. For each GEMINI variant, we calculated the number of new patients in the US per year 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 US patients per year (Fig. 4a). *PRIM1*^{rs2277339}

Fig. 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, Rrp43⁸⁹ (gray) shows the amino acid encoded by rs117135638 (teal) lies on the surface of the Rrp43 protein near the interface with exosome complex subunit Mtr3 (orange). **c** 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. Unaltered alleles (black), alleles with in-frame insertions or deletions (gray), and alleles with frameshift alterations (yellow) were assessed by deep sequencing of EXOSC8 four days post-infection with sgRNA. **d** Immunoblot of EXOSC8 protein levels in indicated patient-derived and isogenic cell lines expressing LacZ, EXOSC8 NA, or EXOSC8 AS sgRNA ($n = 2$ technical replicates of 1 biological sample). **e** 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. Data are presented as mean values \pm s.d. See Supplementary Fig. 4 for additional biological replicates. **f** Immunoblot of EXOSC8 protein levels in indicated isogenic cell lines expressing LacZ, EXOSC8 NA, or EXOSC8 AS sgRNA ($n = 2$ technical replicates of 1 biological sample). **g** 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. Data are presented as mean values \pm s.d. See Supplementary Fig. 4 for additional biological replicates. Source data for Fig. 3c–g are provided as a Source Data file.

and EXOSC8^{rs117135638} could be targetable in a theoretical 22,470 and 5,307 patients per year in the US, respectively (Supplementary Figs. 2a and 4a).

The major challenge to exploiting GEMINI vulnerabilities is identifying means to target them in humans. Three 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 and, when practical, performed proof-of-concept in vitro experiments testing each approach.

We first focused on CRISPR effectors as a potential therapeutic modality for targeting GEMINI vulnerabilities. Because we had already demonstrated that *S. pyogenes* Cas9 can disrupt *PRIMI* and *EXOSC8* in an allele-specific manner, we next analyzed the list of GEMINI vulnerabilities to identify the full range of polymorphisms whose targeting on the DNA level may enable allele-selective gene disruption by a CRISPR-based approach. For this analysis, we included both the canonical *S. pyogenes* PAM, NGG, as well as the weaker, non-canonical PAM, NAG^{17,24}. Of the 4648 GEMINI vulnerabilities in open reading frames, 23% (1088/4648, or 19% of all GEMINI vulnerabilities) generate a PAM site in one allele but not the other, suggesting the potential for allele-specific knockout (Supplementary Data 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 *PRIMI*^{rs2277339} may be targetable in an allele-specific manner using RNAi. For these experiments, we refer to the *PRIMI* alleles by their identifying nucleotide; for example, heterozygous cells are referred to as *PRIMI*^{T/G}, and cells hemizygous for the minor allele are referred to as *PRIMI*^{G/-}. We first sought to determine the level of *PRIMI* knockdown required to substantially decrease cell proliferation. Accordingly, we infected hemizygous and heterozygous isogenic cells with non-allele specific *PRIMI*-targeting shRNAs and assessed *PRIMI* expression and cell growth. We observed that substantial decreases in cell growth were possible under conditions of robust *PRIMI* knockdown (>80%) (Fig. 4b).

We then asked whether allele-specific shRNAs targeting the *PRIMI*^{rs2277339} locus could decrease growth in cells representing the fully matched genotype. *PRIMI*^{T/-} and *PRIMI*^{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 Fig. 5a, b). We then selected the four *PRIMI* 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 (*PRIMI*^{T/G}) and hemizygous cells of the targeted genotype (*PRIMI*^{T/-} or *PRIMI*^{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 Fig. 5c, d). We conclude that *PRIMI*^{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^{25–27}. 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 lines²⁸, 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 *PRIMI* (median DEMETER2 score = -0.52), fit this category, representing 35% (1804/5196) of GEMINI vulnerabilities (Supplementary Data 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^{21,22,23}. 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. In comparison to the

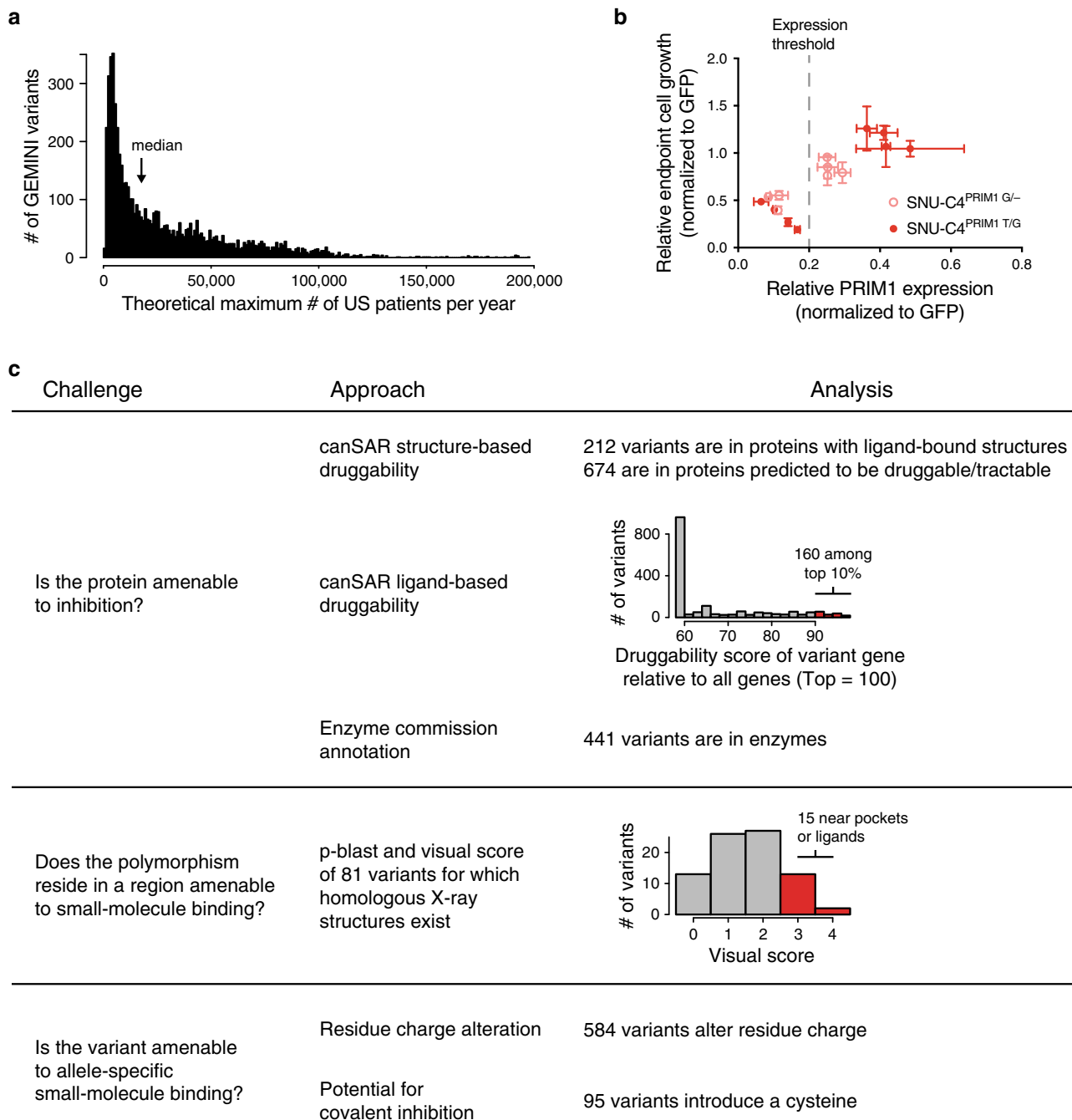


Fig. 4 Potential therapeutic approaches to targeting GEMINI vulnerabilities. **a** Number of GEMINI variants (vertical axis) plotted against the number of patients per year in the US whose tumors might respond to therapeutics targeting those variants (i.e., have lost the resistant allele from a heterozygous germline; horizontal axis). Bin width = 1000 patients. **b** Growth of heterozygous (red circles) and hemizygous cells (pink circles) 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). Data are presented as mean values \pm s.d. Dashed gray 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. Source data for Fig. 4b is provided as a Source Data file.

CRISPR- and RNAi-based approaches explored above, allele-specific small molecules are more tractable for delivery/efficacy. However, no clear allele-specific small molecules exist for any of the 1749 protein-altering GEMINI variants identified in our analysis. Therefore, we focused on in silico analyses to identify and prioritize these GEMINI vulnerabilities (missense, insertion, and deletion variants) for potential allele-specific drug development (Supplementary Data 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^{29,30} (cansar.icr.ac.uk; Methods). This tool uses publicly available structural and chemical information to generate structure- 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 (Fig. 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) (Fig. 4c). Furthermore, 160 GEMINI variants reside in proteins in the top 90th percentile of ligand-based druggability as assessed by the physicochemical properties of small molecules tested against the protein or its homologs (Fig. 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) number³¹ (Fig. 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 Bank³² (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) (Fig. 4c).

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 (Fig. 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, *TGSI*^{rs7823773}, earned a score of 6, including in the visual scoring and cysteine categories. Nine additional variants earned a score of 5 (Supplementary Data 5). These may be among the highest-priority candidates for further exploration.

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 lethality³³. 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^{34,35}. 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 lethality with genetic alterations affecting non-driver genes (also termed “collateral lethality”³⁶) would increase the scope of potential therapeutic approaches. While individual GEMINI genes have been described previously³, 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 lethality. 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 pathways³⁶; and hemizygous loss of essential genes can result in dependency on the remaining copy (CYCLOPS)^{25,26}.

Prior analyses have indicated CYCLOPS genes to be the most frequent class of these synthetic lethal interactions^{26,27}, but we find that GEMINI vulnerabilities provide similar numbers of potential targets. In comparison, fewer paralog dependencies have been described^{27,37–40}. Larger numbers of paralog vulnerabilities have been predicted⁴¹, but it is unclear whether these predictions represent viable candidates³⁶. The 1278 GEMINI genes that we identified also exceed the 299 known driver genes⁴², many of which are proposed therapeutic targets. Expanding the search for GEMINI vulnerabilities beyond pan-essential genes to include variants in lineage-specific essential genes may also increase the number of potential GEMINI 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 genes²⁶ (Supplementary Data 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.

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 division⁴³. Additional alterations affecting nearby nucleic or amino acids could interfere with genetic (e.g., CRISPR and RNAi) or protein-targeting approaches. 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 loss⁴⁴.

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^{43,45–47}, a majority of somatic copy-number alterations, including LOH events, appeared to be clonal, although the fraction of clonal events can be lower in some

loci in some tissues⁴⁸. 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)³⁷.

While we show that cells heterozygous for *PRIMI*^{rs2277339} exhibit no substantial proliferation defects upon ablation of the targeted allele (Fig. 2e, f), systemic knockout of one allele of the 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^{49–52}. Additionally, *S. pyogenes* Cas9 enzymes with altered⁵³ or expanded^{54,55} PAM specificities or CRISPR effectors from other species^{53,56–58} have broadened the total number of targetable loci in the genome and, thereby, the number of targetable variants.

GEMINI vulnerabilities could be also targetable by reversible genetic approaches. Such reversible genetic inhibitors have seen recent success in other disease indications (e.g., the RNAi-based patisiran for treatment of hereditary transthyretin amyloidosis;⁵⁹ the antisense oligonucleotide [ASO]-based IONIS-HTT_{RX} for treatment of Huntington's Disease⁶⁰). Notably, early studies of the GEMINI genes *POLR2A* and *RPA1* achieved allele-specific growth suppression of cancer cells using ASO^{4,61,62} and RNAi⁶³ reagents. Peptide nucleic acids, or PNAs, which can suppress both transcription and translation of target genes, represent another potential allele-specific genetic approach^{64–67}. Finally, the RNA-targeting CRISPR effector Cas13 has shown the ability to knock down target genes⁶⁸ and decrease proliferation of cancer cells⁶⁹ 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 motif⁶⁸, potentially expanding the number of targetable sites beyond those tractable with DNA-targeting CRISPR effectors. The use of genetic targeting approaches would dramatically increase the number of potentially targetable GEMINI vulnerabilities by including silent as well as protein-altering variants. However, unlike the promising therapies for genetic disorders mentioned above, a genetic inhibitor of a GEMINI vulnerability would need to be delivered to all cells in a tumor. Thus, like Cas9-based modalities, the use of 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 inhibitors⁷⁰) as well as those whose genotype-specific effects were identified by pharmacogenomic studies (e.g., warfarin and *VKORC1*⁷¹). However, GEMINI vulnerabilities present an additional challenge for allele-specific inhibitor development because most variants in cell-essential genes do not reside in or near an active site (Fig. 4c) or other functionally critical protein region. This challenge may be addressed through alternative small-molecule approaches, such as proteolysis-targeting chimera (PROTAC)-mediated degradation⁷². SNPs for which one allele is a cysteine could be prioritized for this approach because of the possibility of engineering a covalent inhibitor⁷³.

While we have rigorously validated *PRIMI* and *EXOSC8* as genetic dependencies in cancer, further work is necessary to explore potential therapeutic modalities for targeting them (Supplementary Discussion). The design 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.

Methods

Variant lists. A list of 228,440 potentially targetable variants was downloaded from the Exome Aggregation Consortium (ExAC) database (exac.broadinstitute.org)⁸. 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 in¹². 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 analyses¹². Briefly, SNP array and exome sequencing data from both tumor samples and paired normal DNA were used as inputs to ABSOLUTE⁷⁴, 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_segtab.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 trapping⁷⁵, and pluripotent stem cells with CRISPR-Cas9 gene inactivation⁷⁶). 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”⁷⁷). 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/>)⁷⁸, 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.)

Cell line identification and cell culture. Human cancer cell lines of the appropriate genotypes for *PRIMI* 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>)⁷⁹. 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 mis-identified 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 Zhang⁸⁰. A Cas9 construct co-expressing GFP and two sgRNAs was a gift from Peter Choi²⁶. pLKO.1-TRC cloning vector was a gift from David Root (Addgene plasmid # 10878)⁸¹.

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 polymorphic 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^{80,82}. Briefly, plasmids were cut and dephosphorylated with BsmBI (New England Biolabs) and FastAP (Fermentas) at 37 °C for 2 h. Oligonucleotides for each sgRNA guide sequence (Integrated DNA Technologies) were phosphorylated using T4 polynucleotide kinase (New England Biolabs) at 37 °C for 30 min and then annealed by heating to 95 °C for 5 min and cooling to 25 °C at 1.5 °C/min. Using Quick Ligase (New England Biolabs), annealed oligos were ligated into gel purified vectors (Qiagen) at 25 °C for 5 min. Cloned plasmids were amplified using an endotoxin-free maxi-prep kit (Qiagen).

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 (Corning) at 500, 1000, or 2500 cells per well on the indicated day post-lentiviral infection. Cell number was inferred by ATP-dependent luminescence using CellTiter-Glo reagent (Promega) and normalized to the relative luminescence on the day of plating.

Generation of *PRIM1*-loss and *EXOSC8*-loss cells. A Cas9 construct co-expressing GFP and two sgRNAs with target sites flanking *PRIM1* was used to delete a 20.6 kb region encoding *PRIM1*. Cell lines heterozygous for *PRIM1*^{rs2277339} (SNU-C4 and SNU-175) 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 10 μM ROCK inhibitor Y-27632. Clones were expanded and validated by PCR to harbor the 20.6 kb 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 7.1 kb deletion and *EXOSC8*^{R/R} control lines.

The sgRNA sequences were as follows:

PRIM1 upstream: GCGCGGAAGCTCGCCACGGTA
 PRIM1 downstream: CAGAGCTCCTCAAACCATTTG
 EXOSC8 upstream: GGTTCCTCGGCCGAGCGCCG
 EXOSC8 downstream: TGTACCCATCTACTTAAGTT
 Primers used to verify gene deletion by PCR were as follows:
 PRIM1 deletion genotyping F: ACTGTATGCACCACACAC
 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 previously⁸³. Briefly, 100 pmol of Cas9-NLS was diluted to a final volume of 5 μL with Cas9 buffer (20 mM HEPES [pH 7.5], 150 mM KCl, 1 mM MgCl₂, 10% glycerol, and 1 mM TCEP) and mixed with 5 μL of Cas9 buffer containing 120 pmol of sgRNA. This mixture was incubated for 10 min at RT to allow RNP formation. 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 *MnlI* 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 10 μM 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 (NGS; 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. Non-altered alleles as well as those containing in-frame or frameshift 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 MGH F: AGCTGCAGAGTGTTCCTTTCA
 EXOSC8 MGH R: AGAGCAAAGTAAATGAAAAGCCCCA

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 15,000g 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

See Source Data file for unprocessed scans of the most important blots.

shRNA sequences. pLKO.1 GFP shRNA (target sequence: GCAAGCTGACCC-TGAAGTTCAT) was a gift from David Sabatini (Addgene plasmid # 30323)⁸⁴. 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: CCGAGCTGCTTAACTTTATT
 TRCN0000275194: AGCATCGTCTCTGGGTATATT
 TRCN0000275195: GATTGATATAGCGCAGTATA
 TRCN0000275196: CCGAGCTGCTTAACTTTATT

Allele-specific shRNA sequences were cloned into the vector pLKO.1 as described previously⁸¹. Briefly, the pLKO.1 plasmid was cut with AgeI and EcoRI (New England Biolabs) at 37 °C for 2 h. Oligonucleotides for each shRNA sequence (Integrated DNA Technologies) were annealed by heating to 95 °C for 5 min and cooling to 25 °C at 1.5 °C/min. Using T4 ligase (New England Biolabs), annealed oligos were ligated into gel-purified vectors (Qiagen) at 25 °C for 30 min. Cloned plasmids were amplified using an endotoxin-free maxi-prep kit (Qiagen). The shRNA sequences were as follows:

PRIM1^{rs2277339} major-allele (T) targeting:
 sh3T: TCAATGGAGACGTTTGACC
 sh4T: CAATGGAGACGTTTGACCC
 sh5T: AATGGAGACGTTTGACCCC
 sh6T: ATGGAGACGTTTGACCCCA
 sh7T: TGGAGACGTTTGACCCAC
 sh8T: GGAGACGTTTGACCCACC
 sh9T: GAGACGTTTGACCCACCG
 sh10T: AGACGTTTGACCCACCGA
 sh11T: GACGTTTGACCCACCGAG
 sh16T: TTGACCCACCGAGCTGCC
PRIM1^{rs2277339} minor-allele (G) targeting:
 sh3G: TCAATGGAGACGTTTGCCC
 sh4G: CAATGGAGACGTTTGCCCC
 sh5G: AATGGAGACGTTTGCCCC
 sh6G: ATGGAGACGTTTGCCCCA
 sh7G: TGGAGACGTTTGCCCCAC
 sh8G: GGAGACGTTTGCCCCACC
 sh9G: GAGACGTTTGCCCCACCG
 sh10G: AGACGTTTGCCCCACCGA
 sh11G: GACGTTTGCCCCACCGAG
 sh16G: TTGCCCCACCGAGCTGCC

Quantitative and reverse transcription PCR. RNA was extracted using the RNeasy Mini 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: TTGGATGCCATTACACGACG.

Calculations of theoretical patient numbers. To determine number of patients in the US 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

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

χ = rate of heterozygosity of GEMINI variant

λ = 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

Π = theoretical number of new patients per year in the US.

Estimate of new US pan-cancer cases per year derived from SEER Cancer Statistics Review⁸⁵. Rate of heterozygosity estimated using 2pq from Hardy-Weinberg equation^{86,87}.

DEMETER2 analyses. For a detailed description of the screening and analysis methodology used to generate DEMETER2 scores, please see²⁸. 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 Dependency 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.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The datasets analyzed during the current study are available in the following repositories:

Exome Aggregation Consortium, <http://exac.broadinstitute.org/downloads>

NCI Genomic Data Commons: TCGA copy number and LOH data, <https://gdc.cancer.gov/about-data/publications/pancan-aneuploidy>; CCLE whole exome sequencing data, <https://portal.gdc.cancer.gov/legacy-archive/search/f>

Cancer Cell Line Encyclopedia Portal: <https://portals.broadinstitute.org/ccle>

Cancer Dependency Map Portal: <https://depmap.org/portal/download/>

The source data underlying Fig. 2d–g, 3c–g, 4b and Supplementary Figs 2c–l, 4c–j, and 5a–d are provided as a Source Data file.

Received: 29 January 2019; Accepted: 27 April 2020;

Published online: 20 May 2020

References

- Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
- Martincorena, I. et al. Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029–1041 (2017).
- 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).
- 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).
- 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).
- 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).
- 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).
- Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
- Beroukhi, R. et al. The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
- Zack, T. I. et al. Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
- Holland, A. J. & Cleveland, D. W. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nat. Rev. Mol. Cell Biol.* **10**, 478–487 (2009).
- Taylor, A. M. et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* **33**, 676–689 (2018).
- 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).
- 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).
- Christie, K. A. et al. Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders. *Sci. Rep.* **7**, 1–11 (2017).
- 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).
- Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
- Lucchini, G. et al. polymerase–DNA primase complex: cloning of PRI 1, a single essential gene related to DNA primase activity. *EMBO J.* **6**, 737–742 (1987).
- Francesconi, S. et al. Mutations in conserved yeast DNA primase domains impair DNA replication in vivo. *Proc. Natl Acad. Sci. USA* **88**, 3877–3881 (1991).
- Frick, D. N. & Richardson, C. C. DNA primases. *Annu. Rev. Biochem.* **70**, 39–80 (2001).
- 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).
- Schmid, M. & Jensen, T. H. The exosome: a multipurpose RNA-decay machine. *Trends Biochem. Sci.* **33**, 501–510 (2008).
- 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).
- Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
- Nijhawan, D. et al. Cancer vulnerabilities unveiled by genomic loss. *Cell* **150**, 842–854 (2012).
- Paoletta, 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**, e23268 (2017).
- Tsherniak, A. et al. Defining a cancer dependency map. *Cell* **170**, 564–576 (2017).
- McFarland, J. M. et al. Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. *Nat. Commun.* **9**, 4610 (2018).
- 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).
- Tym, J. E. et al. canSAR: an updated cancer research and drug discovery knowledgebase. *Nucleic Acids Res.* **44**, D938–D943 (2016).
- Webb, E. *Enzyme Nomenclature 1992* (Academic Press, 1992).
- Berman, H. M. et al. The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
- Nijman, S. M. B. Synthetic lethality: general principles, utility and detection using genetic screens in human cells. *FEBS Lett.* **585**, 1–6 (2011).
- Hartwell, L. H., Szankasi, P., Roberts, C. J., Murray, A. W. & Stephen, H. Friend. integrating genetic approaches into the discovery of anticancer drugs. *Science* **278**, 1064–1068 (1997).
- 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).
- Muller, F. L., Aquilanti, E. A. & DePinho, R. A. Collateral lethality: a new therapeutic strategy in oncology. *Trends Cancer* **1**, 161–173 (2015).
- Muller, F. L. et al. Passenger deletions generate therapeutic vulnerabilities in cancer. *Nature* **488**, 337–342 (2012).
- Helming, K. C. et al. ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nat. Med.* **20**, 251–254 (2014).

39. Dey, P. et al. Genomic deletion of malic enzyme 2 confers collateral lethality in pancreatic cancer. *Nature* **542**, 119–123 (2017).
40. 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).
41. Aksoy, B. A. et al. Prediction of individualized therapeutic vulnerabilities in cancer from genomic profiles. *Bioinformatics* **30**, 2051–2059 (2014).
42. Bailey, M. H. et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* **173**, 371–385 (2018).
43. Wang, Y. et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155–160 (2014).
44. Liu, G. et al. Gene essentiality is a quantitative property linked to cellular evolvability. *Cell* **163**, 1388–1399 (2015).
45. 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).
46. Gibson, W. J. et al. The genomic landscape and evolution of endometrial carcinoma progression and abdominopelvic metastasis. *Nat. Genet.* **48**, 848–855 (2016).
47. Jamal-Hanjani, M. et al. Tracking the evolution of non-small-cell. *Lung Cancer N. Engl. J. Med.* **376**, 2109–2121 (2017).
48. McGranahan, N. et al. Allele-specific HLA loss and immune escape in Lung cancer evolution. *Cell* **171**, 1259–1271 (2017).
49. 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).
50. 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. Chem. Int. Ed.* **53**, 2893–2898 (2014).
51. Sun, W. et al. Self-assembled DNA nanoclews for the efficient delivery of CRISPR-Cas9 for genome editing. *Angew. Chem. Int. Ed. Engl.* **54**, 12029–12033 (2015).
52. Wang, H.-X. et al. Nonviral gene editing via CRISPR/Cas9 delivery by membrane-disruptive and endosomolytic helical polypeptide. *Proc. Natl Acad. Sci. USA* **115**, 4903–4908 (2018).
53. Kleinstiver, B. P. et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
54. Hu, J. H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
55. Nishimasu, H. et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **9129**, 1259–1262 (2018).
56. Hou, Z. et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc. Natl Acad. Sci.* **110**, 15644–15649 (2013).
57. Kim, E. et al. In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.* **8**, 1–12 (2017).
58. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-cas system. *Cell* **163**, 759–771 (2015).
59. Adams, D. et al. Patisiran, an RNAi therapeutic, for hereditary transthyretin amyloidosis. *N. Engl. J. Med.* **379**, 11–21 (2018).
60. Tabrizi, S. J. et al. Targeting huntingtin expression in patients with Huntington's disease. *N. Engl. J. Med.* **380**, 2307–2316 (2019).
61. 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).
62. Fluiter, K. et al. In vivo tumor growth inhibition and biodistribution studies of locked nucleic acid (LNA) antisense oligonucleotides. *Nucleic Acids Res.* **31**, 953–962 (2003).
63. 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).
64. Hanvey, J. C. et al. Antisense and antigene properties of peptide nucleic acids. *Science* **258**, 1481–1485 (1992).
65. 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).
66. 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).
67. Egholm, M. et al. PNA hybridizes to complementary oligonucleotides obeying the Watson–Crick hydrogen-bonding rules. *Nature* **365**, 566–568 (1993).
68. Abudayyeh, O. O. et al. RNA targeting with CRISPR–Cas13. *Nature* **550**, 280–284 (2017).
69. 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).
70. 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).
71. Rost, S. et al. Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature* **427**, 537–541 (2004).
72. Toure, M. & Crews, C. M. Small-molecule PROTACS: New approaches to protein degradation. *Angew. Chem. Int. Ed. Engl.* **55**, 1966–1973 (2016).
73. Lonsdale, R. & Ward, R. A. Structure-based design of targeted covalent inhibitors. *Chem. Soc. Rev.* **47**, 3816–3830 (2018).
74. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
75. Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
76. 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**, 610–619 (2018).
77. 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).
78. Sondka, Z. et al. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **18**, 696–705 (2018).
79. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
80. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
81. 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).
82. Shalem, O. et al. Genome-scale CRISPR–Cas9 knockout. *Science* **343**, 84–88 (2014).
83. 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).
84. Sancak, Y. et al. The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1. *Science* **320**, 1496–1501 (2008).
85. Noone, A. et al. *SEER Cancer Statistics Review 1975–2015* (National Cancer Institute, 2017).
86. Hardy, G. H. Mendelian proportions in a mixed population. *Science* **28**, 49–50 (1908).
87. Weinberg, W. Über den Nachweis der Vererbung beim Menschen. *Jahresh. des. Ver. für. Vater.ändische Naturkd. Württemberg.* **64**, 368–382 (1908).
88. 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).
89. Liu, Q., Greimann, J. C. & Lima, C. D. Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* **127**, 1223–1237 (2006).

Acknowledgements

The authors would like to thank Benjamin Ebert, Matthew Meyerson, David Weinstock, Florian Muller, David Barbie, Alexander Gimelbrandt, John Doench, John Alberta, Pratiti (Mimi) Bandopadhyay, John Daley, and Jeremiah Wala for helpful discussions and suggestions. Peter Choi kindly provided a Cas9 construct co-expressing GFP and two sgRNAs used in isogenic cell line generation.

Author contributions

C.A.N. and W.J.G.: designed and performed experiments, analyzed data, and prepared the manuscript; M.S.B., J.P.B. and H.W.: designed and performed experiments and analyzed data; L.M.U. and N.C.: designed and performed experiments; J.A.K. and A.D.C.: analyzed data and provided technical support and conceptual advice; A.B. and G.G.: analyzed data; B.R.P.: designed and performed experiments, analyzed data, prepared the manuscript, and provided technical support and conceptual advice; S.D.P. and R.B.: analyzed data, prepared the manuscript, and provided technical support and conceptual advice.

Competing interests

W.J.G. received consulting fees from nference and ImmPact Bio and holds equity in nference and Ampressa Therapeutics. The remaining authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-16399-y>.

Correspondence and requests for materials should be addressed to B.R.P. or R.B.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020