A systematic genome-wide mapping of oncogenic mutation selection during CRISPR-Cas9 genome editing

Sanju Sinha1,2,3,9, Karina Barbosa4,9, Kuoyuan Cheng1,3,9, Mark D. M. Leiserson3, Prashant Jain4, Anagha Deshpande4, David M. Wilson III5, Bríd M. Ryan2, Ji Luo6, Ze’ev A. Ronai4, Joo Sang Lee7,8,10, Aniruddha J. Deshpande4,10✉ & Eytan Ruppin1,10✉

Recent studies have reported that genome editing by CRISPR-Cas9 induces a DNA damage response mediated by p53 in primary cells hampering their growth. This could lead to a selection of cells with pre-existing p53 mutations. In this study, employing an integrated computational and experimental framework, we systematically investigated the possibility of selection of additional cancer driver mutations during CRISPR-Cas9 gene editing. We first confirm the previous findings of the selection for pre-existing p53 mutations by CRISPR-Cas9. We next demonstrate that similar to p53, wildtype KRAS may also hamper the growth of Cas9-edited cells, potentially conferring a selective advantage to pre-existing KRAS-mutant cells. These selective effects are widespread, extending across cell-types and methods of CRISPR-Cas9 delivery and the strength of selection depends on the sgRNA sequence and the gene being edited. The selection for pre-existing p53 or KRAS mutations may confound CRISPR-Cas9 screens in cancer cells and more importantly, calls for monitoring patients undergoing CRISPR-Cas9-based editing for clinical therapeutics for pre-existing p53 and KRAS mutations.
Clustered regularly interspaced short palindromic repeats (CRISPR) enables targeted gene disruption and editing, a powerful technology that expands our understanding of fundamental biological processes. Beyond its impact on biological research, CRISPR-based approaches have been considered for various applications in medicine, from reparative editing of primary cells to the development of new strategies to treat a variety of genetic diseases, including cancer. However, several clinical trials based on CRISPR technology have been deferred due to significant potential risks, including off-target effects, generation of unexpected chromosomal alterations, and potential immunogenicity. Other studies have demonstrated that double-stranded breaks (DSBs) induced during CRISPR-Cas9-based gene knockout (CRISPR-KO) can lead to DNA damage response, whose level can either be associated with the copy number of the targeted gene or in some cases structural rearrangements in the region.

Recent studies have shown that the DNA damage response following CRISPR-KO can be mediated by p53, a tumor-suppressor gene mutated in over 50% of all human cancers. Genome-wide CRISPR screening in immortalized human retinal pigment epithelial (RPE1) cells revealed that a p53-mediated DNA damage response, followed by cell cycle arrest, is induced upon generation of DSBs by the Cas9 endonuclease, favoring the survival of cells that have inactivated the p53 pathway. Most recently, a study showed that exogenous expression of Cas9 can also activate this p53-mediated DNA damage response. While these studies indicate that CRISPR-Cas9 genome editing techniques may select for p53-mutated cells, several outstanding questions remain unaddressed: First, since most of these p53 studies have involved only a small number of primary or transformed cells, it is unclear whether p53 selection can happen broadly across multiple different cell types including transformed cancer cells. Second, it is not clear whether stronger p53 selection can happen when certain genes or parts of the genome are targeted, or the level of selection is more homogenous regardless of the genes being edited. And finally, it remains to be investigated whether this selection is limited to p53 only or that other cancer driver genes can also be selected for during CRISPR-Cas9 genome editing.

To address these questions, here we employ a computational framework coupled with experimental validations to conduct a comprehensive evaluation of each cancer driver mutation-selection associated with CRISPR-Cas9. We first demonstrate that CRISPR-KO-induced mutant p53 selection can be observed in transformed and non-transformed cells of diverse lineages via both lentivirus and ribonucleoprotein (RNP)-based Cas9 delivery. More importantly, we systematically characterized mutation-selection in other cancer driver genes during CRISPR-Cas9 identifying that KRAS mutants can also be selected for, as demonstrated in large-scale genetic screens and Cas9-expressing cell lines. We further identified the underlying pathways that are likely to mediate this selection.

**Results**

**CRISPR-Cas9 gene-knockouts selects for p53 mutations in a vast variety of transformed and non-transformed cell types.** We first sought to address two important gaps in our understanding of CRISPR-KO-driven mutant p53 selection—firstly, we wanted to investigate whether this selection generalizes across cell types. Secondly, we wanted to understand what type of sgRNAs, genes, and gene networks drive this selection. We analyzed the DepMap16 genome-wide gene essentiality data across 248 cancer cell lines (Supplementary Data 1), where both CRISPR-Cas9 (AVANA16) and shRNA-based (Achilles16) genetic screens were conducted. We searched for genes whose CRISPR-Cas9-based knockout (CRISPR-KO) reduced cell viability more (i.e., more essential) in p53-wildtype (WT; N = 75) than p53-mutant (N = 173) cell lines, but do not exhibit such differential essentiality in the shRNA-based screens (see the “Methods” section). The KO of such genes may select for p53 mutants specifically during CRISPR-Cas9 editing. In the CRISPR-Cas9 screen, we find many more genes (981) that are more essential in p53-WT vs. p53-mutant cell lines, compared to the genes that are more essential in p53-mutant cells (237 genes). In contrast, the numbers of such differentially essential genes in the shRNA screens were balanced (~1500 each). Such significantly different patterns between CRISPR-Cas9 and shRNA screens (Fig. 1a left panel, Chi-squared test P < 1.4E–284) points to a bias that knockout/knockdown of a gene is more likely to impair the fitness of p53-WT cells specifically with CRISPR-Cas9 but not with shRNA.

Among the 981 genes that are more essential in p53-WT cells with CRISPR-KO, 861 genes (87%) do not exhibit this differential essentiality in shRNA screens. We hence termed these CRISPR-specific differentially essential positive (CDE+) genes (Fig. 1a right panel; genes listed in Supplementary Data 2). We find that these CDE+ genes are preferentially located in chromosomal bands containing common fragile sites (CFSs; hypergeometric P < 2.3E–4, Fig. 1b, Supplementary Data 3), which are prone to replicative stress, fork collapse and DNA breaks that cause genomic instability (GI). As CRISPR-KO could induce kilobase-scale structural alterations near the targeted site, this finding suggests that CRISPR-targeting near CFSs may enhance DNA damage, promote the p53-dependent cell death response and provide a selective advantage to p53-mutant cells. The sgRNAs of the CDE+ genes also tend to be highly accessible chromatin (HAC) (hypergeometric P < 0.02; see the “Methods” section), thus inducing a strong damage response. The top pathways enriched within CDE+ genes include DNA damage response, DNA repair, and Fanconi anemia (FA; hypergeometric test adjusted P < 0.01, Supplementary Data 2B). This is consistent with the recent report that the FA pathway is involved in repairing Cas9-induced DNA double-strand breaks (DSBs) and that their KO may further enhance DNA damage.

Analogous to CDE+, we defined CDE− genes, which are more essential in p53-mutant (vs. WT) cells with CRISPR-KO, but do not show such difference in shRNA screens (185 genes, right panel of Fig. 1a). CDE− genes are involved in cellular processes that engage p53, including mitotic checkpoints, DNA replication and cell cycle (Supplementary Data 2B, Fig. 1d, hypergeometric test adjusted P < 0.1), with the top hit being the key cyclin cycle regulator CDKN1A (aka. p21, Wilcoxon rank-sum P < 1.85E–08, Fig. 1c). Transiently inhibiting CDE− genes during CRISPR-KO may mitigate p53 mutation selection and could be of interest from a translational point of view. Top CDE+/− genes are highlighted in Fig. 1c.

We next tested for a series of confounding factors that could potentially lead to this skewness (expanded details in Supplementary Notes 1, 2, 6). We repeated our analysis in the following modes to control for: 1. Gene copy number variations in cell lines, by correcting for gene copy number via a linear regression while testing whether a gene is differentially essential in WT vs. mutant cell lines (Supplementary Fig. 1), 2. Functional effect of different p53 variants, by focusing on cell lines harboring known loss-of-function (LOF) p53 mutations solely (see the “Methods” section, N = 78) vs. WT (Supplementary Note 1a), 3. Effect of partial vs. complete silencing of gene expression in case shRNA-KD and CRISPR-Cas9 KO, respectively, by using genes that are not expressed at all (Supplementary Note 1b), 4. Effect of a potential functional relationship, e.g., synthetic lethal or rescue interaction with p53, by using non-essential genes (Supplementary Note 1c), 5. Effect of different reagents and sgRNA depletion time, by using...
independent CRISPR-Cas9 screen (Sanger Screen) performed using 326 cancer cell lines (Supplementary Note 2, Supplementary Fig. 2), 6. Off target effect of sgRNA, by computing and taking into account an off-target score of the sgRNA used in the screens (Supplementary Note 6). We showed that our findings via the CDE identification process remain valid after controlling for this broad array of potentially confounding factors.

We next performed our own CRISPR-Cas9 essentiality screen, employing CRISPRi-based essentiality screens as a control in a pair of p53-isogenic MOLM13 leukemia cell lines (WT and p53R248Q mutant). We used a deep (10 guides per gene) and focused sgRNA library targeting top p53 CDE+ and CDE− genes (see the “Methods” section; Fig. 2a, details in Supplementary Note 10, Supplementary Data 4). Here, we observed in the CRISPR-Cas9 screen that the CDE+ genes are more essential in p53-WT vs. mutant cells, and vice versa for the CDE− genes (Wilcoxon signed-rank test P < 0.08 and P = 0.03 for CDE+/− genes respectively). Reassuringly, we do not see such differential essentiality in the CRISPRi screens (Wilcoxon signed-rank P = 0.32 and 0.29; Top 10% CDE+− genes are depicted in Fig. 2b).

To further assess whether such selection effects can be observed in non-transformed cells, we next tested and observed that indeed our p53 CDE+ genes have higher essentiality in WT vs. isogenic-mutant cells in published CRISPR-Cas9 but not shRNA genome-wide screens performed in non-transformed RPE1 cells.
A competition assay shows selection for p53 mutant over wildtype cells following CRISPR-Cas9 knockout of CDE+ genes. To test whether the CRISPR-KO of CDE+ genes leads to the selection of p53-mutant cells in a competitive setting,22–24, we silenced the top five predicted CDE+ genes using CRISPR-Cas9 and CRISPRi in the p53-isogenic MOLM13 cells (see the "Methods" section). Following a lentiviral sgRNA transduction, the WT and mutant cells were mixed at an initial ratio of 95:5, and monitored by flow cytometry for up to 25 days (illustrated in Fig. 3a; Supplementary Data 5A). Silencing 2/5 CDE+ genes (NDUFB6 and NDUFB10) induced a strong progressive p53 mutant enrichment of up to five folds over WT at day 25 specifically in CRISPR-KO, across several independent sgRNAs and not for NTC (Fig. 3b, blue lines). No inverse enrichment in p53 WT cells was observed in the competitive assays involving the three other CDE+ genes (Supplementary Fig. 6). We observed that sgRNAs targeting NDUFB6 induced significantly higher DNA damage compared to NTC-treated cells specifically in p53 WT cells (Supplementary Fig. 7, despite higher editing efficiency in the mutant cells as shown in Fig. 4c), demonstrating that the DNA damage was not just due to Cas9 expression. This may partly explain their selective competitive advantage upon the CDE+ gene KO. Testing the robustness of this competitive selection advantage for p53-mutant cells, we repeated the CRISPR-KO competitive assay for a larger number of 18 top CDE+ genes with up to four unique sgRNAs per gene and monitored the assay up to 15 days (Supplementary Data 5B). Using the non-targeting sgRNA as a baseline, we observed the competitive outgrowth of p53-mutant cells for 15 out of 28 sgRNAs and 10 out of 18 CDE+ genes tested (Fig. 3c).

p53 mutation-selection phenomena extend to transient knockout and primary cells. We next asked whether our top CDE+ genes may also select for p53 mutants under CRISPR-Cas9 transient knockout. We delivered Cas9 and the sgRNA as an RNP. We observed that upon Cas9-RNP mediated transfection of a sgRNA targeting our top CDE+ gene from our pooled and competition assays, NDUF6 (see the "Methods" section), there was a higher loss of edited cells in the p53 WT vs isogenic p53-mutant MOLM13 cells over 10 days of culture, as measured by the change in ICE scores (Fig. 4a top panel). Using an orthogonal method of proliferation monitoring by dye-dilution25, we observed that there was a progressive slowing down in cell proliferation of p53 WT, but not p53 mutant MOLM13 cells upon Cas9-RNP based KO of NDUF6 vs. respective non-targeting controls (NTCs) (Fig. 4b, c). Similar to the lentiviral system, this is likely due to the DNA damage induced by the NDUF6 sgRNA compared to the Cas9 only or Cas9 with NTC controls in p53 WT cells (Figs. 4b and S8). We repeated this transient knockout in non-transformed cells (RPE1) and consistently observed an increased loss of edited p53 WT over p53-mutant cells (Fig. 4a, bottom panel). Notably, we also observed a selection of p53 mutant over WT in patient tumors (TCGA) based on the copy number alteration patterns of CDE+ genes (details in Supplementary Note 11A).

KRAS mutant cell lines exhibit a selection advantage in large-scale genetic screens. To determine whether additional cancer
Driver mutations may be selected for the following CRISPR-KO, we focused on a list of 61 cancer driver genes from Vogelstein et al.\textsuperscript{26} that are mutated in at least 10 of the cell lines screened in the AVANA\textsuperscript{10} and Achilles\textsuperscript{16} datasets. For each of these cancer genes, we identified the differentially essential genes between its WT and mutant cell lines in the CRISPR-Cas9 (AVANA) and shRNA (Achilles) screens, as described above for p53. We ranked the cancer genes by the significance of skewness in the numbers of differentially essential genes from CRISPR-Cas9 vs. shRNA screens similar to that shown in Fig. 1a for p53 (with Fisher’s exact tests, see the “Methods” section; results shown in Fig. 5a and Supplementary Data 6). The mutants of these genes may be selected for during CRISPR-KO, as their WT cells are overall more vulnerable during CRISPR-KO compared to the mutants. We term these genes “(potential) CRISPR-selected cancer drivers” (CCDs). The top significant CCD in addition to p53 is the oncogene KRAS. Like for p53, potential confounding factors including copy number were controlled for (Supplementary Note 1, Supplementary Fig. 1b), and there is no significant correlation between the mutation profiles of KRAS and p53 (Fisher’s test \( P = 0.67 \)), suggesting that KRAS might be a CCD independent of p53. We thus next focused on investigating the selection of mutant KRAS as another major CCD.

KRAS is a major oncogene whose gain of function mutation is known to activate various DNA repair pathways and may override the trigger of cell death upon DNA damage\textsuperscript{27,28}, supporting its role as a CCD. We computationally identified the CDE\textsuperscript{+} and CDE\textsuperscript{−} genes of KRAS in a similar way described above for p53 (Fig. 5b, Supplementary Data 2A). KRAS has high numbers of CDE\textsuperscript{+}/− genes, while only very few KRAS mutation-associated genes are identified in the shRNA screen. The predicted median mutant selection levels are comparable to those of p53 (Supplementary Note 4), i.e. the CRISPR-KO of its CDE\textsuperscript{+} genes is likely to drive comparable levels of mutant

---

**Fig. 3 Selection for p53 mutant cells under CRISPR-Cas9 knockout of CDE\textsuperscript{+} genes in a co-culture of p53 WT/mutant cells.**

- **a** An illustration showing the experimental design of the competition assay where isogenic p53 WT/mutant MOLM13 cell lines were mixed with a ratio of 95:5 and top p53 CDE\textsuperscript{+} genes were knocked out by CRISPR-Cas9. The population ratio was monitored for 25 days at a 5-day interval starting from the day of sgRNA transduction. Change in the percentage of p53 mutant cells in the p53-mutant WT cells (y-axis) co-culture with time (x-axis, number of days in co-culture), under the CRISPR-KO or CRISPRi of individual selected top p53 CDE\textsuperscript{+} genes or with non-targeting control sgRNA. The p-values are calculated using two-sided Wilcoxon Rank Sum tests. Data are presented as mean values ± SEM across \( N = 2 \) independent biological replicates (unique sgRNAs).

- **b** The difference of the percentage of p53-mutant cells between Day 15 and Day 0 in co-culture (y-axis), under the CRISPR-KO of a larger set of top p53 CDE\textsuperscript{+} genes (x-axis, by individual sgRNAs, specified by number suffixes after gene symbols). Similar to panel (a), data are presented as mean values ± SEM across \( N = 3 \) independent biological replicates (unique sgRNAs). The horizontal dashed line represents the value for non-targeting control sgRNA (NTC).
Fig. 4 Transient knockout in both MOLM13 and RPE1 leads to preferential loss of p53 WT over mutant cells. 

(a) The editing efficiency of NDUFB6 around the sgRNA cut site was determined in p53 mutant and wildtype isogenic pair of MOLM13 (top panels, transformed cells) and RPE1 cells (bottom panels, primary cells) using ICE protocol (see the “Methods” section) at day 0 (orange) and day 10 (green) after Cas9-RNP-sgRNA nucleofection. Differences in day 0 compared to day 10 editing efficiency can be used as a measure of relative fitness of edited compared to non-edited cells. The p-values are calculated using two-sided Wilcoxon Rank Sum tests. Data are presented as mean values ± SEM across N = 4 independent biological replicates (unique sgRNAs).

(b) DNA damage is quantified from gamma H2AX (γ-H2AX) staining images and measured by γH2AX staining in p53 wildtype MOLM13 cells with no Cas9, Cas9 + sgRNA for a non-targeting control (NTC) or NDUFB6. γ-H2AX foci (y-axis) in all three conditions (x-axis) are enumerated in the violin plot.

(c) Mean fluorescence intensity of the CellTrace™ dye (APC) in MOLM13 p53 mutant (top panel) vs. wildtype cells (bottom panel) is shown for NTC (light blue) or NDUFB6 targeting sgRNAs (dark blue). CellTrace™ APC fluorescence is inversely correlated with proliferation. The error bars denote standard error (mean ± standard deviation) across three replicates. The p-values are calculated using a two-sided t-test given a small number of data points (n = 3). 

(d) Proliferative effects of NDUFB6 editing in RNP-transfected p53 mutants compared to wild-type cells. A histogram of MOLM13 p53 wildtype (top panel) cells transfected with an NTC or an NDUFB6 sgRNA is shown with the fluorescence intensity of the CellTrace™ dye (APC) on the x-axis. Similarly, MOLM13 p53 mutant cells are plotted in the bottom panel. The error bars are presented and computed as panel (a).
selection as the KO of the CDE+ genes of p53. Fourteen genes are CDE+ genes of both p53 and KRAS, and thus their CRISPR-KO may impose considerable selection for both KRAS and p53 mutants (Supplementary Data 2A). Consistent with the knowledge of downstream pathways regulated by activated KRAS, its CDE− genes are significantly enriched for DNA DSB repair pathways (FDR < 0.02, see the “Methods” section, visualized in Fig. 5c, Supplementary Data 2E).

Based on the number of CDE+ genes (Fig. 5a), the third-ranked CCD is VHL, having a large number of CDE+ genes. Like p53 and KRAS, we controlled for a series of potential confounding factors (Supplementary Notes 1, 2, 6). We also note that the CDE+ genes of VHL were also enriched in chromosomal bands of CFSs (hypergeometric \( P < 2.4e^{-2} \)). The known function of VHL as a positive regulator of p53 in DNA damage-induced cell cycle arrest or apoptosis possibly accounts for its role as a CCD. However, in this study, we chose to carefully study one additional CCD and our top-ranked hit—KRAS and its potential mutant selection during CRISPR-Cas9 gene editing and thus will focus on that from here onwards.

**Pooled essentiality screens show that KRAS CDE+ genes are more essential in WT than mutants and vice versa for CDE− genes specifically in CRISPR-Cas9 screens.** Similar to p53, we next performed our own CRISPR-Cas9 and a control CRISPRi gene essentiality screens, but on a smaller scale, in a pair of isogenic KRAS WT and KRAS G12D mutant MOLM13 cell lines using a sgRNA library targeting top KRAS CDE+ and CDE− genes (details in Supplementary Note 10, Supplementary Data 7). Here, we observed that the KRAS CDE+ genes are more essential in WT than mutants and vice versa for CDE− genes specifically in CRISPR-Cas9 screens (Wilcoxon signed-rank test \( P < 0.074 \) and \( P < 0.042 \) for CDE+/−, respectively, Fig. 6a left panel), but not in CRISPRi screens (Wilcoxon signed-rank \( P < 0.22 \) and \( P < 0.49 \); Fig. 6a right panel). Similar results were obtained from analyzing published genome-wide CRISPR-Cas9 and shRNA genetic screens performed in a different pair of KRAS isogenic cell lines (WT and G12D mutation in DLD1 cell line; Fig. 6b). Similar to p53, we also observed a selection of KRAS mutants in patient tumor profiles (TCGA) based on the copy number alteration patterns of its CDE+ genes (details in Supplementary Note 11B).

**A competition assay shows selection for KRAS mutant over wildtype cells following CRISPR-Cas9 knockout of KRAS CDE+ genes.** To test whether, like the p53 case, CRISPR-KO of KRAS CDE+ genes can confer a selective advantage to KRAS mutant over WT cells in co-culture, we conducted a similar competition assay using a pair of WT and KRAS G12D mutant isogenic MOLM13 cell lines. As in the experiment for p53, we mixed the WT and KRAS mutant cells at an initial ratio of 95:5 following KRAS CDE+ sgRNA transduction and monitored the population for 15 days to track the percentage of KRAS-mutant cells (TdTomato+) with flow cytometry (see the “Methods” section;
Supplementary Data 8). A total of 10 KRAS CDE+ genes were tested, in addition to NTC. In the control group, the KRAS mutant cell fraction decreased with time, indicating that the mutant cells have lower baseline fitness levels than the WT cells.

In comparison, in 8 out of 10 CDE+ genes tested, there is a gain in the fitness of the mutant cells (Fig. 6c; see the “Methods” section), testifying that even though the KRAS-mutant cells have a lower baseline fitness level, the CRISPR-KO of the majority of
CDE+ genes can enhance their fitness and in a subset of cases lead to selective outgrowth of KRAS mutant over WT cells in a mixed population.

Cas9 expression in cancer cell lines selects for KRAS mutations. Multiple studies have reported a higher editing efficiency of Cas9 in p53 mutated versus p53 WT cell lines. We first asked if this may also extend to KRAS as an equally important CDD. Analyzing induced exogenous Cas9 activity in 1601 cancer cell lines from DepMap (1375 and 226 KRAS WT and mutant, respectively), we found that, like p53, Cas9 activity is significantly higher in KRAS-mutant cells than in KRAS WT cells (P = 2.9E-05, Wilcoxon Rank-Sum test, Fig. 6d; see the “Methods” section). We repeated the above analysis modeling Cas9 activity vs. KRAS status adjusting for p53 status in a linear model, yielding concordant findings (P = 2.43E-04, Wilcoxon Rank-Sum test). Importantly, across all the 61 cancer driver genes we analyzed above from Vogelstein et al., KRAS and p53 are the only ones showing such a significant difference in Cas9 activity between WT and mutant cells after FDR correction (FDR = 9.1E-04 for KRAS and 7.1E-06 for p53, Wilcoxon Rank-Sum test, Fig. 6e). This further shows that in addition to p53, KRAS WT status can also hamper the efficiency of CRISPR-Cas9.

Based on the above findings and a recent report of selection for p53 mutant due to DNA damage upon Cas9 expression (without sgRNA), we asked whether DNA damage induced by Cas9 alone can also lead to a mutation-selection of KRAS and/or other cancer drivers. To this end, we re-analyzed deep sequencing profiles from 42 Cas9-expressed vs. matched parental (i.e. without Cas9) cell lines (see the “Methods” section) from Enache et al., and identified a total of 9 cases involving 5 unique KRAS mutations, occurring in 7 different cell lines with moderate to high Cas9 activity (Methods). Seven out of these 9 cases show increased mutant allele frequency after induced Cas9 expression (Wilcoxon signed-rank test P = 0.027, Fig. 6f). Four of the 5 KRAS mutations are missense mutations; the other mutation is an intronic mutation occurring 100 bp from the splicing site. This mutation is not present in the parental cell lines but emerges independently after Cas9 expression in four different cell lines.

While the results suggest that CRISPR-Cas9 may select for KRAS mutations, the functional role of these mutations needs to be interpreted with caution. Among the 61 cancer driver genes from Vogelstein et al., KRAS is a top gene (ranked the second) along with p53 (ranked fourth) that shows significant mutant sub-clonal expansion (Fig. 6g). Notably, the top genes identified to be involved in mutant sub-clonal expansions have a significant overlap with our previously identified top CDD genes (Fisher’s exact test P = 0.04). We note that when we repeated the above analysis excluding the intronic deletion, KRAS is ranked 16th out of 61 genes tested (Wilcoxon rank-sum P = 0.28).

KRAS mutant cells downregulate the G2M checkpoint pathway in response to Cas9 induction. To investigate the mechanism underlying the potential selective advantage of KRAS mutant vs. WT cells during CRISPR-KO, we analyzed gene expression data of 163 pairs of parental (without Cas9) and the corresponding Cas9-expressed cell lines (138 KRAS WT, 25 KRAS mutant) and identified the pathways that are differentially regulated upon Cas9 expression between KRAS WT and mutant cells (i.e. up/down-regulated in KRAS mutant but inversely or non-significantly regulated in KRAS WT cells; Supplementary Fig. 9a, Supplementary Note 5 and Supplementary Fig. 10 for p53). A major differentially regulated pathway was the G2M checkpoint with the highest difference in the normalized enrichment score (Supplementary Fig. 9b), which is strongly down-regulated (rank 2/50) in KRAS mutants but strongly upregulated in KRAS WT cells (4/50). This is a canonical pathway that serves to prevent the cells with genomic DNA damage from entering mitosis (M-phase) and thus its downregulation in KRAS mutant cells may provide them with a proliferative advantage. Another top pathway, E2F Targets, which primarily regulates G1/S transition and DNA replication was also found to be downregulated in KRAS mutant cells but upregulated in KRAS WT cells upon Cas9 expression (Supplementary Fig. 9b). Thus, Cas9-induction may similarly underlie the selective advantage of KRAS mutant cells by selectively activating cell cycle checkpoint pathways in response to DNA damage.
Discussion
In this study, we systematically investigated the possibility of selection of pre-existing cancer driver mutations during CRISPR-Cas9 gene editing. First, we confirmed and extended upon previous findings that selection of CRISPR-Cas9 can happen, showing it in a large set of transformed and non-transformed cell lines. We identified the specific CDE+ genes whose CRISPR-KO is likely to mediate such selection, and further tested and validated some of these predictions in new screens and competitive assays that we have performed. After studying and validating our integrated computational and experimental pipeline in the known case of p53, we turned to applying it to study a collection of major cancer driver genes and discovered that KRAS is another major cancer driver gene whose pre-existing mutants have a selective advantage during CRISPR-Cas9 gene editing. We demonstrated the selective advantage of KRAS mutant cells performing a CRISPR-KO/CRISPRi screen in isogenic cells with pooled CDE+ gene-targeting sgRNAs, and further in competition assays during the CRISPR-KO of top predicted KRAS CDE+ genes. We also observed that KRAS WT cells have lower Cas9 activity and thus a lower editing efficiency, similar to that observed for p53, which may limit CRISPR-mediated gene-editing in such cells. Analyzing recently published KRAS screens, we also find a subclonal expansion of KRAS mutant cancer cells following Cas9 expression. Finally, our study also shows that the introduction of the Cas9 protein down-regulates the G2M checkpoint and E2F targets in KRAS mutant, but not KRAS WT cells, which may confer a selective advantage to KRAS-mutant cells.

Multiple factors can contribute to the identity of CDE+ genes, including involvement in DNA repair and cell cycle pathways, being located in chromosomal fragile sites or HAC regions, supporting that their CRISPR-KO can lead to augmented DNA damage. We find that these factors can together account for up to 15% of our CDE+ genes. We also observed that a gene-targeted by highly off-target guides can also lead to high DNA damage, which reassuringly only accounts for up to 10% of the CDE+ genes (details in Supplementary Note 6). Taken together, these putative mechanisms can explain about 25% of the CDE+ genes we have identified, however, the mechanisms underlying the rest are yet open to further studies.

Overall, our results point to a need for accounting for CDE effects in the analysis of dependencies in CRISPR screens. More importantly, our studies point to the need for careful selection of sgRNAs for therapeutic genome editing and recommend cautionary monitoring of KRAS status in addition to that of p53 during therapies utilizing CRISPR-Cas9. Lastly, in the publicly available CRISPR-Cas9 screens that we have analyzed, the current small numbers of cell lines with mutations in other cancer drivers, such as VHL, limits our ability to reliably determine whether these cancer genes could also be selected during CRISPR-Cas9 genome editing. The investigation of the latter thus awaits specifically designed screens in designated isogenic cell lines.

Methods
CRISPR and shRNA essentiality screen data. We obtained CRISPR-Cas9 essentiality screen (or dependency profile) data in 436 cell lines from Meyers et al.19 for 16,368 genes, whose expression, CNV, and mutation data are available via the CCLE portal24. We obtained the shRNA essentiality screen data in 501 cell lines from the DepMap portal19 for 16,165 genes, whose expression, CNV, and mutation data are available publicly via the CCLE portal24. The 248 cell lines and 14,718 genes that appear in both datasets were used in this analysis (Supplementary Data 1). For mutation data, only non-synonymous mutations were considered. Synonymous (silent) mutations were removed from the pre-processed MAF files downloaded from the CCLE portal24.

Identifying CRISPR specific differentially essential genes of a potential CRISPR-selected cancer driver. For a given CDE (e.g. p53 or KRAS), we checked how many CDE genes are significantly associated with the mutational status of the CDD using a Wilcoxon rank-sum test in the CRISPR and shRNA datasets, respectively (FDR < 0.1). CRISPR-specific differentially essential positive (CDE+) genes are those whose CRISPR-KO is significantly more viable when the CDD is mutant while their shRNA silencing is not, whereas the CDE– genes are those whose CRISPR-KO is significantly more viable when the CDD is WT while their shRNA silencing is not. We filtered out any candidate CDE genes whose copy number was also significantly associated with the given mutation to control for potentially spurious associations coming from copy number (we removed genes showing significant association (FDR < 0.1)—the exact procedure used is described below in the section titled “Identifying potential CRISPR-selected cancer drivers”).

Identifying CDEs considering the functional impact of mutations. Out of a total of 248 cell lines that we analyzed, 173 cell lines (69.7%) have p53 non-synonymous mutations. In addition to identifying CDEs by considering all non-synonymous mutations, we additionally employed a more conservative approach where we aimed to consider only p53 LOF mutations in the CDE identification process. To this end, we considered a mutation to be LOF if it was classified as non-sense, indel, frame-shift, or among the 4 most frequent non-functional hotspot mutations (R248Q, R273H, R248W and R175H within the DNA-binding domain, determined as pathogenic by COSMIC36). Using this definition we obtained new mutation values for all the CDE genes and identified Enriched CDE genes via the same method described in the section titled “Identifying CRISPR specific differentially essential genes of a potential CRISPR-selected cancer driver”. We repeated a similar process with the top three known gain-of-function hotspot mutation variants of KRAS.

Identifying potential CRISPR-selected cancer drivers of CRISPR-KO. To identify additional CDD genes like p53, we considered 121 cancer driver genes identified by Vogelstein et al.18, whose nonsynonymous mutation is observed in at least 10 cell lines (N ≥ 10). We determined whether each of these genes is a CDD as follows: for each of the 61 candidate genes, we tested the association between the mutational status of each gene in the genome (reflected by post-KO cell viability) with the mutational status of the candidate CDD gene using a Wilcoxon rank-sum test. We then counted the number of genes, whose essentiality is: (i) significantly positively associated with the candidate CDD mutation status (FDR-corrected P-value < 0.1, median essentiality of WT > mutant of the cancer gene), (ii) significantly negatively associated with the candidate CDD mutation status (FDR-corrected P-value < 0.1, median essentiality of WT < mutant of the cancer gene), and (iii) not associated (FDR-corrected P-value > 0.1) with the candidate CDD mutation status; we performed this computation separately for the CRISPR and the shRNA screens, respectively. This computation results in a 3-by-2 contingency table for each candidate CDD gene. We then checked whether the distribution of the above three counts in the CRISPR dataset significantly deviates from that in the shRNA dataset via a Fisher’s exact test on the contingency table. If each of the values in the contingency table was >30, we used the exact version of Fisher’s exact test. We further filtered out any candidate CDE genes whose copy number was also significantly associated with the given mutation to control for potentially spurious associations coming from copy number (we removed genes showing significant association (FDR < 0.1)). We performed this procedure for all 61 candidate genes, one by one, and selected those with FDR corrected t-test < 0.1. We further filtered out the candidate CDD whose mutation profile is correlated with p53 mutation profile via a pairwise Fisher test of independence (FDR < 0.1). We finally report the CDD genes that have a substantial number of CDE+ genes (N > 300).

Pathway enrichment analysis of CDE+–CDE– genes. We analyzed the CDE+–CDE– genes of each of the CDDs for their pathway enrichment with annotations from the Reactome database27 in two different ways. First, we tested for significant overlap between our CDE genes with the pathways with hypergeometric tests (FDR < 0.1). Second, we ranked all the genes in the CRISPR-KO screen by the differences in their median post-KO cell viability values in mutant vs. WT cells, and the standard GSEA method18 was employed to test whether the genes of each Reactome pathway have significantly higher act over ranks vs. the rest of the genes (FDR < 0.1). We repeated the GSEA analysis with the genes ranked by differential post-KO cell viability in the shRNA screen and only reported significant pathways specific to CRISPR but not shRNA screens. We confirmed that for p53, the GSEA method was able to recover the top significant pathways identified by the hypergeometric test (e.g., those in Fig. 1d), although extra significant pathways were identified (Supplementary Data 2). For p53 and KRAS CDE– genes respectively, the enriched pathways were clustered based on the Jaccard index and the number of overlapping genes with Enrichment Map39, and the largest clusters were visualized as network diagrams with Cytoscape40. To study the potential enrichment of CDE genes in CDDs, we obtained chromosomal band locations of CFS22 and defined the CFS gene set as the set of all
genes located within these chromosomal bands (obtained from BioMart)44. We tested for significant overlap between our CDE genes and the CFS gene set with a hypergeometric test confirmed the lack of significant overlap with the corresponding shRNA-DE genesets. Similarly, for the common HAC regions, we obtained a list of these regions defined by a consensus of DNAasel and FAIRE across seven different cancer cell lines from a previous study44. Next, we identified sgRNAs that are expected to target such HAC regions (see the "Calculating off-target scores" section) and ranked genes based on the number of targeting such sgRNAs. Taking the top genes equal to the number of p53 CDE+ genes, we computed the enrichment for p53 CDE+ genes via a hypergeometric test.

Testing the clinical relevance of copy number alterations of the p53 or KRAS CDE genes. We tested the hypothesis that copy number alterations in CDE+ genes (as a potential surrogate for the number of DSBs in these genes) can reduce the fitness of the CCD (p53 or KRAS) WT tumor in the patient data. This was done by testing whether the GI of CDE+ genes is significantly lower than that of control non-CDE genes in CCD-WT but not in CCD-mutant tumors. Further, we tested if higher absolute levels of SCNA of the CDE+ genes are associated with increased rate of CCD (p53 or KRAS) mutation accumulation with cancer stage, as this would further suggest that such amplification/deletion events in the CDE+ genes can drive the selection for CCD mutants. To this end, the following logistic regression model was used to identify the genes whose high absolute SCNA levels were associated as above with a higher rate of CCD mutation accumulation with cancer stage, while controlling for cancer type and overall mutation load:

\[
\logit(P(\text{CCD})) = \beta_0 + \sum_{i} \beta_{\text{cancer type}_i} \text{cancer type}_i + \beta_{\text{mutation load}} \text{mutation load} + \beta_{\text{stage}} \text{stage} + \beta_{\text{interaction}} \text{interaction},
\]

where $\beta_i$ is the absolute log ratio of SCNA of gene $i$ in a sample relative to normal control, and $\beta$ is the indicator function. Wilcoxon rank-sum test was then used to test whether the GI of CDE+ genets is significantly lower than that of control non-CDE genes in CCD-WT but not in CCD-mutant tumors. Further, we tested if higher absolute levels of SCNA of the CDE+ genes are associated with increased rate of CCD (p53 or KRAS) mutation accumulation with cancer stage, as this would further suggest that such amplification/deletion events in the CDE+ genes can drive the selection for CCD mutants. To this end, the following logistic regression model was used to identify the genes whose high absolute SCNA levels were associated as above with a higher rate of CCD mutation accumulation with cancer stage, while controlling for cancer type and overall mutation load:

\[
\logit(P(\text{CCD})) = \beta_0 + \sum_{i} \beta_{\text{cancer type}_i} \text{cancer type}_i + \beta_{\text{mutation load}} \text{mutation load} + \beta_{\text{stage}} \text{stage} + \beta_{\text{interaction}} \text{interaction},
\]

\[
GL = \frac{1}{N} \sum_{i} (1),
\]

corresponding shRNA-DE genesets. Similarly, for the common HAC regions, we obtained gold-standard essential and non-essential geneset from Hart et al.46. To test the quality of each genetic screen we computed an area under the precision-recall curve (AUPRC) using the average logFC across replicates and cell lines. In this study, we only considered the genetic screens with an AUROC $> 0.6$ (random model AUPRC $= 0.5$). We also employed this method to test the quality of our in-house generated genetic screens.

Crispr-Cas9 RNP transfection experiments. We generated sgRNAs by in vitro transcription using the HiScribe™ T7 Quick High Yield RNA Synthesis Kit (New England Biolabs, Beverly, MA) and performed the RNP complex formation using TrueCut Cas9 Protein v2 (ThermoFisher Scientific, Waltham, MA) according to published protocols. We transfected with Cas9 RNP complexed with NTC or NDUFB6 RNA, according to the manufacturer’s instructions. Cells were maintained in culture for 48 h before harvest for imaging, dye-dilution, and editing estimation assays. For NDUFB6 and NDUFS6 editing estimation assays, we used the Syngro Performance Analysis IEC tool according to the instructions, using un-transfected parental MOLM13 samples as controls and samples from 48 h post-transfection as the Day 0 initial timepoint and Day 10 as a final time point, in triplicates. For RPE1 experiments, mutant cells with p53R248Q or p53-NDUFS6 were transfected using Lipofectamine 2000 (Invitrogen) and tested for the presence of the p53R248Q mutation using the p53R248Q MaxQ Cas9 Transfection Reagent (ThermoFisher Scientific) according to the manufacturer’s instructions for 12-well plate format, in triplicates. Cell harvesting time-points were similar to those of MOLM-13.

Dye-dilution experiments. We used the CellTrace™ Violet Cell Proliferation Kit (ThermoFisher Scientific) to stain MOLM13-WT and MOLM13-p53 mutant cells transfected with Cas9 RNP complexed with NTC or NDUFB6 RNA, according to the manufacturer’s instructions. Cells were maintained in culture in the dark and assayed by flow cytometry using the LSR Fortessa every 2 days for 14 days. FCS files were analyzed using FlowJo software.

Analysis of γ-H2AX foci in MOLM13-Cas9 and MOLM13-p53 mutant cells. MOLM13-WT and MOLM13-p53 mutant cells were left untreated or treated with 1 μg/mL MPM-224 for 2 h at 37°C and then fixed with paraformaldehyde. DNA was stained with 4',6-diamidino-2-phenylindole (DAPI) and γ-H2AX was stained with the monoclonal antibody (clone DB-1, Novus Biologicals) diluted 1:1000. Cells were imaged using a V10.0 microscope (Zeiss Oberkochen) with 40× objective. Images were analyzed using the ImageJ Fiji plugin, and the number of γ-H2AX foci per nucleus was counted in 100 cells for each condition.
complexed with NTC or NDUFB6, and negative control cells were pelleted at 400g for 5 min at 4 °C. washed two times in PBS, and fixed in 4% paraformaldehyde in PBS for overnight at 4 °C. Cells were washed two times in PBS and permeabilized in 0.25% triton X-100 in PBS for 5 min at room temperature. Following two washes with PBS, the cells were incubated in blocking buffer (3% BSA in PBS) for 30 min at room temperature and subsequently incubated with APC conjugated H2AX phospho (Serine 139) antibody (BioLegend; Cat #613415) at an antibody dilution of 1:200 in blocking buffer for overnight at 4 °C in dark. Cells were washed two times with PBS and resuspended in 150 μl PBS. Cell suspensions were spotted on poly-lysine-coated glass slides using cytospin (Cytospin 4; Thermo Scientific) centrifugation at 72× g for 4 min. Coverslips were mounted onto the slides using ProLong Gold antifade reagent with DAPI (Invitrogen) and cured overnight at room temperature in dark. Slides were imaged in Nikon A1R HD confocal microscope. Sequential z-sections were imaged using a x60 oil objective and maximum projection images were obtained using the Nikon NIS-Elements platform.

**Cas9 activity in cancer cell lines with KRAS (or another cancer driver) WT vs. mutant.** We downloaded the exogenous Cas9 activity of 1601 cancer cell lines from the DepMap portal and their KRAS mutation status considering only non-synonymous variants profiled using whole-exome sequencing (1375 and 226 KRAS WT and mutant, respectively)16. We tested whether the Cas9 activity is higher in KRAS WT vs. KRAS WT cell lines using a one-sided Wilcoxon rank-sum test. We repeated this process for each cancer driver gene and used the FDR corrected significance to rank them in addition to the fold change of Cas9 expression.

**Subclonal expansion of KRAS mutant in parent vs. high Cas9-expressed cell lines.** We downloaded the deep targeted sequencing of cancer driver genes performed on 42 parental and matched Cas9-expressed cancer cell lines from Enache et al.15. In this analysis, we discarded the cell lines with <20% Cas9 activity and thus low DNA damage. We asked whether mutant allele frequency of a cancer driver (e.g. KRAS) significantly increased in Cas9-expressed cell lines compared to matched parental cell lines using Wilcoxon signed-rank test. In this analysis, we have considered both intronic and exonic variants provided by sequencing.

**Analysis of differentially expressed pathways in KRAS WT and mutant cells in response to Cas9 induction.** Gene expression profiles of 163 pairs of parents (without Cas9) and the corresponding Cas9-expressed cell lines (138 KRAS WT, 25 KRAS mutant) were obtained from Enache et al.15. Differential expression analysis between the Cas9-expressed cells and the parental cells was performed for the KRAS WT and mutant cells separately, and GSEA analysis (genes ranked by logFC) was performed to identify the hallmark pathways from MSigDB (https://www.gsea-msigdb.org/gsea/msigdb/)48. Cas9 activity in 1601 cell lines from DepMap and deep sequencing profiles of these Cas9-expressed vs. matched parental cells were downloaded from Enache et al.15 (Supplementary Data) and the raw data can be found at BioProject accession number PRJNA545458. Source data are provided with this paper.

**Code availability**

We have provided the scripts to reproduce each step (numbered) of results and main text and supplementary figures in a GitHub repository which can be accessed here: https://doi.org/10.5281/zenodo.5478587. Source data are provided with this paper.

Received: 2 February 2021; Accepted: 23 September 2021; Published online: 11 November 2021
33. Knauf, J. A. et al. Oncogenic RAS induces accelerated transition through G2/M and promotes defects in the G2 DNA damage and mitotic spindle checkpoints. *Journal of Biological Chemistry*, 281, 3800–3809 (2006).
34. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607 (2012).
35. McFarland, J. M. et al. Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. *Nucleic Acids Res.* 43, D805–D811 (2015).
36. Forbes, S. A. et al. COSMIC: exploring the world’s knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* 37, D805–D811 (2015).
37. Matthews, L. et al. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* 37, D619–D622 (2009).
38. Korotkevich, G. et al. Fast gene set enrichment analysis. Preprint at bioRxiv https://doi.org/10.1101/060012 (2021).
39. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* 5, e13984 (2010).
40. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504 (2003).
41. Smedley, D. et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* 43, W589–W598 (2015).
42. Song, L. et al. Open chromatin defined by DNasel and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* 21, 1757–1767 (2011).
43. Bilal, E. et al. Improving breast cancer survival analysis through competition-based multidimensional modeling. *PLoS Comput. Biol.* 9, e1003047 (2013).
44. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* 42, e168 (2014).
45. Spahn, P. N. et al. PinAPL-Py: a comprehensive web-application for the analysis of CRISPR/Cas9 screens. *Sci. Rep.*, 15854 (2017).
46. Hart, T. et al. Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol. Syst. Biol.* 10, 733 (2014).
47. Brunetti, L., Gundry, M. C., Kitano, A., Nakada, D. & Goodell, M. A. Highly efficient gene disruption of murine and human hematopoietic progenitor cells by CRISPR/Cas9. *J. Vis. Exp.* 134, 57278 (2018).
48. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 6, 417–425 (2015).
49. Sinha, S. & Cheng, K. *Zenodo* https://doi.org/10.5281/zenodo.5478587 (2021). ruppinlab/crispr_risk: CRISPR-riskv1.0.0 (Published_v1.0.0).

Acknowledgements

We acknowledge and thank the National Cancer Institute for providing financial and infrastructural support. We thank Curtis Harris, Andre Nussenzweig, Sridhar Hannenhalli and the members of Cancer Data Science Lab for insightful feedback. This research was supported in part by the Intramural Research Program of the National Institutes of Health, NCI. S.S. and K.C. are supported by the NCI-UMD Partnership for Integrative Cancer Research Program. A.J.D. would like to acknowledge the support of the National Cancer Institute of the National Institutes of Health under Award Number P30 CA030199, the Rally Foundation for Childhood Cancer Research, and Luke Tatsu Johnson Foundation under Award Number 19FIN45, an Emerging Scientist Award from the Children’s Cancer Research Fund, and the V Foundation for Cancer Research (TVF) under Award Number DVP2019-015.

Author contributions

E.R., A.J.D., and J.S.L. supervised the study. S.S., K.C., K.B., E.R., A.J.D., J.S.L. conceived and designed the study. S.S., K.B., and K.C. developed the methodology. K.B., S.S., K.C., P.J., A.D., A.J.D. designed and performed experiments. S.S., K.B., K.C., M.D.L., P.J., A.D., D.M.W., B.M.R., J.L., Z.A.R. analyzed and interpreted the data. S.S., K.B., K.C., M.D.L., P.J., A.D., D.M.W., B.M.R., J.L., Z.A.R., J.S.L., A.J.D., E.R. wrote, reviewed and revised the manuscript.

Funding

Open Access funding provided by the National Institutes of Health (NIH).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-26788-6.

Correspondence and requests for materials should be addressed to Aniruddha J. Deshpande or Eytan Ruppin.

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

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 processes.

© The Author(s) 2021, corrected publication 2022