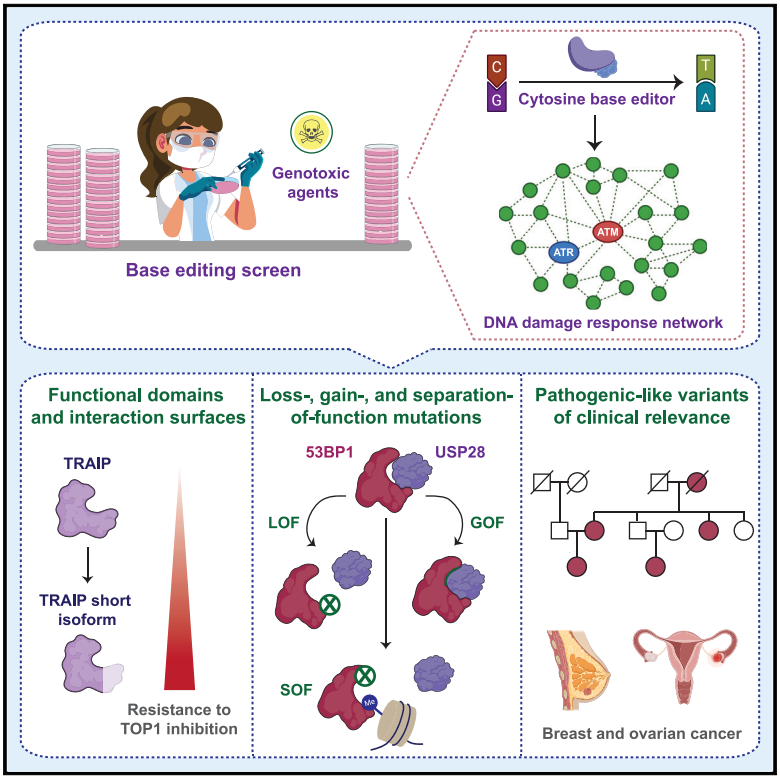
Cell

Functional interrogation of DNA damage response variants with base editing screens

Graphical Abstract



Highlights

- Base editing screens characterize mutations in 86 DNA damage response genes
- Mutations in 53BP1 Tudor domain modulate USP28 binding without altering DNA repair
- Mutation cluster analyses define a TRAIP domain required for TOP1 inhibitor toxicity
- Cancer variants of pathogenic-like behavior are uncovered by base editing screens

Authors

Raquel Cuella-Martin, Samuel B. Hayward, Xiao Fan, ..., Chao Lu, Yufeng Shen, Alberto Ciccia

Correspondence

ac3685@cumc.columbia.edu

In Brief

Base editing screens enable large-scale functional characterization of nucleotide variants in genes of the DNA damage response, allowing study of mutation-specific phenotypes and providing a higher degree of genotype-phenotype resolution relative to standard gene knockout screens.

Cuella-Martin et al., 2021, Cell *184*, 1081–1097
February 18, 2021
https://doi.org/10.1016/j.cell.2021.01.041

CellPress

Cell

CellPress

Resource

Functional interrogation of DNA damage response variants with base editing screens

Raquel Cuella-Martin,[1] Samuel B. Hayward,[1,5] Xiao Fan,[2,3,5] Xiao Chen,[1] Jen-Wei Huang,[1] Angelo Taglialatela,[1] Giuseppe Leuzzi,[1] Junfei Zhao,[2,3,4] Raul Rabadan,[2,3,4] Chao Lu,[1] Yufeng Shen,[2,3] and Alberto Ciccia[1,6,*]

[1]Department of Genetics and Development, Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, NY 10032, USA

[2]Department of Systems Biology, Columbia University Irving Medical Center, New York, NY 10032, USA

[3]Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY 10032, USA

[4]Program for Mathematical Genomics, Columbia University Irving Medical Center, New York, NY 10032, USA

[5]These authors contributed equally

[6]Lead contact

*Correspondence: ac3685@cumc.columbia.edu

https://doi.org/10.1016/j.cell.2021.01.041

SUMMARY

Mutations in DNA damage response (DDR) genes endanger genome integrity and predispose to cancer and genetic disorders. Here, using CRISPR-dependent cytosine base editing screens, we identify > 2,000 sgRNAs that generate nucleotide variants in 86 DDR genes, resulting in altered cellular fitness upon DNA damage. Among those variants, we discover loss- and gain-of-function mutants in the Tudor domain of the DDR regulator 53BP1 that define a non-canonical surface required for binding the deubiquitinase USP28. Moreover, we characterize variants of the TRAIP ubiquitin ligase that define a domain, whose loss renders cells resistant to topoisomerase I inhibition. Finally, we identify mutations in the ATM kinase with opposing genome stability phenotypes and loss-of-function mutations in the CHK2 kinase previously categorized as variants of uncertain significance for breast cancer. We anticipate that this resource will enable the discovery of additional DDR gene functions and expedite studies of DDR variants in human disease.

INTRODUCTION

Maintenance of genome stability is promoted by a network of cellular pathways known as the DNA damage response (DDR). DDR proteins often exert numerous distinct functions, with some factors regulating both DNA repair and DNA damage checkpoints (e.g., ATM, 53BP1) and others controlling multiple DNA repair processes (e.g., BRCA1) (Ciccia and Elledge, 2010). Since many DDR proteins are also essential for cell viability and/or possess large multidomain architectures, dissecting their functions often requires analysis of point mutations. As current methods to study mutations rely on conventional structure-function approaches, only a few separation-of-function (SOF) mutations in DDR genes have been defined to date (Billing et al., 2018; Cuella-Martin et al., 2016; Feng and Jasin, 2017; Ghezraoui et al., 2018; Schlacher et al., 2011).

Genetic alterations of DDR genes have been implicated in numerous human genetic disorders and cancer predisposition syndromes (Ciccia and Elledge, 2010; Jackson and Bartek, 2009; Khanna and Jackson, 2001; McKinnon, 2009). According to the *ClinVar* database, >37,000 single nucleotide variants (SNVs) in DDR genes are listed as variants of uncertain significance (VUSs). Standard methods for classifying DDR SNVs rely on the study of wild-type and mutant open reading frames (ORFs). However, these studies have been limited to a small number of genes (e.g., *BRCA1*, *BRCA2*, *PALB2*, and *BARD1*) and entail analyses of ectopically expressed proteins (Adamovich et al., 2019; Boonen et al., 2019; Bouwman et al., 2013; Farrugia et al., 2008; Guidugli et al., 2014; Lee et al., 2015; Wiltshire et al., 2020). Alternatively, SNVs can be introduced at endogenous loci using CRISPR-dependent homology-directed repair (HDR) (Hsu et al., 2014). This strategy was recently used to examine the clinical significance of ~4,000 *BRCA1* SNVs (Findlay et al., 2018). However, despite its high accuracy, CRISPR-dependent HDR displays limited efficiency and requires the formation of DNA double-strand breaks (DSBs), which can elicit genomic rearrangements and p53-mediated cell cycle arrest or cell death (Haapaniemi et al., 2018; Ihry et al., 2018; Kosicki et al., 2018).

Recently, DSB-independent CRISPR-based methods have been developed to generate nucleotide variants by direct modification of DNA bases. In particular, base editors resulting from the fusion of nickase Cas9 with either the cytosine deaminase APOBEC1 (e.g., BE3) or the adenine deaminase TadA (e.g., ABE) have been shown to introduce, respectively, C→T or A→G transitions within a defined nucleotide window (Gaudelli et al., 2017; Komor et al., 2016). Consequently, BE3 and ABE can generate 64 distinct amino acid (AA) substitutions with

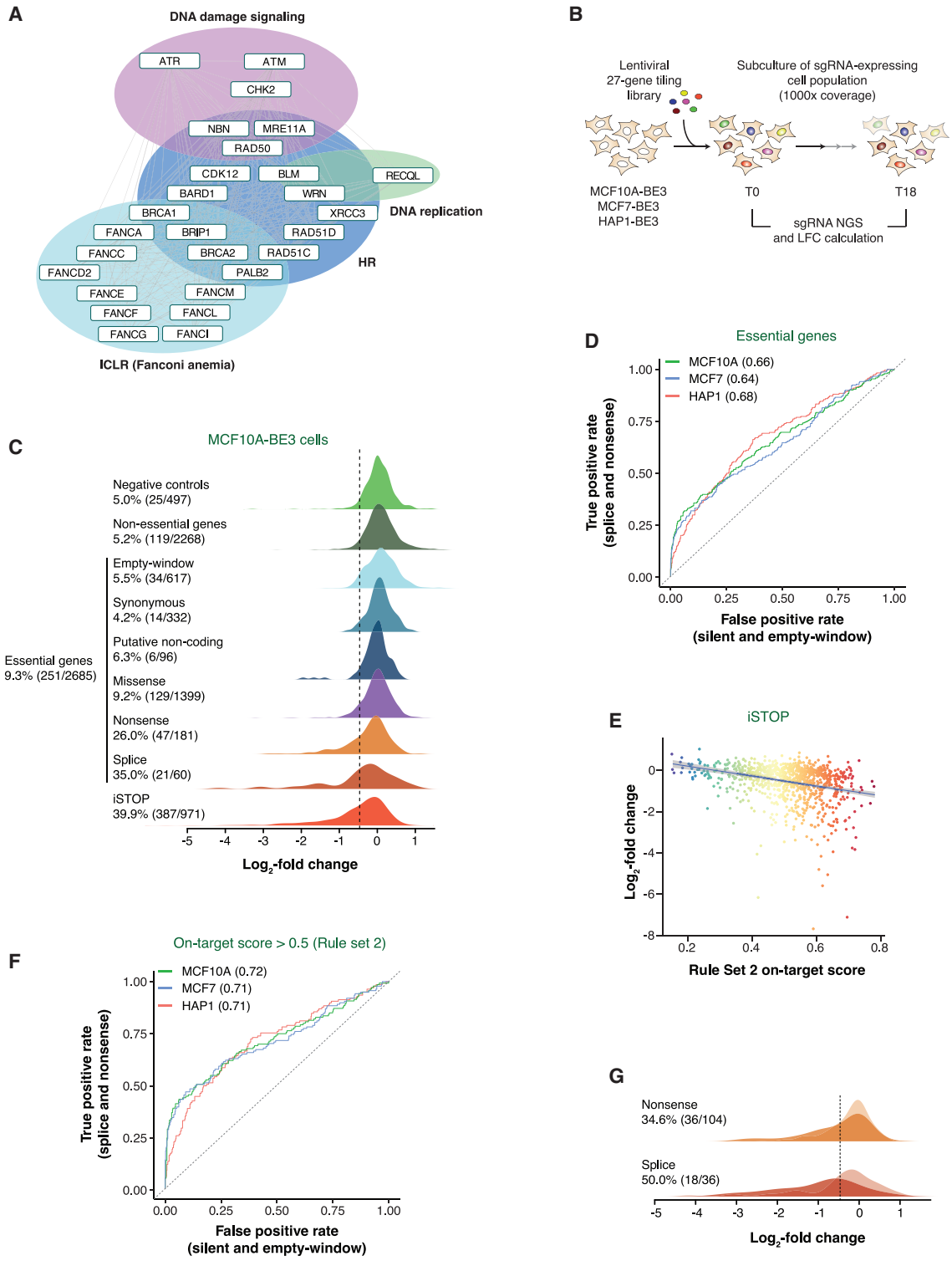


Figure 1. Performance of CRISPR-dependent base editing screens targeting DDR genes

(A) 27-gene DDR network targeted by the sgRNA sublibrary 1, organized by pathways, and their genetic and physical interactions (gray lines), according to the STRING database.

(B) Schematic of the protocol utilized for base editing screens in untreated conditions. MCF10A, MCF7, and HAP1 cells expressing BE3 were transduced with a lentiviral sgRNA library targeting the gene network in (A), and then cultured for 18 days after selection. The LFC in sgRNA abundance between day 0 (T0) and day 18 (T18) was then determined following next-generation sequencing (NGS).

(legend continued on next page)

higher efficiency than HDR-based approaches (Yeh et al., 2018). Base editing has been applied to generate loss-of-function mutations or revert pathogenic variants (Billon et al., 2017; Gapinske et al., 2018; Gaudelli et al., 2017; Koblan et al., 2021; Komor et al., 2016; Kuscu et al., 2017). Furthermore, base editing screening approaches have been employed to generate mutations in yeast and evaluate *BRCA1*, *MAP2K1*, *KRAS*, and *NRAS* variants in mammalian cells (Després et al., 2020; Jun et al., 2020; Kweon et al., 2020).

Here, we investigate the potential of CRISPR-dependent base editing to define the function of human nucleotide variants in a high-throughput manner, both for structure-function analyses of functionally complex DDR genes, and for identification of DDR VUSs with pathogenic potential. In particular, we employ BE3-dependent base editing to induce mutational tiling of 86 human DDR genes and analyze the effects of these mutations in response to DNA damaging agents. Through these studies we identify loss-, gain-, and separation-of-function (LOF, GOF, and SOF) mutations in 53BP1, mutations in the TRAIP ubiquitin ligase with drug-specific phenotypes, and VUSs of pathogenic-like behavior in the CHK2 kinase. Together, this work highlights the versatility of base editing screens for large-scale phenotypic characterization of human nucleotide variants.

RESULTS

CRISPR-dependent base editing allows large-scale separation of mutational outcomes in DDR genes

To evaluate the performance of base editing screens, we designed a lentiviral single guide RNA (sgRNA) library targeting every NGG PAM sequence within the coding region of 27 DDR genes primarily involved in homologous recombination (HR) and/or inter-strand crosslink repair (ICLR) (Figure 1A; Table S1). Additionally, sgRNAs designed to generate stop codons in common essential genes were included as positive controls for cell lethality (hereafter, iSTOP controls) (Hart et al., 2015), while non-targeting sgRNAs, *AAVS1*-targeting sgRNAs, and sgRNAs targeting essential loci without cytosines within the editing window (hereafter, empty-window) were used as negative controls. This sgRNA library (~11,000 sgRNAs) was transduced into MCF10A, MCF7, and HAP1 cells stably expressing BE3 (hereafter, MCF10A-BE3, MCF7-BE3, and HAP1-BE3 cells) (Zafra et al., 2018) (Figures 1B, S1A, and S1B). The transduced cells were then cultured for 18 days, and $\log_2$-fold change (LFC), p value, and gene rank for each sgRNA were computed at day 18 (T18) relative to day 0 (T0) (Tables S2–S4). We then determined all possible DNA editing outcomes for each sgRNA within the defined 6-nucleotide BE3 window (Zafra et al., 2018) and grouped sgRNAs into five unique categories, ranked from the least to the most damaging mutational outcome (i.e., empty-window, silent, missense, nonsense, and splice), and a sixth category for sgRNAs targeting non-productive transcripts (i.e., putative non-coding). Next, we selected sgRNAs targeting the top 7 essential genes within our library (Table S1) and examined the distribution of their LFCs in each mutational category. As expected, iSTOP controls were depleted across all cell lines tested, while empty-window and synonymous categories presented LFC distributions comparable to the negative controls (Figures 1C and S1C). The distribution of missense mutations reflected variable mutational outcomes, whereas nonsense and splice categories presented the highest percentages of sgRNAs below the established threshold (bottom 5% of negative controls), reflecting the ability of nonsense and splice mutations to induce gene knockout (Figures 1C and S1C).

Screen performance was also evaluated by computing receiving operator characteristic (ROC) curves that compared true positives, namely sgRNAs predicted to introduce splice and nonsense mutations (hereafter, deleterious mutations) in the top 7 essential genes, with false positives, as defined by silent and empty-window sgRNAs (hereafter, neutral mutations) targeting the same gene set (Figure 1D). Area under the curve (AUC) values reflected consistent sgRNA behavior across all three human cell lines evaluated, establishing the ability of base editing screens to resolve mutational outcomes regardless of cell transformation status and/or ploidy (Figure 1D). Likewise, ROC analyses revealed good performance of the designed positive and negative controls (Figures S1D and S1E). Unlike standard CRISPR-KO screens, which require DSB formation, base editing screens should not activate p53-mediated checkpoint responses upon sgRNA targeting. Accordingly, ROC analyses showed that non-targeting, *AAVS1*-targeting, and empty-window negative controls behaved indistinguishably (Figure S1F).

We additionally observed that the design of positive controls for base editing screens can be optimized by applying the Rule Set 2 on-target efficiency score developed for CRISPR-KO sgRNAs (Doench et al., 2016). Indeed, iSTOP sgRNAs with higher on-target efficiency values were more likely to yield deleterious phenotypes (Figure 1E; Tables S3 and S4). Remarkably, application of a minimum Rule Set 2 score of 0.5 resulted in increased resolution of deleterious versus neutral mutations and led to a shift in the LFC distributions of nonsense and splice sgRNAs for the top 7 essential genes in MCF10A-BE3 cells (Figures 1C, 1D, 1F, 1G, and S1G). As in the case of iSTOP controls, LFC values for splice sgRNAs in the top 7 essential genes negatively correlated with sgRNA on-target efficiency (Figure S1H). Targeting splice donor or acceptor sites yielded similar results, regardless of their relative position in the gene (Figure S1I).

(C) Density plots of LFC values for sgRNAs predicted to generate the indicated mutational outcomes in essential and non-essential genes in MCF10A-BE3 cells (Table S1). LFC density plots are also shown for iSTOP and negative control sgRNAs. Dotted line, LFC for the bottom 5% of negative controls (LFC = −0.46581).
(D) ROC analyses of MAGeCK ranks for sgRNAs predicted to introduce splice and nonsense mutations (true positives) versus silent and empty-window sgRNAs (false positives) in essential genes. AUC values are indicated in brackets.
(E) Graphical representation of LFC values and Rule Set 2 on-target scores for iSTOP controls in MCF10A-BE3 cells.
(F) ROC analyses conducted as in (D) exclusively for sgRNAs with Rule Set 2 on-target scores > 0.5. AUC values are indicated in brackets.
(G) Density plots of LFC values for sgRNAs introducing nonsense and splice variants in essential genes with Rule Set 2 on-target score > 0.5 in MCF10A-BE3 cells. Density plots from (C) are depicted in lighter colors. Dotted line, LFC = −0.46581.
See also Figure S1 and Tables S1–S4.

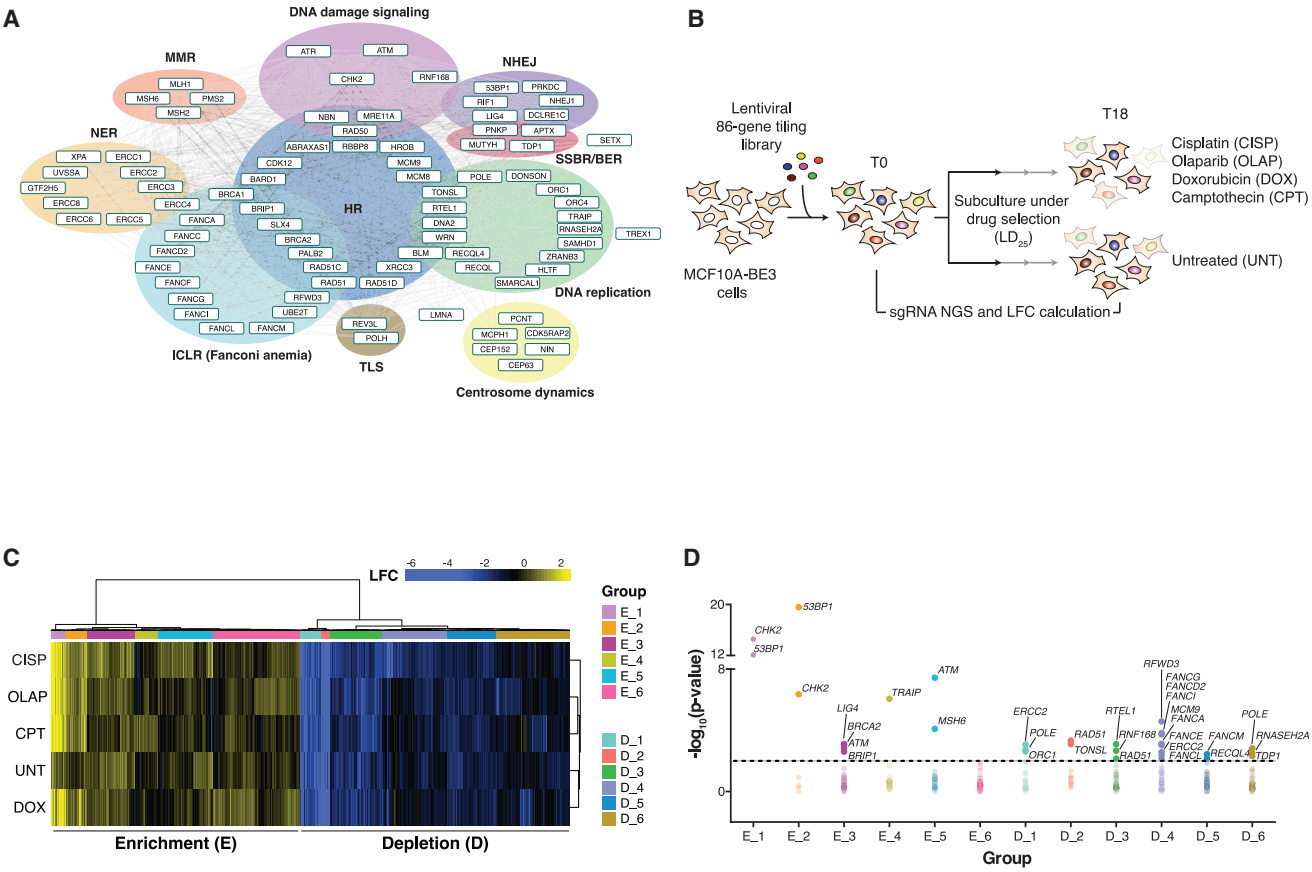


Figure 2. Analyses of drug-sgRNA interactions upon treatment with DNA damaging agents

(A) 86-gene DDR network targeted by sgRNA sublibraries 1 and 2, represented as in Figure 1A.

(B) Schematic of the protocol utilized for base editing screens including treatment with genotoxic agents. MCF10A-BE3 cells were transduced with the lentiviral sgRNA sublibraries 1 and 2 targeting the gene network in (A) and cultured in the presence of the indicated genotoxic agents or left untreated. LFC values in sgRNA abundance between T0 and T18 were then calculated for each individual condition following NGS.

(C) Heatmap of LFC values for biologically and statistically relevant sgRNAs (LFC beyond top or bottom 1% of negative controls and $p < 0.01$, Table S6) hierarchically clustered by Ward's method. E, enrichment; D, depletion.

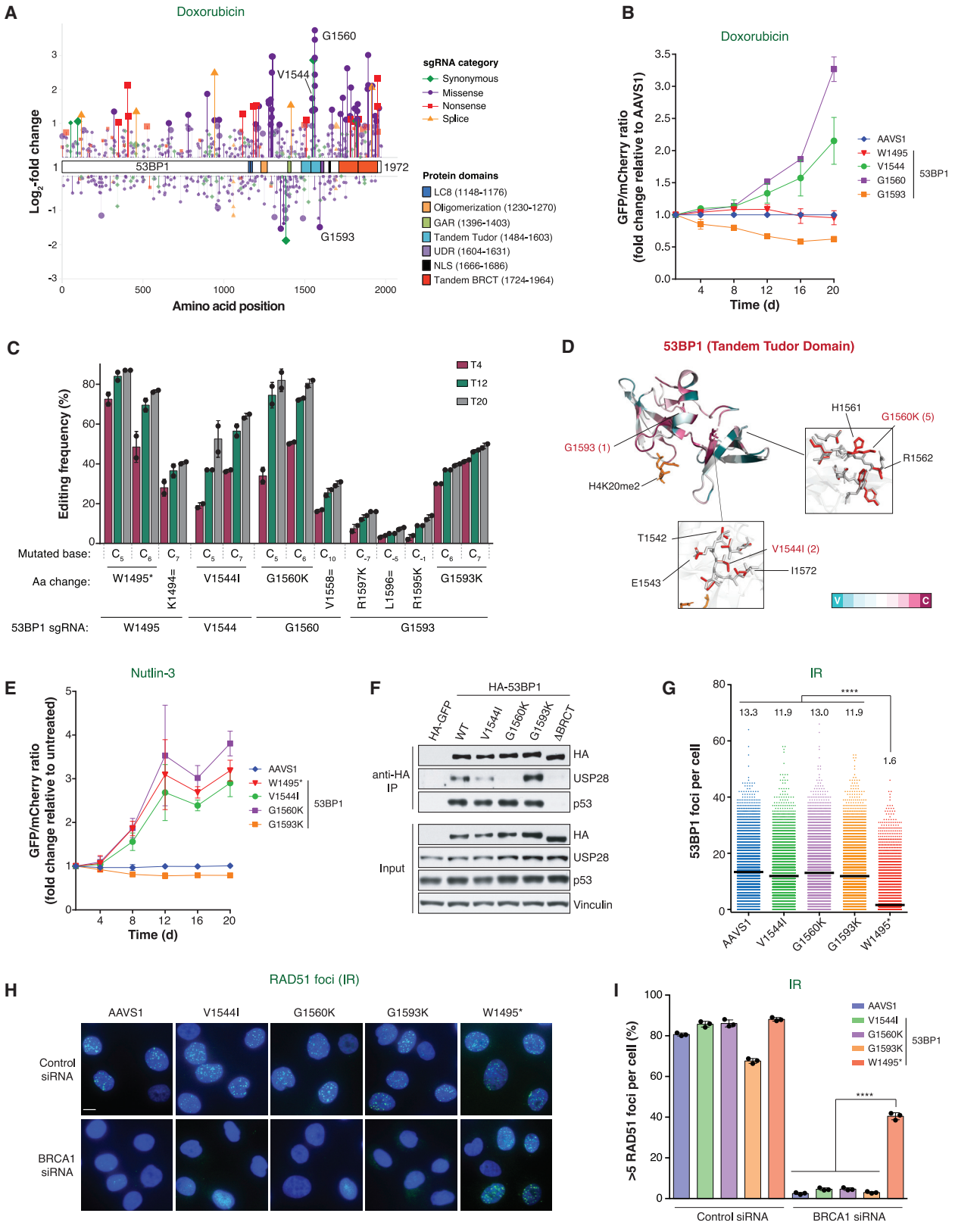
(D) Analyses of the enrichment of sgRNA-targeted genes within the groups defined by the cluster analyses in (C). Dotted line, Fisher's t test $p < 0.01$.

See also Tables S5 and S6.

Collectively, these data demonstrate that base editing screens can distinguish mutational outcomes and show how standard CRISPR-KO metrics can be applied to improve base editing screening.

sgRNA cluster analyses enable the identification of DDR mutation-phenotype associations induced by DNA damaging agents

Having established the potential of base editing screens to separate mutational outcomes in essential genes, we extended our analysis to 59 additional DDR genes (sublibrary 2, Table S1; Figure 2A) with central roles in DNA repair, characterized by well-established or newly proposed (e.g., *TONSL*) disease association or whose essentiality and/or large size complicates traditional molecular biology approaches (e.g., *RIF1*, *REV3L*, *PCNT*). We then transduced MCF10A-BE3 cells with sgRNA libraries (~37,000 sgRNAs, sublibraries 1 and 2) tiling our 86-gene network and executed base editing screens in untreated (UNT) conditions and under treatment with the DNA damaging agents cisplatin (CISP), olaparib (OLAP), doxorubicin (DOX), and camptothecin (CPT) (Figure 2B). sgRNA LFCs and p values were then computed by comparing T18 with T0 for each individual condition. To identify sgRNAs germane to specific drug responses, we selected all sgRNAs that had a biologically relevant change of statistical significance (LFC beyond the top and bottom 1% of negative control distribution and $p < 0.01$, hereafter referred to as relevant sgRNAs) in at least one of the five conditions examined and subjected them to hierarchical cluster analysis (Figure 2C; Tables S5 and S6). A total of 1,767 sgRNAs were selected for this analysis and clustered in 12 distinct groups, depending on enrichment or depletion phenotypes across the five conditions tested. A remarkable enrichment of sgRNAs targeting the *CHEK2* (hereafter, *CHK2*) and *TP53BP1* (hereafter, *53BP1*) genes was observed in groups characterized by pronounced



(legend on next page)

growth advantage in all conditions (Figures 2C and 2D, E_1, E_2), while sgRNAs targeting *ATM* caused milder growth advantage (Figures 2C and 2D, E_3, E_5). Surprisingly, sgRNAs targeting *TRAIP*, a gene that regulates replication fork metabolism (Feng et al., 2016; Hoffmann et al., 2016; Sonneville et al., 2019; Wu et al., 2019; Harley et al., 2016), were significantly present in a group conferring specific survival advantage to camptothecin (Figures 2C and 2D, E_4). As expected, common essential genes, such as *POLE*, *RAD51*, and *TONSL*, were targeted by sgRNAs localizing to groups of strong depletion phenotypes (Figures 2C and 2D, D_1, D_2), and perturbations in Fanconi anemia genes resulted primarily in sensitivity to cisplatin (Figures 2C and 2D, D_4). Together, these studies reveal that base editing screens enable the identification of mutation-phenotype associations in response to DNA damaging agents.

LOF and GOF mutations in 53BP1 define a surface on its tandem Tudor domain required for binding USP28

Given the notable enrichment of sgRNAs targeting *53BP1* in sgRNA groups associated with growth advantage (Figures 2C and 2D), we sought to characterize the functional consequences of *53BP1* nucleotide variants. 53BP1 controls p53-dependent responses through direct binding of p53 and USP28 (Cuella-Martin et al., 2016) and promotes NHEJ-mediated DSB repair by inhibiting DNA end resection (Panier and Boulton, 2014; Setiaputra and Durocher, 2019). Importantly, the distinct roles of 53BP1 in p53-dependent signaling and NHEJ are resolvable by specific SOF mutations (Cuella-Martin et al., 2016).

To accurately define the function of 53BP1 mutants identified in our screen, we conducted small-scale cluster analyses on relevant 53BP1 sgRNAs (Figure S2A). Deleterious sgRNAs largely conferred resistance to DNA damaging agents, indicating that, in the context of our screen, the growth advantage conferred by loss of cell-cycle checkpoint regulation outweighs the disadvantage caused by loss of NHEJ activity (Figure S2A). Interestingly, four sgRNAs conferring growth advantage were predicted to uniquely modify 53BP1 serine residues reported to be phosphorylated (PhosphoSite database, Figures S2A and S2B), including S500, a site recurrently identified in phosphoproteomic datasets, and S1342, an Aurora kinase B substrate implicated in mitotic spindle regulation (Wang et al., 2017). More strikingly, a small group of missense sgRNAs displayed consistent depletion across conditions (Figure S2A). By mapping onto the 53BP1 protein the predicted sgRNA-induced mutations and their LFC values in response to doxorubicin treatment (Figure 3A), we located sgRNAs with depletion phenotypes to two distinct regions: in between the oligomerization domain and the glycine-arginine-rich (GAR) motif (AAs 1288–1383) and within the tandem Tudor domain (TTD) (Figure 3A). For this study, we focused on the TTD, where a relevant sgRNA causing growth disadvantage was predicted to target residue G1593, while sgRNAs targeting V1544 (two sgRNAs) and G1560 (five sgRNAs) resulted in a marked growth advantage (Figures 3A and S2A). The growth phenotypes induced by these sgRNAs were confirmed in two-color competition assays carried out in MCF10A cells in the presence of doxorubicin (Figure 3B). Similar results were also obtained in untreated MCF10A cells and in both untreated and doxorubicin-treated MCF7 cells (Figures S2C and S2D). Interestingly, cells carrying the nonsense mutation W1495* in the TTD displayed survival advantage in untreated conditions, but not following doxorubicin treatment, possibly reflecting simultaneous impairment of p53 signaling and DSB repair (Figures 3B, S2C, and S2D). Sanger sequencing of the edited loci identified V1544I, G1560K, and G1593K as the primary mutational outcomes of the analyzed sgRNAs (Figure 3C). Analysis of the 53BP1 crystal structure predicted that the V1544I and G1560K mutations induce local alterations of the TTD surface without disrupting the overall TTD conformation (Figure 3D).

Given the role of 53BP1's TTD in p53-dependent signaling, we examined whether the V1544I, G1560K, and G1593K mutations

Figure 3. Characterization of LOF and GOF mutations in 53BP1's tandem Tudor domain that define a previously uncharacterized surface required for USP28 interaction

(A) Lollipop plot of 53BP1 sgRNAs and their LFC values in the presence of doxorubicin mapped to the canonical 53BP1 protein isoform. Lollipop transparency reflects sgRNA biological relevance, with non-transparent lollipops indicating sgRNAs with a LFC value beyond the threshold of 1% of negative controls. Symbol size reflects sgRNA statistical significance, with larger symbols indicating sgRNAs with $p < 0.01$. Residues of interest are highlighted next to their corresponding lollipop(s).

(B) Competitive growth assay in the presence of doxorubicin (2.5 nM) conducted on a BE3-expressing MCF10A cell population expressing the indicated sgRNAs. Data represent the sgRNA_of_interest_GFP/*AAVS1*_sgRNA_mCherry ratio normalized to the day 1 (T1) time point and to the corresponding GFP/mCherry *AAVS1* sgRNA ratio at each experimental time point. Mean ± SD for n = 2.

(C) Time course analysis of editing frequency induced by the indicated 53BP1 sgRNAs at their target loci, as determined by Sanger sequencing and ICE analysis (Hsiau et al., 2019). Mean ± SD for n = 2.

(D) Cartoon of the crystal structure of 53BP1's TTD bound to a H4K20me2 peptide (orange) (PDB: 2IG0, Botuyan et al. [2006]), colored based on amino acid conservation (V, variable; C, conserved). Representations of overlapping WT (gray) and mutant structures for V1544I and G1560K predicted by Missense3D (red) are shown, and AAs structurally affected are highlighted, along with the number of sgRNAs targeting each residue (in brackets).

(E) Competitive growth assays in the presence of N3 (1 μM) conducted on MCF7-BE3 cells expressing the indicated 53BP1 sgRNAs. GFP/mCherry sgRNA ratios for N3-treated conditions were normalized to the corresponding untreated controls. Mean ± SD for n = 3.

(F) Immunoblot showing USP28 and p53 co-immunoprecipitated by HA-GFP, HA-53BP1 WT, and mutants from HEK293T cell lysates using an anti-HA antibody. Images are representative of two independent experiments.

(G) Dot plot of the number of IR-induced 53BP1 foci per cell in MCF10A-BE3 cells *AAVS1*-targeted or carrying the indicated 53BP1 mutations (Table S7). Cells were fixed 4 h post-IR treatment (5 Gy) prior to imaging by high-content microscopy. Data belong to two independent experiments, and mean values for each condition are indicated. Statistical analysis was conducted using one-way ANOVA (****$p < 0.0001$).

(H) Representative images of IR-induced RAD51 foci in control and 53BP1 mutant cells upon treatment with control or BRCA1 siRNA (Table S7). Cells were treated and imaged as in (G). Scale bar, 10 μm.

(I) Quantification of IR-induced RAD51 foci per cell from (H). Mean ± SD for n = 3. Statistical analysis was conducted using unpaired t test (****$p < 0.0001$).

See also Figure S2 and Tables S5–S7.

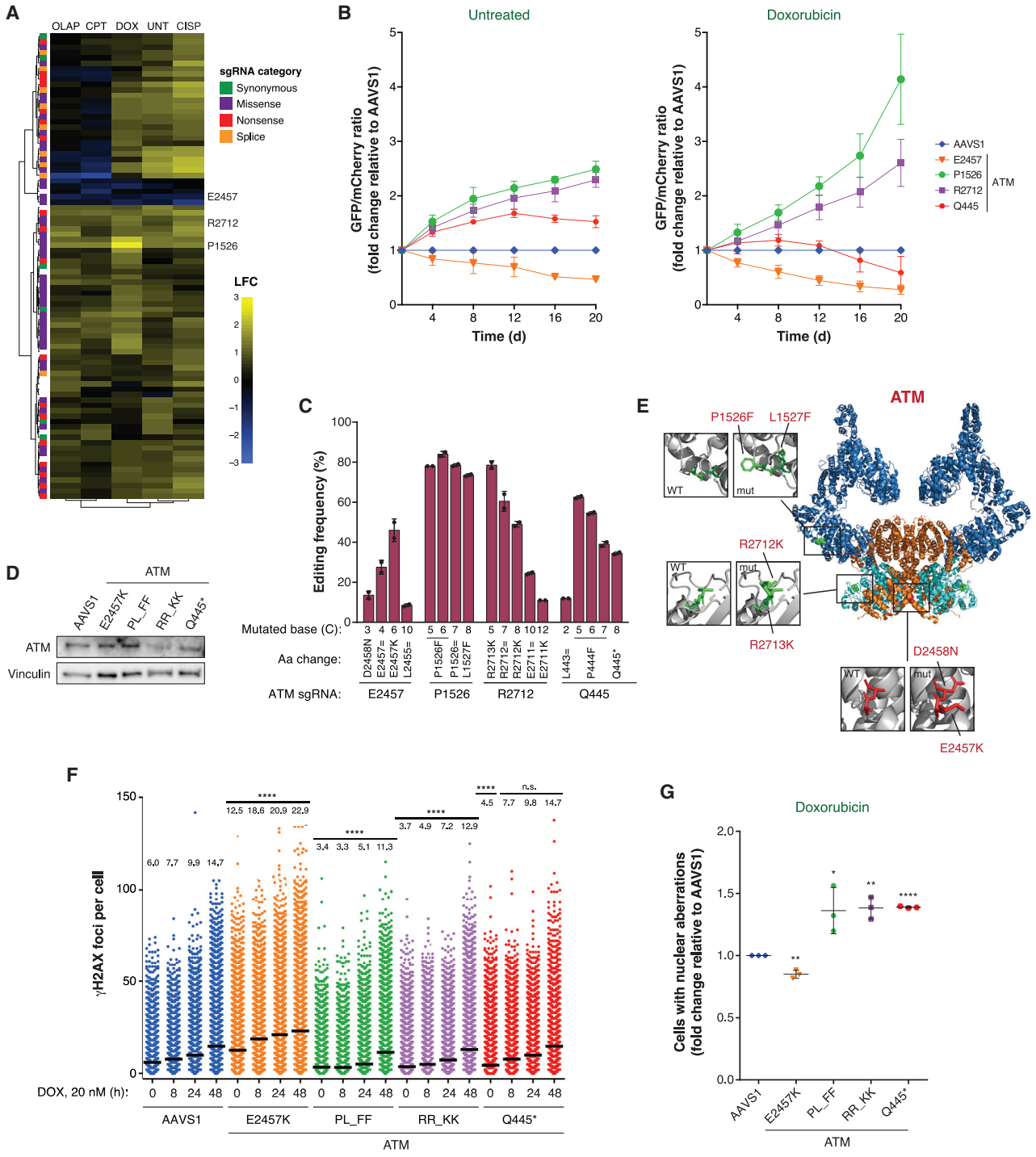


Figure 4. Identification of variants of the ATM protein with opposing phenotypes

(A) Heatmap of LFC values for relevant ATM sgRNAs in MCF10A-BE3 cells hierarchically clustered by the Ward's method. sgRNA categories are color coded, and empty-window/non-coding sgRNAs are in white. sgRNAs selected for validation are highlighted.

(B) Competitive growth assays in untreated conditions or upon doxorubicin (2.5 nM) treatment conducted on MCF10A-BE3 cells expressing the indicated ATM sgRNAs. Data are represented as in Figure 3B. Mean ± SD for n = 4 (UNT) and n = 3 (DOX).

(C) Analysis of editing frequency induced by the indicated ATM sgRNAs at their target loci at day 4 post-selection, as determined by Sanger sequencing and ICE analyses. Mean ± SD for n = 2.

(D) Immunoblot showing ATM expression in MCF10A-BE3 cells, either *AAVS1* targeted or edited with the indicated ATM sgRNAs.

(legend continued on next page)

altered the response of MCF7 cells to the p53-stabilizing small molecule Nutlin-3 (N3). Without affecting 53BP1 protein levels (Figure S2E), the V1544I and G1560K mutations conferred resistance to N3 similar to the W1495* mutation, indicating a loss of 53BP1-dependent p53 regulation (Figure 3E). Conversely, G1593K displayed a subtle but consistent increase in N3 sensitivity relative to the *AAVS1*-transduced control, suggesting that G1593K might enhance 53BP1-dependent p53 responses (Figure 3E). To determine whether V1544, G1560 and/or G1593 regulate the interaction of 53BP1 with p53 and/or USP28, we immunoprecipitated HA-tagged 53BP1 WT and mutant protein complexes from HEK293T cells (Figure 3F). Remarkably, the interaction between 53BP1 and USP28 was completely abolished by the G1560K mutation and largely diminished by the V1544I mutation, while the G1593K mutation increased 53BP1 binding to USP28, thus acting as a GOF mutation (Figure 3F). Unlike a 53BP1 mutant lacking the tandem BRCT domain, none of the analyzed mutations hindered 53BP1-p53 interactions (Figure 3F).

To examine whether the above TTD mutations interfered with 53BP1's role in DSB repair, we first evaluated the ability of 53BP1 to form nuclear foci upon ionizing radiation (IR), which depends on the interaction of the TTD with H4K20me2 (Botuyan et al., 2006). Whereas 53BP1 focus formation was largely abrogated by the W1495* mutation, V1544I, G1560K, and G1593K did not interfere with 53BP1 localization to DSBs (Figures 3G, S2F, and S2G). To evaluate the functionality of 53BP1 mutants in DSB repair, we then assessed the formation of IR-induced RAD51 foci in BRCA1-deficient cells, which is re-established upon 53BP1 deficiency (Bunting et al., 2010). Unlike the W1495* mutation, the V1544I, G1560K, and G1593K mutations failed to restore RAD51 foci formation in BRCA1-deficient cells, indicating that these 53BP1 mutants retain the ability to inhibit DSB resection and suppress HR (Figures 3H, 3I, and S2H). Collectively, these findings identify SOF mutations in the TTD of 53BP1 that uniquely abolish (V1540I, G1560K) or enhance (G1593K) 53BP1-dependent p53 regulation without interfering with 53BP1's function in DSB repair. Furthermore, this work defines a non-canonical surface of 53BP1's TTD required for binding USP28.

Identification of mutations with opposing phenotypes in the ATM kinase

Hierarchical clustering analyses revealed the presence of ATM sgRNAs that conferred mild growth advantage in response to DNA damage (Figures 2C and 2D, E_3, E_5). ATM is a large multifunctional kinase that controls cell-cycle checkpoint responses and promotes DSB repair (Blackford and Jackson, 2017). Cluster analysis of relevant ATM sgRNAs showed that deleterious sgRNAs largely elicited a growth advantage in untreated conditions and upon doxorubicin and cisplatin treatment, but not in the presence of olaparib and camptothecin (Figure 4A). In contrast, a sgRNA predicted to target E2457 conferred growth disadvantage across all conditions, while sgRNAs predicted to target P1526 and R2712 displayed consistent growth advantage (Figures 4A and S3A). These growth phenotypes were confirmed in cell competition assays conducted in MCF10A-BE3 cells in the presence of doxorubicin and olaparib and in a BE3-expressing MCF10A polyclonal cell population, albeit in a less pronounced manner (Figures 4B, S3B, and S3C). As a LOF control for these assays, we used a sgRNA generating the nonsense mutation Q445*, which caused growth advantage in untreated conditions and sensitivity to olaparib, similar to two CRISPR-KO ATM sgRNAs (Figures 4B, S3B, S3D and S3E). Sanger sequencing identified E2457K in ~50% of the alleles edited by the E2457 sgRNA and P1526_L1527delinsFF (PL_FF) and R2712_R2713delinsKK (RR_KK) in up to ~80% and ~50% of the alleles edited by the P1526 and R2712 sgRNAs, respectively (Figure 4C). Introduction of RR_KK, but not PL_FF or E2457K, reduced the levels of full-length ATM protein (Figure 4D). E2457K is located in a groove within the FAT domain of ATM (Figure 4E, in red), while PL_FF and RR_KK are surface mutations located in ATM's N-terminal solenoid and kinase domains, respectively (Figure 4E, in green). Notably, R2712K and R2713K are indexed in *ClinVar* as VUSs in a hereditary cancer-predisposing syndrome.

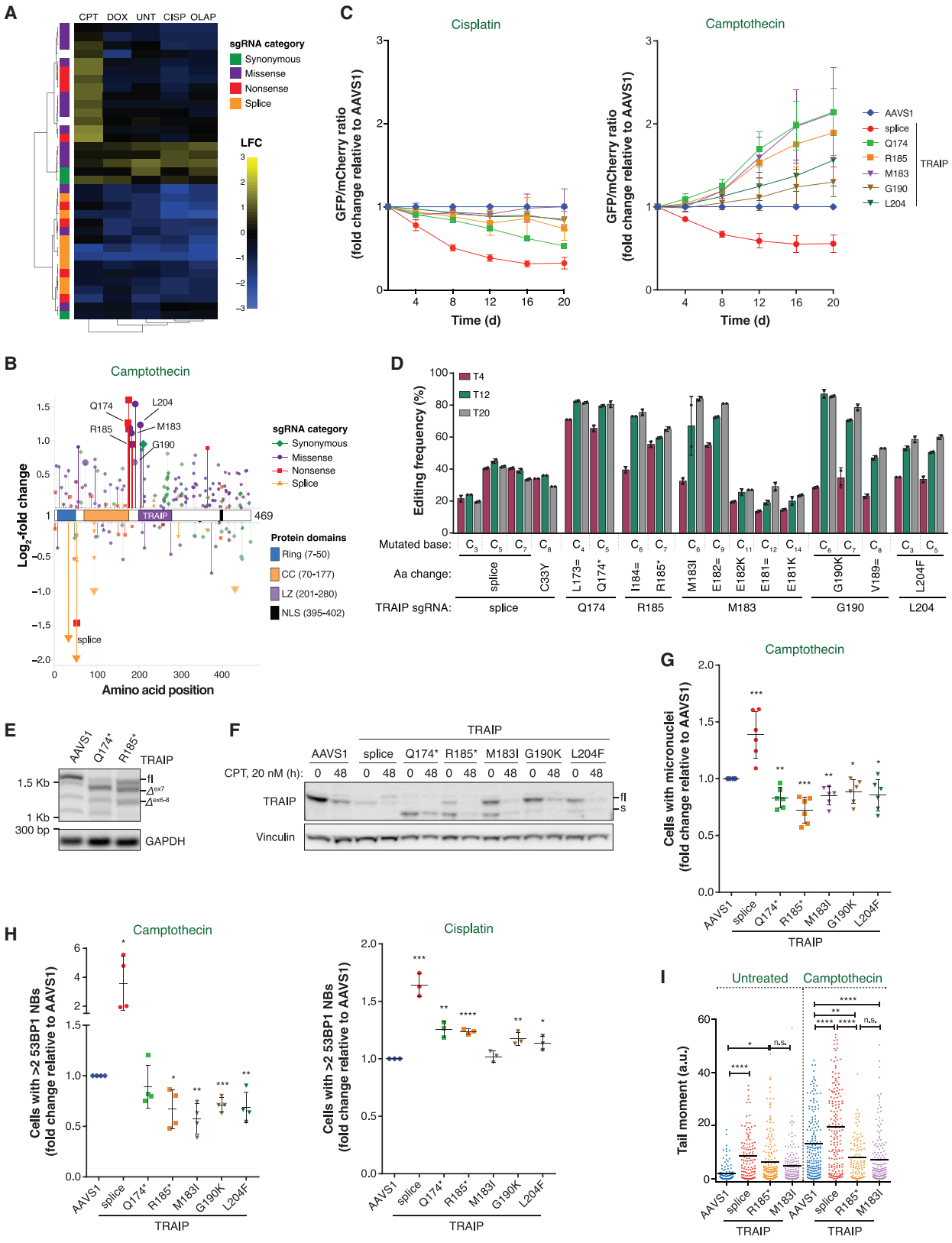
To determine the impact of the above ATM mutations on the response to DNA damage, we monitored the phosphorylation of H2AX, which is induced by ATM upon DNA damage (Ciccia and Elledge, 2010). Interestingly, E2457K mutant cells displayed increased levels of phosphorylated H2AX (γH2AX) at DNA damage sites relative to *AAVS1*-targeted cells in basal conditions and upon doxorubicin treatment (Figure 4F). The increase in γH2AX levels in E2457K mutant cells was accompanied by a mild but consistent reduction in nuclear aberrations in response to doxorubicin relative to control cells, possibly resulting from enhanced ATM-dependent DNA damage checkpoint responses (Figure 4G). Conversely, cells carrying the PL_FF and RR_KK mutations displayed lower γH2AX levels (Figure 4F), along with increased accumulation of doxorubicin-induced nuclear aberrations, likely due to reduced ATM-dependent signaling upon DNA damage (Figure 4G). Lower γH2AX levels in basal, but not doxorubicin-treated conditions, were also observed in cells carrying the Q445* mutation (Figure 4F). These studies identify mutations with opposing genome stability phenotypes in the ATM kinase.

(E) Cartoon of the crystal structure of the ATM dimer (PDB: 5NP0, Baretić et al. [2017]) with domains colored as in Figure S3A. ATM mutations verified by Sanger sequencing are highlighted in green (growth advantage) or red (growth disadvantage) in the protein structure.

(F) Dot plot of the number of doxorubicin-induced γH2AX foci per cell in MCF10A-BE3 cells *AAVS1* targeted or carrying the indicated ATM mutations (Table S7). Cells were fixed 8, 24, and 48 h after doxorubicin (20 nM) addition prior to imaging by high-content microscopy. Data belong to two independent experiments, and mean values for each condition are indicated. Statistical analysis for each condition relative to its respective *AAVS1*-targeted time point was conducted using one-way ANOVA (****$p < 0.0001$).

(G) Analyses of nuclear aberrations (micronuclei, nuclear fragments) in MCF10A cells carrying the indicated ATM mutations upon doxorubicin treatment (20 nM) for 48 h. Mean ± SD for n = 3. Statistical analysis for each sample relative to the *AAVS1*-targeted control was conducted using unpaired t test (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

See also Figure S3 and Tables S5–S7.



(legend on next page)

Characterization of a functional region in the TRAIP ubiquitin ligase whose loss or mutation results in resistance to topoisomerase I inhibition

Next, we examined the *TRAIP* gene, whose mutations caused unexpected resistance to the topoisomerase I (TOP1) inhibitor camptothecin (Figures 2C and 2D, E_4). TRAIP is an E3 ubiquitin ligase essential for mammalian development and cell viability that localizes to sites of replication stress and DSBs, where it promotes replication fork progression and HR, respectively (Feng et al., 2016; Harley et al., 2016; Hoffmann et al., 2016; Park et al., 2007; Soo Lee et al., 2016; Sonneville et al., 2019). More recently, TRAIP has also been shown to regulate ICLR (Wu et al., 2019).

In line with TRAIP's requirement for mammalian cell viability, cluster analysis showed that most sgRNAs generating deleterious mutations in TRAIP were depleted consistently across all conditions (Figure 5A). However, a distinct set of sgRNAs targeting a region of TRAIP between its coiled-coil (CC) and leucine zipper (LZ) domains (AAs 174–204) caused specific resistance to camptothecin (Figures 5A, 5B, and S4A). This set included a sgRNA predicted to insert the R185* mutation, previously described as a deleterious hypomorphic mutation that causes primordial dwarfism (Harley et al., 2016). To study the functional impact of TRAIP mutations in the 174–204 AA region, we selected sgRNAs predicted to uniquely make two nonsense (Q174* and R185*) and three missense (M183I, G190K, and L204F) mutations. As a LOF control, we selected a TRAIP sgRNA generating a splice mutation consistently depleting across conditions (splice, Figures 5A, 5B, and S4A). Competition assays in MCF10A cell populations confirmed the deleterious behavior of the splice sgRNA, similar to that of CRISPR-KO TRAIP sgRNAs (Figures 5C and S4B–S4D). Conversely, while the evaluated TRAIP sgRNAs targeting the 174–204 AA region caused mild to no growth retardation in the presence of cisplatin, they conferred a growth advantage in the presence of camptothecin relative to the *AAVS1*-targeted control, indicating the relevance of the 174–204 AA region for camptothecin-induced toxicity (Figures 5C and S4B). The predicted mutational outcomes were confirmed by Sanger sequencing for all the above sgRNAs, with the exception of the sgRNA inserting M183I, which also introduced the out-of-window mutations E181K and E182K at lower efficiency (Figure 5D).

Previous studies using *TRAIP* mutant patient cells showed that the R185* mutation was skipped by alternative splicing of *TRAIP* exon 7 (AAs 169–206), potentially resulting in shorter TRAIP protein isoforms (Harley et al., 2016). In line with this observation, Q174* or R185* caused a reduction of full-length *TRAIP* transcripts and a concomitant increase of alternatively spliced isoforms lacking exon 7 or exons 6 to 8, as determined by Sanger sequencing (Figure 5E; data not shown). Notably, sequence alignment revealed that exon 7 encodes for an evolutionarily conserved protein region (Figure S4E). Consistent with the alternative splicing results, cells carrying Q174* and R185* mutations largely lacked full-length TRAIP ($TRAIP^{fl}$) protein, instead expressing a smaller TRAIP isoform ($TRAIP^{S}$) not detected in *AAVS1*-targeted control cells (Figures 5E, 5F, and S4F). $TRAIP^{S}$ was also present, albeit at lower levels, in cells carrying TRAIP missense mutations (Figure 5F). Interestingly, treatment with camptothecin induced downregulation of both $TRAIP^{fl}$ and $TRAIP^{s}$ isoforms, and a similar effect was also observed upon UV radiation (Figures 5F and S4G) (Harley et al., 2016). However, cells expressing these TRAIP mutants were proficient for UV-induced phosphorylation of $RPA2^{S4/8}$ and H2AX, unlike TRAIP-depleted cells (Figure S4G) (Harley et al., 2016).

Next, we evaluated genomic instability in our mutants upon camptothecin or cisplatin treatment, by monitoring the formation of 53BP1 nuclear bodies (NBs) and micronuclei. TRAIP deficiency induced by the splice variant significantly increased micronuclei and 53BP1 NB formation after cisplatin and camptothecin treatments relative to the *AAVS1*-targeted control (Figures 5G, 5H, and S4H) (Sonneville et al., 2019). In contrast, the tested TRAIP mutations caused a mild decrease in the formation of micronuclei and 53BP1 NBs in response to camptothecin, but not cisplatin (Figures 5G, 5H, and S4H). In line with these findings, the R185* and M183I mutations caused fewer camptothecin-induced

Figure 5. Discovery of a functional region in the TRAIP ubiquitin ligase that promotes sensitivity to camptothecin

(A) Heatmap of LFC values for relevant TRAIP sgRNAs hierarchically clustered by the Ward's method, as shown in Figure 4A for ATM sgRNAs.

(B) Lollipop plot of TRAIP sgRNAs and their LFC values for camptothecin-treated conditions mapped to the canonical TRAIP protein isoform, as shown in Figure 3A for 53BP1. sgRNAs targeting residues selected for validation are highlighted.

(C) Competitive growth assays upon cisplatin (1 μM) or camptothecin (5 nM) treatment in BE3-expressing MCF10A cell populations carrying the indicated TRAIP sgRNAs. Data are represented as in Figure 3B. Mean ± SD for n = 3 (CISP) and n = 4 (CPT).

(D) Time course analysis of editing frequency induced by the indicated TRAIP sgRNAs at their target sites, as determined by Sanger sequencing and ICE analyses. Mean ± SD for n = 2.

(E) RT-PCR analyses of *TRAIP* transcript levels in MCF10A cells carrying the Q174* and R185* mutations. The identity of the indicated *TRAIP* splice variants was verified by Sanger sequencing. fl, full-length transcript; Δ^{ex7}, delta exon 7 transcript; Δ^{ex6-8}, delta exons 6–8 transcript.

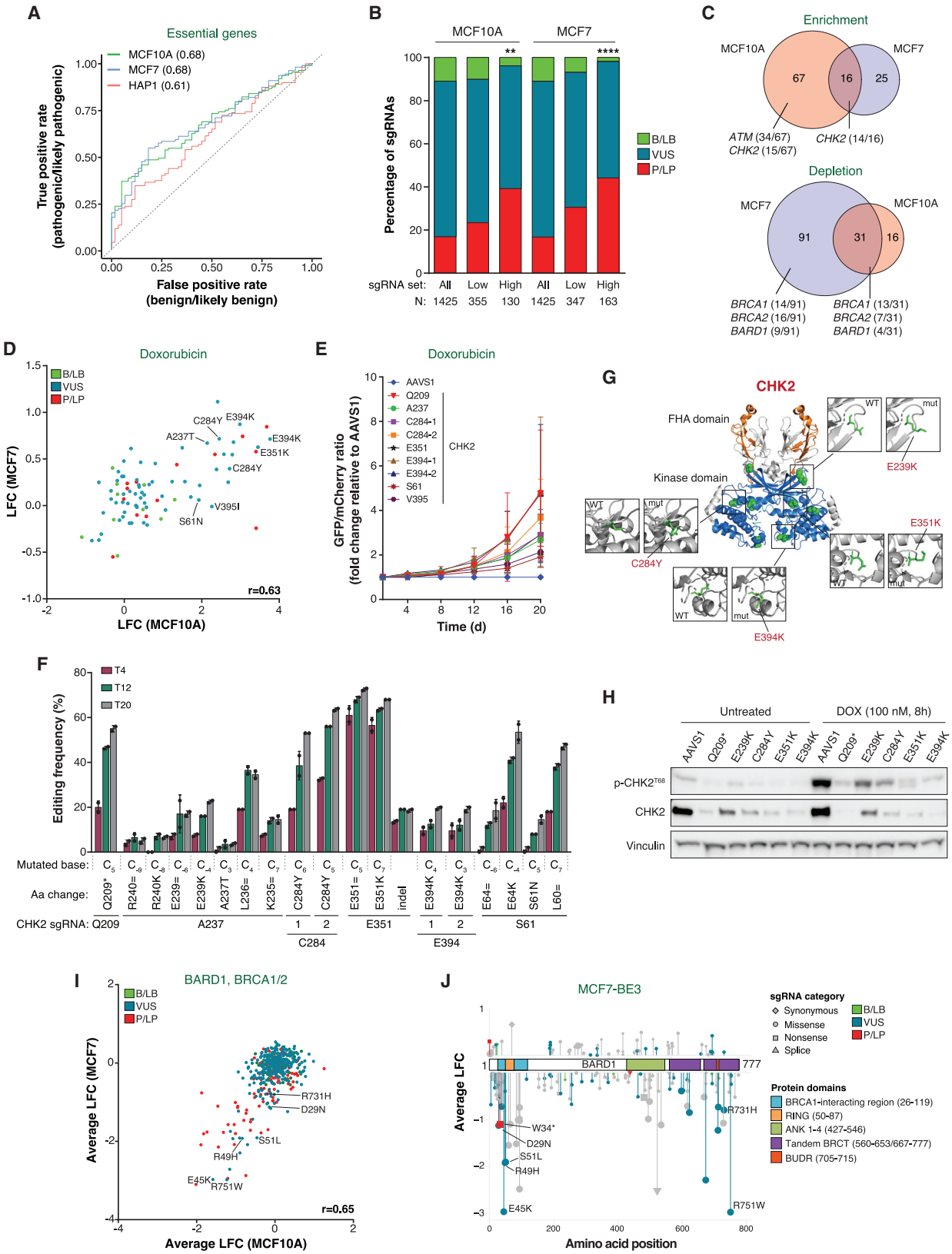
(F) Immunoblot showing TRAIP expression in MCF10A-BE3 cells targeted with the indicated sgRNAs and subjected to camptothecin (20 nM, 48 h) treatment or left untreated. fl, full-length isoform; s, short isoform.

(G) Analyses of micronuclei in MCF10A-BE3 cells targeted with the indicated sgRNAs upon camptothecin treatment (10 nM, 48 h). Values are represented as fold change relative to the *AAVS1*-targeted control. Mean ± SD for n = 6. Statistical analysis for each sample relative to the *AAVS1*-targeted control was performed using unpaired t test (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

(H) Analyses of 53BP1 NBs in MCF10A-BE3 cells *AAVS1*-targeted or carrying the indicated TRAIP mutations, subjected to camptothecin (10 nM, 48 h) or cisplatin (2 μM, 48 h) treatment. The fold change in the number of cells with >2 53BP1 NBs in each condition relative to the *AAVS1*-targeted control is shown. Mean ± SD for n = 3 (CISP) or n = 4 (CPT). Statistical analysis was performed as in (G).

(I) Neutral comet assay on MCF10A-BE3 cells *AAVS1*-targeted or carrying the indicated TRAIP mutations, subjected to mock or camptothecin (10 nM, 48 h) treatment. Mean tail moment is indicated and data represent three independent replicates. Statistical analysis was conducted by one-way ANOVA (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

See also Figure S4 and Tables S5–S7.



(legend on next page)

DSBs than the *AAVS1*-targeted control, as measured by neutral comet assay (Figure 5I). Together, these data uncover a region of functional relevance in the TRAIP ubiquitin ligase.

Application of base editing screens to identify LOF mutations of clinical relevance in DDR genes

Having established the utility of base editing screens to define the functions of DDR genes, we explored their applicability to determine the clinical relevance of VUSs. For this purpose, we focused on the 27-gene network of sublibrary 1, which included genes associated with Fanconi anemia and breast and ovarian cancer (Figure 1A). According to the *ClinVar* database, this sublibrary contained 157 sgRNAs generating benign/likely benign (B/LB) mutations, 241 sgRNAs introducing pathogenic/likely pathogenic (P/LP) variants, and 1,029 sgRNAs producing VUSs. ROC analyses on clinically relevant variants in essential genes in MCF10A-BE3, MCF7-BE3, and HAP1-BE3 cells showed that base editing screens are able to resolve P/LP from B/LB mutations (Figure 6A). To strengthen our analysis, we conducted base editing screens in MCF7-BE3 cells treated with CISP, OLAP, DOX, and CPT. We then compared the distribution of clinically relevant sgRNAs in the whole library versus sets of low-priority (LFC beyond top or bottom 5% of negative controls and $p < 0.05$) or high-priority (LFC beyond top or bottom 1% of negative controls and $p < 0.01$, relevant sgRNAs, Table S6) sgRNAs in MCF10A-BE3 and MCF7-BE3 cells. These analyses revealed significant shifts in the distribution of clinically relevant variants in both cell lines at the higher priority threshold, with the percentage of P/LP mutations consistently increasing with threshold stringency (Figure 6B).

To identify VUSs of pathogenic potential with high confidence, we explored the overlap between the relevant sgRNA sets in MCF10A-BE3 and MCF7-BE3. The MCF10A-BE3 screens identified a higher number of clinically relevant sgRNAs associated with growth advantage phenotypes, especially sgRNAs targeting DNA damage checkpoint genes (i.e., *ATM*, *CHK2*) (Figure 6C, Enrichment; Table S6). Strikingly, 14 out of the 16 clinically relevant sgRNAs conferring growth advantage in both cell lines targeted *CHK2*. Mutations in *CHK2* predispose to breast and ovarian cancer and have also been linked to Li-Fraumeni syndrome (Antoni et al., 2007; Apostolou and Papasotiriou, 2017; Bell et al., 1999). The CHK2 kinase is activated by ATM to establish p53-dependent cell-cycle arrest in response to DNA damage, and its deficiency confers resistance to DSB-inducing agents (Falck et al., 2001; Hirao et al., 2000; Takai et al., 2002).

In line with the above findings, sgRNAs targeting *CHK2* caused pronounced resistance to genotoxic agents (Figures 2C and 2D, E_1, E_2). In particular, the growth effects induced by clinically relevant CHK2 sgRNAs under doxorubicin treatment exhibited good correlation in MCF7-BE3 and MCF10A-BE3 cells, albeit they were more pronounced in MCF10A-BE3 cells (Pearson $r = 0.63$, Figure 6D). The greater phenotypic penetrance of CHK2 sgRNAs in MCF10A cells may be due to a previously unreported heterozygous *CHK2* mutation (Y390C) associated with increased breast cancer risk (Wang et al., 2015), which is expected to inactivate one *CHK2* allele in MCF10A cells (Figure S5A).

Based on the above analysis, we selected as high-confidence hits sgRNAs predicted to uniquely introduce a *CHK2* VUS (i.e., A237T, C284Y (2), E351K and E394K (2)) and displaying enrichment behavior in the presence of doxorubicin in both MCF10A-BE3 and MCF7-BE3 cells (Figure 6D). sgRNAs enriched only in MCF10A-BE3 cells upon doxorubicin treatment (i.e., sgRNAs predicted to generate S61N and V395I) were selected as low-confidence hits (Figure 6D). Validation of the screen phenotypes by competition assay confirmed that all the evaluated sgRNAs caused a growth advantage upon doxorubicin and olaparib treatment, similar to a sgRNA generating the LOF nonsense mutation Q209* (Figures 6E and S5C). Sanger sequencing verified

Figure 6. Analysis of variants of uncertain significance with pathogenic-like behavior

(A) ROC analyses of MAGeCK ranks for sgRNAs predicted to introduce pathogenic and likely pathogenic mutations (true positives) versus benign and likely benign sgRNAs (false positives) in essential genes from sublibrary 1. AUC values are indicated in brackets.

(B) Distribution of low-priority (LFC beyond top or bottom 5% of negative controls and $p < 0.05$ in any treatment) or high-priority (LFC beyond top or bottom 1% of negative controls and $p < 0.01$ in any treatment, Table S6) sgRNAs predicted to introduce clinically relevant mutations in MCF10A-BE3 and MCF7-BE3 cells. Statistical analysis relative to the full set of relevant sgRNAs in MCF10A-BE3 or MCF7-BE3 cells was conducted using a chi-square test (**$p < 0.01$, ****$p < 0.0001$).

(C) Venn diagrams of high-priority sgRNAs generating clinically relevant mutations in MCF10A-BE3 and MCF7-BE3 screens (Table S6). LFC values for each individual sgRNA across the 5 conditions (UNT, CISP, OLAP, DOX, CPT) were averaged and sgRNAs with average LFC > 0 were assigned to the enrichment set, while sgRNAs with average LFC < 0 were assigned to the depletion set.

(D) Correlation of LFC values for clinically relevant sgRNAs targeting the *CHK2* gene upon doxorubicin treatment in MCF10A-BE3 versus MCF7-BE3 cells. Pearson correlation value is shown. sgRNAs targeting residues selected for validation are highlighted.

(E) Competitive growth assays in the presence of doxorubicin (2.5 nM) conducted on MCF10A-BE3 cells expressing the indicated sgRNAs. Data are represented as in Figure 3B. Mean ± SD for n = 4.

(F) Time course analysis of editing frequency induced by the indicated CHK2 sgRNAs at their target loci, as determined by Sanger sequencing and ICE analysis. Mean ± SD for n = 2.

(G) Cartoon of the crystal structure of the CHK2 dimer (PDB: 3I6W, Cai et al. [2009]) with domains colored as in Figure S5B. Mutations verified by Sanger sequencing are highlighted in green.

(H) Immunoblot showing CHK2 expression and its phosphorylation on T68 in MCF10A-BE3 cells, either *AAVS1* targeted or edited by CHK2 sgRNAs, with or without doxorubicin treatment (100 nM, 8 h).

(I) Correlation of averaged LFC values (CISP, OLAP) for clinically relevant sgRNAs targeting *BARD1*, *BRCA1*, and *BRCA2* in MCF10A-BE3 versus MCF7-BE3 cells. Pearson correlation value is shown. sgRNAs that uniquely generate a VUS in the *BARD1* gene are indicated.

(J) Lollipop plot of BARD1 sgRNAs and their average LFC values (CISP, OLAP) in MCF7-BE3 cells mapped to the canonical BARD1 protein isoform. sgRNA categories are distinguished by symbol shape, and larger symbols indicate sgRNAs with $p < 0.01$ in olaparib and/or cisplatin treatments. Colors indicate clinical relevance. sgRNAs that uniquely generate a VUS with pathogenic-like behavior in the *BARD1* gene are indicated.

See also Figures S5 and S6 and Tables S5–S7.

that the selected sgRNAs generated the predicted mutations, with the exception of sgRNAs targeting A237 and S61, which resulted in the generation of a second VUS (E239K and E64K, respectively) with higher efficiency than the predicted mutation (A237T and S61N, respectively), and sgRNA targeting V395 (no editing) (Figure 6F). The analyzed high-confidence CHK2 mutations, which localized to the kinase domain, significantly decreased CHK2 protein expression comparably to Q209* (Figures 6G and 6H). Accordingly, only residual ATM-dependent phosphorylation of CHK2 on T68 was observed in response to doxorubicin treatment (Figure 6H). In the case of the E64K mutation, the effect on CHK2 protein levels and ATM-dependent CHK2 phosphorylation was less pronounced, consistent with the milder growth advantage of cells carrying this mutation (Figures 6D, 6E, S5C, and S5E). Together, these findings show that all sgRNAs selected (9/9) validated the screen results (Figures 6E, S5B, and S5C) and illustrate the importance of confirming sgRNA-induced mutational outcomes, since 3 of 9 tested CHK2 sgRNAs did not introduce the expected mutations at high frequency.

Clinically relevant sgRNAs with depletion phenotypes also exhibited notable overlap between MCF7-BE3 and MCF10A-BE3 cells (Figure 6C, Depletion; Table S6). MCF7-BE3 cells displayed a greater number of clinically relevant sgRNAs causing growth disadvantage, with more than 50% of them targeting breast cancer-associated HR genes (i.e., *BRCA1*, *BRCA2*, and *BARD1*) (Figure 6C). LFC values for clinically relevant sgRNAs in *BRCA1*, *BRCA2*, and *BARD1* under cisplatin and olaparib treatment showed high correlation for both cell lines (Pearson r = 0.6716 for MCF10A-BE3, r = 0.809 for MCF7-BE3). We then averaged the LFC values under cisplatin and olaparib treatment to draw an inter-cell line correlation (Pearson r = 0.65) and pinpoint VUSs in *BRCA1*, *BRCA2*, and *BARD1* with pathogenic-like behavior (Figure 6I). High-confidence sgRNAs generating VUSs primarily targeted *BARD1* and localized to known functional domains, i.e., the BRCA1-interacting region and the tandem BRCT motifs (Figures 6I and 6J). Interestingly, two BARD1 sgRNAs localized to the UDR motif within the BRCT domain that recognizes H2AK15ub (Becker et al., 2020) (Figure 6J). By selecting only sgRNAs predicted to uniquely make VUSs, we identified three high-confidence (E45K, R49H, R751W) and three low-confidence (D29N, S51L, R731H) mutations that displayed P/LP behavior. Similar analysis on the *BRCA1* and *BRCA2* genes revealed that missense sgRNAs of P/LP behavior localized to critical protein domains, coinciding with results from similar base editing screens conducted by Doench and colleagues (Figures S5F and S5G) (Hanna et al., 2021). Indeed, 3 of 7 deleterious, clinically relevant, missense sgRNAs in BRCA1 localized to the RING domain and the tandem BRCT motifs, while 8 of 10 clinically relevant, missense BRCA2 sgRNAs targeted the helical and OB-fold domains (Figures S5F and S5G). Collectively, these studies indicate that base editing screens can identify VUSs that exhibit P/LP behavior.

Additional DDR genes with mutational patterns of interest

We then sought to extend our analyses of relevant sgRNAs in MCF10A-BE3 and MCF7-BE3 to all sublibrary 1 sgRNAs to pinpoint additional overlapping and cell-line-specific hits. Interestingly, sgRNAs targeting *BRIP1* conferred growth advantage in MCF7-BE3, but not MCF10A-BE3 cells (Figure S6A, Enrichment; Table S6). BRIP1 is a DNA helicase mutated in breast and ovarian cancer and Fanconi anemia that promotes HR and ICLR (Bridge et al., 2005; Cantor et al., 2001; Litman et al., 2005). Unlike other HR genes, LOF mutations in BRIP1 conferred unexpected resistance to camptothecin and, to a lesser extent, olaparib in MCF7-BE3 cells (Figure S6B). Missense mutations with LOF behavior targeted critical residues, such as the conserved C284 in the Fe-S domain, or surface residues of the arch and helicase domains (i.e., T411, T722, L810, or H835) (Figures S6C and S6D).

MCF10A-BE3 and MCF7-BE3 cells shared depletion phenotypes for sgRNAs targeting the ATR kinase (Figure S6A, Depletion; Table S6). The averaged cisplatin and olaparib LFC values for ATR sgRNAs showed good inter-cell line correlation (Pearson $r > 0.69$), enabling us to pinpoint high-confidence deleterious sgRNAs generating missense variants (Figures S6E and S6F) in the FAT and kinase domains (8/14), and/or targeting surface residues (5/14), three of which located in a lateral groove (S1115_P1116, H1587_L1588, and R1653_V1655) (Figures S6F and S6G).

Finally, we explored the depletion phenotypes of sgRNAs targeting *CDK12* (Figure S6A, Depletion; Table S6), a kinase that promotes the expression of DNA repair genes, including *BRCA1* and *ATR* (Blazek et al., 2011). Accordingly, nonsense and deleterious missense mutations within CDK12's proline-rich motifs (PRM1-2) and kinase domain caused growth disadvantage across conditions (Figures S6H and S6I). Among them, Q1088* and G909E mutations are considered pathogenic according to TCGA pancancer analyses, and sgRNAs generating the VUSs R344W, A284V, and E1041K display LOF behavior (Figure S6H, bold font; Table S5). Collectively, these data highlight the potential of base editing screens for dissecting the functionality of DDR genes and determining the impact of clinically relevant mutations.

DISCUSSION

In this study, we combined CRISPR-mediated base editing and pooled screening technologies for large-scale analyses of human nucleotide variants across multiple cell lines. Application of base editing screens to an 86-gene DDR network under the selective pressure of DNA damaging agents allowed us to uncover phenotypes distinct from those obtained using traditional CRISPR-KO screens. To facilitate the interrogation of our dataset, we developed an online database including the predicted mutational outcome and protein localization for the studied sgRNAs, alongside their clinical relevance, and statistical and biological significance (https://www.ciccialab-database.com/ddr-variants). This resource will facilitate the study of DDR variants and the evaluation of their impact on human pathology.

Base editing screening: current features and future optimization

At least three aspects need to be considered for optimizing library design, targeting range, phenotypic resolution, and editing

specificity in base editing screens. First, the design of base editing sgRNA libraries can be optimized by applying CRISPR-KO off- and on-target scores (Doench et al., 2016) (Figures 1E–1G, S1G, and S1H). Similarly, existing CRISPR-KO analysis algorithms (MAGeCK, DrugZ) coupled to SNV annotation tools (ANNOVAR) can be used to evaluate base editing screen data (Colic et al., 2019; Li et al., 2014; Wang et al., 2010). Second, the repertoire of variants inserted and the targeting range of base editing screens can be broadened with cytosine and adenine base editors with relaxed PAM requirements (Gaudelli et al., 2017; Hu et al., 2018; Huang et al., 2019; Nishimasu et al., 2018; Walton et al., 2020). Third, base editors with narrow (1–2 nucleotide) activity windows can limit the number of mutational outcomes caused by multiple editable bases within and/or outside of the editing window (Figures 3C, 4C, 5D, and 6F) (Kim et al., 2017; Tan et al., 2019). In addition, coupling base editing screens with individual validation of candidate variants enables the characterization of the precise mutation(s) responsible for the observed phenotype.

Applications of base editing screening technologies

Through our studies, we demonstrate the versatility of base editing screening technologies and highlight their use for the applications discussed below.

Identification of LOF, GOF, and SOF mutations

Our work shows that base editing screens enable the identification of SOF mutations critical for functional dissection of complex multidomain proteins. To exemplify this point, we identified and characterized mutations of 53BP1's TTD domain that confer gain (G1593K) or loss (V1544I, G1560K) of 53BP1-dependent p53 regulation, without interfering with the role of 53BP1 in DSB repair (Figures 3 and S2). The 53BP1 TTD has been implicated in regulating the USP28-53BP1-p53 axis independently of p53 binding (Cuella-Martin et al., 2016). Here, we describe the TTD as the second 53BP1 domain, in addition to the tandem BRCT, that regulates the USP28-53BP1 interaction (Figure 3F) (Cuella-Martin et al., 2016; Knobel et al., 2014). In particular, the TTD residues V1544 and G1560, located opposite to the methyl-K/R binding pocket (Figure 3D) (Botuyan et al., 2006), define a non-canonical TTD interaction surface that promotes the association of 53BP1 with USP28. Our mutational analyses also revealed a possible role in p53 regulation for an uncharacterized 53BP1 region (AAs 1288–1383) located in the vicinity of its oligomerization domain (Figure 3A). Two clusters of 5 and 4 sgRNAs within this region conferred growth advantage and disadvantage, respectively (Figures 3A and S2A), suggesting that the resulting 53BP1 variants might regulate p53 signaling similarly to the above TTD mutations. Further characterization of these and other 53BP1 motifs (e.g., phosphosites) will expedite our understanding of the role of 53BP1 in p53-mediated responses.

Identification of functional domains that mediate treatment-specific phenotypes

Our observations also highlight the potential of base editing screens to identify treatment-specific phenotypes and discover functional domains. In particular, we characterized a subset of TRAIP mutations, clustered in a previously undefined but evolutionary-conserved 30 AA (174–204) region, that cause growth advantage and enhanced genomic stability uniquely in response to camptothecin (Figures 5B, 5C, 5G–5I, S4E, and S4H). Camptothecin acts by trapping TOP1 onto DNA, thus generating DNA single-strand breaks that can be converted into DSBs during DNA replication (Liu et al., 2000). Since TRAIP interacts with TOP1 (Feng et al., 2016), TRAIP mutations in this region may interfere with the loading of TOP1 on chromatin or enhance the removal of TOP1-DNA adducts, resulting in the formation of fewer camptothecin-induced DSBs and a consequent growth advantage. Alternatively, these TRAIP mutants might possess elevated HR-promoting activities relative to the wild-type protein (Soo Lee et al., 2016), which would enable more efficient repair of camptothecin-induced DSBs. Given that the TRAIP R185* mutation causes primordial dwarfism, further mechanistic studies defining the 174–204 AA region may provide insights into the etiology of this disease.

Phenotypic characterization of disease-associated genetic variants

Our study analyzes clinically relevant variants in 75 DDR genes associated with genetic syndromes and/or cancer (Tables S1 and S5). We demonstrate that base editing screens can separate pathogenic from benign mutations and identify > 240 clinically relevant DDR variants exhibiting altered cellular growth and/or response to genotoxic agents (Figures 6A–6C; Table S6). In particular, we characterized 5 *CHK2* VUSs and identified 12 *BARD1*, *BRCA1*, and *BRCA2* VUSs that display behavior comparable to that of P/LP variants predisposing to breast cancer (Figures 6C–6J and S5B–S5G). Importantly, we show that accurate variant classification through base editing screens relies on the generation of cellular scenarios for optimal penetrance of mutation-induced phenotypes and requires the verification of sgRNA-induced mutational outcomes (Figures 6E and 6F). Future improvements in editing scope and precision will increase the ability of base editing screens to categorize clinically relevant variants.

Conclusions and future perspectives

Collectively, our work highlights the potential of base editing screens to query nucleotide variants, allowing for the separation of phenotypes that are often confounded in analyses based on full protein loss. Mutational tiling of endogenous loci enables the study of essential and large genes, overcoming limitations of current structure-function approaches. From a clinical standpoint, this technology can guide the identification of mutations with pathogenic-like behavior in genetic disorders and cancer and facilitate the discovery of mutation-drug interactions for the development of personalized therapies. Together, base editing screening technologies will help elucidate the molecular determinants of human disease and accelerate the path to precision medicine.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cell.2021.01.041.

ACKNOWLEDGMENTS

We thank Richard Baer and Alex Chavez for comments on the manuscript, Han S. Ryu for technical support, Eric Bryant and Juan Patiño Galindo for bioinformatic support, Enrique Vazquez de Luis for NGS advice, and Aaron and Sarah Meltzer for the online database. LentiGuide fluorescent constructs and TRAIP antibodies were kindly provided by Daniel Durocher, Niels Mailand, and Andrew Jackson. Graphical abstract was created using BioRender. This work was supported by the NIH grants R01CA197774 and R01CA227450 to A.C., P01CA174653 to A.C. and R.R., the EMBO Long-Term Fellowship (ALTF 366-2019) to R.C.-M., and the Damon Runyon Cancer Research Foundation (DFS-28-18) and Pew-Stewart Scholar for Cancer Research program to C.L.

AUTHOR CONTRIBUTIONS

Conceptualization, R.C.-M. and A.C.; Methodology, R.C.-M., S.B.H., J.-W.H., A.T., and G.L.; Formal Analyses, R.C.-M., S.B.H., X.F., X.C., and J.Z.; Investigation, R.C.-M.; Writing – Original Draft, R.C.-M. and A.C.; Writing – Review & Editing, R.C.-M., S.B.H., X.F., X.C., J.-W.H., A.T., G.L., J.Z., R.R., C.L., Y.S., and A.C.; Supervision, R.R., C.L., Y.S., and A.C.; Funding Acquisition, R.C.-M. and A.C.

DECLARATION OF INTERESTS

R.C.-M., S.B.H., and A.C. have filed a provisional patent application (63/068,928) derived from this work.

Received: May 6, 2020
Revised: November 16, 2020
Accepted: January 22, 2021
Published: February 18, 2021

WEB RESOURCES

Functional analysis of nucleotide variants in DDR genes, https://www.ciccialab-database.com/ddr-variants

REFERENCES

Adamovich, A.I., Banerjee, T., Wingo, M., Duncan, K., Ning, J., Martins Rodrigues, F., Huang, K.-L., Lee, C., Chen, F., Ding, L., and Parvin, J.D. (2019). Functional analysis of BARD1 missense variants in homology-directed repair and damage sensitivity. PLoS Genet. *15*, e1008049.

Antoni, L., Sodha, N., Collins, I., and Garrett, M.D. (2007). CHK2 kinase: cancer susceptibility and cancer therapy - two sides of the same coin? Nat. Rev. Cancer *7*, 925–936.

Apostolou, P., and Papasotiriou, I. (2017). Current perspectives on CHEK2 mutations in breast cancer. Breast Cancer (Dove Med. Press) *9*, 331–335.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. *44* (W1), W344-50.

Baretić, D., Pollard, H.K., Fisher, D.I., Johnson, C.M., Santhanam, B., Truman, C.M., Kouba, T., Fersht, A.R., Phillips, C., and Williams, R.L. (2017). Structures of closed and open conformations of dimeric human ATM. Sci. Adv. *3*, e1700933.

Becker, J.R., Bonnet, C., Clifford, G., Groth, A., Wilson, M.D., and Chapman, J.R. (2020). BARD1 links histone H2A Lysine-15 ubiquitination to initiation of BRCA1-dependent homologous recombination. bioRxiv, 2020.06.01.127951.

Bell, D.W., Varley, J.M., Szydlo, T.E., Kang, D.H., Wahrer, D.C., Shannon, K.E., Lubratovich, M., Verselis, S.J., Isselbacher, K.J., Fraumeni, J.F., et al. (1999). Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. Science *286*, 2528–2531.

Billing, D., Horiguchi, M., Wu-Baer, F., Taglialatela, A., Leuzzi, G., Nanez, S.A., Jiang, W., Zha, S., Szabolcs, M., Lin, C.-S., et al. (2018). The BRCT Domains of the BRCA1 and BARD1 Tumor Suppressors Differentially Regulate Homology-Directed Repair and Stalled Fork Protection. Mol. Cell *72*, 127–139.e8.

Billon, P., Bryant, E.E., Joseph, S.A., Nambiar, T.S., Hayward, S.B., Rothstein, R., and Ciccia, A. (2017). CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons. Mol. Cell *67*, 1068–1079.e4.

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

Blazek, D., Kohoutek, J., Bartholomeeusen, K., Johansen, E., Hulinkova, P., Luo, Z., Cimermancic, P., Ule, J., and Peterlin, B.M. (2011). The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. Genes Dev. *25*, 2158–2172.

Boonen, R.A.C.M., Rodrigue, A., Stoepker, C., Wiegant, W.W., Vroling, B., Sharma, M., Rother, M.B., Celosse, N., Vreeswijk, M.P.G., Couch, F., et al. (2019). Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. Nat. Commun. *10*, 5296.

Botuyan, M.V., Lee, J., Ward, I.M., Kim, J.-E., Thompson, J.R., Chen, J., and Mer, G. (2006). Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. Cell *127*, 1361–1373.

Bouwman, P., van der Gulden, H., van der Heijden, I., Drost, R., Klijn, C.N., Prasetyanti, P., Pieterse, M., Wientjens, E., Seibler, J., Hogervorst, F.B.L., and Jonkers, J. (2013). A high-throughput functional complementation assay for classification of BRCA1 missense variants. Cancer Discov. *3*, 1142–1155.

Bridge, W.L., Vandenberg, C.J., Franklin, R.J., and Hiom, K. (2005). The BRIP1 helicase functions independently of BRCA1 in the Fanconi anemia pathway for DNA crosslink repair. Nat. Genet. *37*, 953–957.

Bunting, S.F., Callén, E., Wong, N., Chen, H.-T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L., et al. (2010). 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. Cell *141*, 243–254.

Cai, Z., Chehab, N.H., and Pavletich, N.P. (2009). Structure and activation mechanism of the CHK2 DNA damage checkpoint kinase. Mol. Cell *35*, 818–829.

Cantor, S.B., Bell, D.W., Ganesan, S., Kass, E.M., Drapkin, R., Grossman, S., Wahrer, D.C., Sgroi, D.C., Lane, W.S., Haber, D.A., and Livingston, D.M.

(2001). BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. Cell *105*, 149–160.

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

Colic, M., Wang, G., Zimmermann, M., Mascall, K., McLaughlin, M., Bertolet, L., Lenoir, W.F., Moffat, J., Angers, S., Durocher, D., and Hart, T. (2019). Identifying chemogenetic interactions from CRISPR screens with drugZ. Genome Med. *11*, 52.

Cuella-Martin, R., Oliveira, C., Lockstone, H.E., Snellenberg, S., Grolmusova, N., and Chapman, J.R. (2016). 53BP1 Integrates DNA Repair and p53-Dependent Cell Fate Decisions via Distinct Mechanisms. Mol. Cell *64*, 51–64.

Després, P.C., Dubé, A.K., Seki, M., Yachie, N., and Landry, C.R. (2020). Perturbing proteomes at single residue resolution using base editing. Nat. Commun. *11*, 1871.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

Falck, J., Mailand, N., Syljuåsen, R.G., Bartek, J., and Lukas, J. (2001). The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. Nature *410*, 842–847.

Farrugia, D.J., Agarwal, M.K., Pankratz, V.S., Deffenbaugh, A.M., Pruss, D., Frye, C., Wadum, L., Johnson, K., Mentlick, J., Tavtigian, S.V., et al. (2008). Functional assays for classification of BRCA2 variants of uncertain significance. Cancer Res. *68*, 3523–3531.

Feng, W., and Jasin, M. (2017). BRCA2 suppresses replication stress-induced mitotic and G1 abnormalities through homologous recombination. Nat. Commun. *8*, 525.

Feng, W., Guo, Y., Huang, J., Deng, Y., Zang, J., and Huen, M.S.-Y. (2016). TRAIP regulates replication fork recovery and progression via PCNA. Cell Discov. *2*, 16016.

Findlay, G.M., Daza, R.M., Martin, B., Zhang, M.D., Leith, A.P., Gasperini, M., Janizek, J.D., Huang, X., Starita, L.M., and Shendure, J. (2018). Accurate classification of BRCA1 variants with saturation genome editing. Nature *562*, 217–222.

Gapinske, M., Luu, A., Winter, J., Woods, W.S., Kostan, K.A., Shiva, N., Song, J.S., and Perez-Pinera, P. (2018). CRISPR-SKIP: programmable gene splicing with single base. Genome Biol. *19*, 107–111.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature *551*, 464–471.

Ghezraoui, H., Oliveira, C., Becker, J.R., Bilham, K., Moralli, D., Anzilotti, C., Fischer, R., Deobagkar-Lele, M., Sanchiz-Calvo, M., Fueyo-Marcos, E., et al. (2018). 53BP1 cooperation with the REV7-shieldin complex underpins DNA structure-specific NHEJ. Nature *560*, 122–127.

Guidugli, L., Carreira, A., Caputo, S.M., Ehlen, A., Galli, A., Monteiro, A.N.A., Neuhausen, S.L., Hansen, T.V.O., Couch, F.J., and Vreeswijk, M.P.G.; ENIGMA consortium (2014). Functional assays for analysis of variants of uncertain significance in BRCA2. Hum. Mutat. *35*, 151–164.

Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., and Taipale, J. (2018). CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. Nat. Med. *24*, 927–930.

Hanna, R.E., Hegde, M., Fagre, C.R., DeWeirdt, P.C., Sangree, A.K., Szegletes, Z., Griffith, A., Feeley, M.N., Sanson, K.R., Baidi, Y., et al. (2021). Massively parallel assessment of human variants with base editor screens. Cell *184*, this issue, 1064–1080.

Harley, M.E., Murina, O., Leitch, A., Higgs, M.R., Bicknell, L.S., Yigit, G., Blackford, A.N., Zlatanou, A., Mackenzie, K.J., Reddy, K., et al. (2016). TRAIP promotes DNA damage response during genome replication and is mutated in primordial dwarfism. Nat. Genet. *48*, 36–43.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell *163*, 1515–1526.

Hirao, A., Kong, Y.Y., Matsuoka, S., Wakeham, A., Ruland, J., Yoshida, H., Liu, D., Elledge, S.J., and Mak, T.W. (2000). DNA damage-induced activation of p53 by the checkpoint kinase Chk2. Science *287*, 1824–1827.

Hoffmann, S., Smedegaard, S., Nakamura, K., Mortuza, G.B., Räschle, M., Ibañez de Opakua, A., Oka, Y., Feng, Y., Blanco, F.J., Mann, M., et al. (2016). TRAIP is a PCNA-binding ubiquitin ligase that protects genome stability after replication stress. J. Cell Biol. *212*, 63–75.

Hsiau, T., Conant, D., Rossi, N., Maures, T., Waite, K., Yang, J., Joshi, S., Kelso, R., Holden, K., Enzmann, B.L., et al. (2019). Inference of CRISPR Edits from Sanger Trace Data. bioRxiv. https://doi.org/10.1101/251082.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262–1278.

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

Huang, S., Liao, Z., Li, X., Liu, Z., Li, G., Li, J., Lu, Z., Zhang, Y., Li, X., Ma, X., et al. (2019). Developing ABEmax-NG with Precise Targeting and Expanded Editing Scope to Model Pathogenic Splice Site Mutations In Vivo. iScience *15*, 640–648.

Hustedt, N., Álvarez-Quilón, A., McEwan, A., Yuan, J.Y., Cho, T., Koob, L., Hart, T., and Durocher, D. (2019). A consensus set of genetic vulnerabilities to ATR inhibition. Open Biol. *9*, 190156.

Ihry, R.J., Worringer, K.A., Salick, M.R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C., et al. (2018). p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. Nat. Med. *24*, 939–946.

Ittisoponpisan, S., Islam, S.A., Khanna, T., Alhuzimi, E., David, A., and Sternberg, M.J.E. (2019). Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? J. Mol. Biol. *431*, 2197–2212.

Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature *461*, 1071–1078.

Joung, J., Konermann, S., Gootenberg, J.S., Abudayyeh, O.O., Platt, R.J., Brigham, M.D., Sanjana, N.E., and Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nat. Protoc. *12*, 828–863.

Jun, S., Lim, H., Chun, H., Lee, J.H., and Bang, D. (2020). Single-cell analysis of a mutant library generated using CRISPR-guided deaminase in human melanoma cells. Commun Biol *3*, 154.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. *10*, 845–858.

Khanna, K.K., and Jackson, S.P. (2001). DNA double-strand breaks: signaling, repair and the cancer connection. Nat. Genet. *27*, 247–254.

Kim, Y.B., Komor, A.C., Levy, J.M., Packer, M.S., Zhao, K.T., and Liu, D.R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat. Biotechnol. *35*, 371–376.

Knobel, P.A., Belotserkovskaya, R., Galanty, Y., Schmidt, C.K., Jackson, S.P., and Stracker, T.H. (2014). USP28 is recruited to sites of DNA damage by the tandem BRCT domains of 53BP1 but plays a minor role in double-strand break metabolism. Mol. Cell. Biol. *34*, 2062–2074.

Koblan, L.W., Erdos, M.R., Wilson, C., Cabral, W.A., Levy, J.M., Xiong, Z.-M., Tavarez, U.L., Davison, L.M., Gete, Y.G., Mao, X., et al. (2021). In vivo base editing rescues Hutchinson-Gilford progeria syndrome in mice. Nature *589*, 608–614.

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

Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. Nat. Biotechnol. *36*, 765–771.

Kuscu, C., Parlak, M., Tufan, T., Yang, J., Szlachta, K., Wei, X., Mammadov, R., and Adli, M. (2017). CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. Nat. Methods *14*, 710–712.

Kweon, J., Jang, A.-H., Shin, H.R., See, J.-E., Lee, W., Lee, J.W., Chang, S., Kim, K., and Kim, Y. (2020). A CRISPR-based base-editing screen for the functional assessment of BRCA1 variants. Oncogene *39*, 30–35.

Lee, C., Banerjee, T., Gillespie, J., Ceravolo, A., Parvinsmith, M.R., Starita, L.M., Fields, S., Toland, A.E., and Parvin, J.D. (2015). Functional Analysis of BARD1 Missense Variants in Homology-Directed Repair of DNA Double Strand Breaks. Hum. Mutat. *36*, 1205–1214.

Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol. *15*, 554.

Litman, R., Peng, M., Jin, Z., Zhang, F., Zhang, J., Powell, S., Andreassen, P.R., and Cantor, S.B. (2005). BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCJ. Cancer Cell *8*, 255–265.

Liu, L.F., Desai, S.D., Li, T.K., Mao, Y., Sun, M., and Sim, S.P. (2000). Mechanism of action of camptothecin. Ann. N Y Acad. Sci. *922*, 1–10.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. *47* (W1), W636–W641.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal *17*, 10–12.

McKinnon, P.J. (2009). DNA repair deficiency and neurological disease. Nat. Rev. Neurosci. *10*, 100–112.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O.O., Gootenberg, J.S., Mori, H., et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science *361*, 1259–1262.

Noordermeer, S.M., Adam, S., Setiaputra, D., Barazas, M., Pettitt, S.J., Ling, A.K., Olivieri, M., Álvarez-Quilón, A., Moatti, N., Zimmermann, M., et al. (2018). The shieldin complex mediates 53BP1-dependent DNA repair. Nature *560*, 117–121.

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

Park, E.-S., Choi, S., Kim, J.-M., Jeong, Y., Choe, J., Park, C.-S., Choi, Y., and Rho, J. (2007). Early embryonic lethality caused by targeted disruption of the TRAF-interacting protein (TRIP) gene. Biochem. Biophys. Res. Commun. *363*, 971–977.

Rao, Q., Liu, M., Tian, Y., Wu, Z., Hao, Y., Song, L., Qin, Z., Ding, C., Wang, H.-W., Wang, J., and Xu, Y. (2018). Cryo-EM structure of human ATR-ATRIP complex. Cell Res. *28*, 143–156.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. *42*, W320-4.

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

Schlacher, K., Christ, N., Siaud, N., Egashira, A., Wu, H., and Jasin, M. (2011). Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. Cell *145*, 529–542.

Setiaputra, D., and Durocher, D. (2019). Shieldin - the protector of DNA ends. EMBO Rep. *20*, 122.

Sonneville, R., Bhowmick, R., Hoffmann, S., Mailand, N., Hickson, I.D., and Labib, K. (2019). TRAIP drives replisome disassembly and mitotic DNA repair synthesis at sites of incomplete DNA replication. eLife *8*, E5765.

Soo Lee, N., Jin Chung, H., Kim, H.-J., Yun Lee, S., Ji, J.-H., Seo, Y., Hun Han, S., Choi, M., Yun, M., Lee, S.-G., et al. (2016). TRAIP/RNF206 is required for recruitment of RAP80 to sites of DNA damage. Nat. Commun. *7*, 10463.

Sowa, M.E., Bennett, E.J., Gygi, S.P., and Harper, J.W. (2009). Defining the human deubiquitinating enzyme interaction landscape. Cell *138*, 389–403.

Takai, H., Naka, K., Okada, Y., Watanabe, M., Harada, N., Saito, S., Anderson, C.W., Appella, E., Nakanishi, M., Suzuki, H., et al. (2002). Chk2-deficient mice exhibit radioresistance and defective p53-mediated transcription. EMBO J. *21*, 5195–5205.

Tan, J., Zhang, F., Karcher, D., and Bock, R. (2019). Engineering of high-precision base editors for site-specific single nucleotide replacement. Nat. Commun. *10*, 439.

Walton, R.T., Christie, K.A., Whittaker, M.N., and Kleinstiver, B.P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science *368*, 290–296.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.

Wang, N., Ding, H., Liu, C., Li, X., Wei, L., Yu, J., Liu, M., Ying, M., Gao, W., Jiang, H., and Wang, Y. (2015). A novel recurrent CHEK2 Y390C mutation identified in high-risk Chinese breast cancer patients impairs its activity and is associated with increased breast cancer risk. Oncogene *34*, 5198–5205.

Wang, H., Peng, B., Pandita, R.K., Engler, D.A., Matsunami, R.K., Xu, X., Hegde, P.M., Butler, B.E., Pandita, T.K., Mitra, S., et al. (2017). Aurora kinase B dependent phosphorylation of 53BP1 is required for resolving merotelic kinetochore-microtubule attachment errors during mitosis. Oncotarget *8*, 48671–48687.

Wiltshire, T., Ducy, M., Foo, T.K., Hu, C., Lee, K.Y., Belur Nagaraj, A., Rodrigue, A., Gomes, T.T., Simard, J., Monteiro, A.N.A., et al. (2020). Functional characterization of 84 PALB2 variants of uncertain significance. Genet. Med. *22*, 622–632.

Wu, R.A., Semlow, D.R., Kamimae-Lanning, A.N., Kochenova, O.V., Chistol, G., Hodskinson, M.R., Amunugama, R., Sparks, J.L., Wang, M., Deng, L., et al. (2019). TRAIP is a master regulator of DNA interstrand crosslink repair. Nature *567*, 267–272.

Yeh, W.-H., Chiang, H., Rees, H.A., Edge, A.S.B., and Liu, D.R. (2018). In vivo base editing of post-mitotic sensory cells. Nat. Commun. *9*, 2184.

Zafra, M.P., Schatoff, E.M., Katti, A., Foronda, M., Breinig, M., Schweitzer, A.Y., Simon, A., Han, T., Goswami, S., Montgomery, E., et al. (2018). Optimized base editors enable efficient editing in cells, organoids and mice. Nat. Biotechnol. *36*, 888–893.

Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics *30*, 614–620.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-Flag	Sigma-Aldrich	Cat#F1804; RRID:AB_262044
Mouse monoclonal anti-HA	Sigma-Aldrich	Cat#H3663; RRID:AB_262051
Rabbit polyclonal anti-USP28	Bethyl Laboratories	Cat#A300-898A; RRID:AB_2212916
Mouse monoclonal anti-p53	Santa Cruz Biotechnology	Cat#sc-126; RRID:AB_628082
Sheep polyclonal anti-TRAIP	a gift from Niels Mailand	N/A
Rabbit polyclonal anti-phosphoRPA2$^{S4/8}$	Bethyl Laboratories	Cat#A300-245A; RRID:AB_210547
Rabbit polyclonal anti-RPA2	Bethyl Laboratories	Cat#A300-244A; RRID:AB_185548
Rabbit polyclonal anti-γ-H2AX	Bethyl Laboratories	Cat#A300-081A; RRID:AB_203288
Rabbit polyclonal anti-H2AX	Bethyl Laboratories	Cat#A300-082A; RRID:AB_203287
Rabbit monoclonal anti-ATM	Cell signaling	Cat#2873; RRID:AB_2062659
Rabbit monoclonal anti-phosphoCHK2^{T68}	Cell signaling	Cat#2197; RRID:AB_2080501
Mouse monoclonal anti-CHK2	Cell signaling	Cat#3440; RRID:AB_2229490
Mouse monoclonal anti-vinculin	Sigma-Aldrich	Cat#V9131; RRID:AB_477629
Rabbit polyclonal anti-53BP1	Bethyl laboratories	Cat#A300-272A; RRID:AB_185520
Rabbit polyclonal anti-RAD51	Bioacademia	Cat#70-002; RRID:AB_1056187
Mouse monoclonal anti-BRCA1	Santa Cruz Biotechnology	Cat#sc-6954; RRID:AB_626761
Mouse monoclonal anti-γ-H2AX	BioLegend	Cat#613402; RRID:AB_315795
Mouse monoclonal anti-cyclin A	Santa Cruz Biotechnologies	Cat#sc-271682; RRID:AB_10709300
Goat anti-mouse Alexa Fluor 488	Thermo Fisher Scientific	Cat#A11001; RRID:AB_2534069
Goat anti-rabbit Alexa Fluor 488	Thermo Fisher Scientific	Cat#A11008; RRID:AB_143165
Goat anti-mouse Alexa Fluor 594	Thermo Fisher Scientific	Cat#A11005; RRID:AB_141372
Goat anti-rabbit Alexa Fluor 594	Thermo Fisher Scientific	Cat#A11012; RRID:AB_141359
Bacterial and Virus Strains		
Subcloning Efficiency DH5α	Thermo Fisher Scientific	Cat#18265-017
Endura Electrocompetent Cells	Lucigen	Cat#60242-1
Chemicals, Peptides, and Recombinant Proteins		
Transfection reagent: TransIT-293	Mirus	Cat#MIR 2700
Transfection reagent: Lipofectamine RNAiMAX	Thermo Fisher Scientific	Cat#13778-150
Puromycin	Gold Biotechnology	Cat#P-600-100
Blasticidin	Gold Biotechnology	Cat#B-800-100
Cisplatin	Sigma-Aldrich	Cat#P4394-25MG
Olaparib (AZD2281, Ku-0059436)	Selleck Chemicals	Cat#S1060
Doxorubicin hydrochloride	TCI Chemicals	Cat#D4193-25MG
Camptothecin	Sigma-Aldrich	Cat#C9911-100MG
(±)-Nutlin-3	Cayman Chemicals	Cat#1004372
Gibson Assembly Master Mix	New England Biolabs	Cat#E2611L
Q5 High-Fidelity DNA polymerase	New England Biolabs	Cat#M0491L
Anti-HA agarose beads	Sigma-Aldrich	Cat#A2095; RRID:AB_257974
Critical Commercial Assays		
Quick-DNA Midiprep Plus Kit	Zymo Research	Cat#D4075
Quick Extract DNA Extraction solution	Lucigen	Cat#QE09050
Deposited Data		
Unprocessed gels, blots and microscopy images	This study; Mendeley Data	https://dx.doi.org/10.17632/kfyr3t5v9w.1
sgRNA NGS data	SRA	SRA: SUB8818564; BioProject: PRJNA689217

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Human reference genome NCBI build 38, GRCh38	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
Structure of 53BP1 TTD bound to H4K20me2	(Botuyan et al., 2006)	PDB: 2IG0
Structure of the ATM dimer	(Baretić et al., 2017)	PDB: 5NP0
Structure of the CHK2 dimer	(Cai et al., 2009)	PDB: 3I6W
Structure of ATR bound to ATRIP	(Rao et al., 2018)	PDB: 5YZ0
The ClinVar database (v. 2018-06-03)	NCBI	https://www.ncbi.nlm.nih.gov/clinvar/; RRID:SCR_006169
The Cancer Genome Atlas database (v. gdc-1.0.0)	GDC	https://portal.gdc.cancer.gov/; RRID:SCR_003193
Experimental Models: Cell Lines		
MCF10A	ATCC	Cat#CRL-10317; RRID:CVCL_0598
MCF7	ATCC	Cat#HTB-22; RRID:CVCL_0031
HAP1	A gift from Thijn Brummelkamp	N/A
HEK293T	ATCC	Cat#CRL-11268; RRID:CVCL_1926
Oligonucleotides		
List of oligonucleotides	This study, Table S7	N/A
Recombinant DNA		
Plasmid: BE3-FNLS-P2A-BlastR	(Billon et al., 2017)	N/A
Plasmid: pLenti-Guide-Puro	(Sanjana et al., 2014)	Addgene Cat#52963
Plasmid: LentiGuide-NLS-GFP	(Noordermeer et al., 2018)	N/A
Plasmid: LentiGuide-NLS-mCherry-*AAVS1*	(Hustedt et al., 2019)	N/A
Plasmids: pLenti-Guide-Puro-sgRNA-of-interest	This study, Table S7	N/A
Plasmids: LentiGuide-NLS-GFP-sgRNA-of-interest	This study, Table S7	N/A
Plasmid: pDONR223-53BP1	This study	N/A
Plasmid: pMSCV-FLAG-HA-DEST	(Sowa et al., 2009)	N/A
Plasmid: pMSCV-FLAG-HA-53BP1 WT and mutants	This study	N/A
Plasmid: pMSCV-FLAG-HA-GFP	This study	N/A
sgRNA Library: 27-gene-library in pLenti-Guide-Puro	This study	N/A
sgRNA Library: 56-gene-library in pLenti-Guide-Puro	This study	N/A
Software and Algorithms		
R (3.4.1; 3.5.0)	The R project for statistical computing	https://www.r-project.org; RRID:SCR_001905
RStudio	RStudio	https://rstudio.com; RRID:SCR_000432
PEAR	(Zhang et al., 2014)	https://cme.h-its.org/exelixis/web/software/pear/; RRID:SCR_003776
Cutadapt	(Martin, 2011)	https://cutadapt.readthedocs.io/en/stable/; RRID:SCR_011841
ANNOVAR	(Wang et al., 2010)	https://annovar.openbioinformatics.org/en/latest/; RRID:SCR_012821
MAGeCK	(Li et al., 2014)	https://sourceforge.net/p/mageck/wiki/Home/
Snapgene Viewer	Snapgene	https://www.snapgene.com/snapgene_viewer/; RRID:SCR_015053
Prism	GraphPad Software	https://www.graphpad.com/scientific-software/prism/; RRID:SCR_002798
ImageJ	National Institutes of Health	https://imagej.nih.gov/ij; RRID:SCR_003070
CometScore Software Version 1.5	TriTek Corporation	http://rexhoover.com/index.php?id=cometscore
MetaXpress High-Content Image Acquisition and Analysis Software	Molecular Devices	https://www.moleculardevices.com/products/cellular-imaging-systems/acquisition-and-analysis-software/metaxpress#gref; RRID:SCR_016654

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Inference of CRISPR Editing (ICE)	Synthego	https://ice.synthego.com/#/
PyMOL	Delano Scientific	https://pymol.org/2/; RRID:SCR_000305
Phyre2	(Kelley et al., 2015)	http://www.sbg.bio.ic.ac.uk/~phyre2; RRID:SCR_010270
Missense3D	(Ittisoponpisan et al., 2019)	http://www.sbg.bio.ic.ac.uk/~missense3d/
Consurf server	(Ashkenazy et al., 2016)	https://consurf.tau.ac.il/
ESPript3.0	(Robert and Gouet, 2014)	http://espript.ibcp.fr/ESPript/ESPript; RRID:SCR_006587
Other		
Base editing screen results in MCF10A and MCF7	This study, web resource	https://www.ciccialab-database.com/ddr-variants

RESOURCE AVAILABILITY

Lead contact

Further information and request for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Alberto Ciccia (ac3685@cumc.columbia.edu).

Materials availability

Plasmids and cell lines generated in this study will be made available by the Lead Contact under a Material Transfer Agreement.

Data and code availability

All screening data are provided as Supplemental Tables and can be additionally enquired on a dedicated website available at the following address: https://www.ciccialab-database.com/ddr-variants.

The raw NGS data generated during this study are publicly available at the Single Read Archive (SRA). The accession numbers for the NGS data are SRA: SUB8818564; BioProject: PRJNA689217. Original gel, blots and microscopy images have been deposited to Mendeley Data: https://dx.doi.org/10.17632/kfyr3t5v9w.1. Scripts utilized for analysis and graph generation will be made available upon request.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Cell lines and cell culture

Female human mammary epithelial MCF10A cells were cultured in a high-glucose, GlutaMAX containing 1:1 mixture of DMEM and Ham's F12 medium (GIBCO, Thermo Fisher Scientific, Waltham, MA), supplemented with 5% (v/v) horse serum (GIBCO, Thermo Fisher Scientific), 10 mg/mL insulin (Sigma-Aldrich, St Louis, MO), 0.5 mg/mL hydrocortisone (Sigma-Aldrich), 20 ng/mL human epidermal growth factor (PeproTech, Rocky Hill, NJ) and 100 U/mL penicillin and 100 μg/mL streptomycin (Pen/Strep, Thermo Fisher Scientific). Female breast cancer MCF7 cells were cultured in 2 mM L-glutamine containing DMEM (GIBCO, Thermo Fisher Scientific) supplemented with 10% (v/v) fetal bovine growth serum (GIBCO, Thermo Fisher Scientific), and Pen/Strep (Thermo Fisher Scientific). HAP1 cells were cultured in Iscove Modified Dulbecco's Medium (IMDM, Caisson Labs, Smithfield, UT) supplemented with 10% (v/v) Fetalgro bovine growth serum (BGS, RMBIO, Missoula, MT), 2 mM GlutaMAX (Thermo Fisher Scientific), and Pen/Strep (Thermo Fisher Scientific). Female HEK293T cells were cultured in high-glucose, 2 mM L-glutamine containing DMEM (GIBCO, Thermo Fisher Scientific) supplemented with 10% (v/v) Fetalgro bovine growth serum (BGS, RMBIO). All cells were grown in a humidified atmosphere at 37°C and 5% CO_2. To generate cells constitutively expressing BE3-FNLS-P2A-BlastR (Billon et al., 2017; Zafra et al., 2018), lentiviral transduction was carried out. Briefly, $2x10^6$ HEK293T cells were seeded in 6 cm dishes in antibiotic-free DMEM. Sixteen h later, cells were co-transfected with 4 μg of the BE3-FNLS-P2A-BlastR and the vectors encoding the lentiviral packaging components gag-pol, VSV-G, rev and tat (0.2 μg, 0.4 μg, 0.2 μg, and 0.2 μg, respectively). Vectors were mixed in 500 μL of Opti-MEM (GIBCO, Thermo Fisher Scientific) and 18 μL of TransIT-293 reagent (Mirus Bio, Madison, WI) were added, mixed, incubated for 20 min at room temperature (RT). The mixture was then added dropwise to HEK293T cells. Forty-eight and 72 h post-transfection, viral supernatants were collected, cell residues eliminated by filtration through a 0.45 μm filter and applied in a 1:1 virus/medium mixture over the cell line of interest growing at 50% confluency with 4 μg/mL polybrene for 24 h. After transduction, cells were treated with blasticidin at 10 μg/mL for 72 h. For clone recovery, cells were seeded at low density in 10 cm dishes (MCF10A, MCF7) or at a density of 0.5 cells per well in 96-well plates (HAP1) and outgrown until populations were recovered. BE3-FNLS expression was determined by western blot.

METHOD DETAILS

Plasmids

BE3-FNLS-P2A-BlastR was previously generated (Billon et al., 2017) and pLenti-Guide-Puro (Addgene #52963, (Sanjana et al., 2014)) were utilized for constitutive BE3 and sgRNA expression. For characterization of individual sgRNA-induced perturbations, sgRNAs in Table S7 were cloned into the modified form of LentiCRISPRv2 that replaced Cas9 with NLS-tagged GFP (referred as LentiGuide-NLS-GFP) from (Noordermeer et al., 2018), as previously described (Sanjana et al., 2014). LentiGuide-NLS-mCherry, and LentiGuide-NLS-GFP carrying a sgRNA targeting the *AAVS1* locus were used as controls (Hustedt et al., 2019). 53BP1 full-length ORF (NM_005657) was cloned into pDONR223 entry vector with BP clonase II (Life Technologies). A stop codon and mutations in 53BP1 were introduced by site-directed mutagenesis (inverse PCR) and verified by Sanger sequencing. pDONR223-53BP1 WT, its mutants or a GFP control were subsequently cloned into pMSCV-FLAG-HA-DEST (Sowa et al., 2009) with LR clonase II (Life Technologies).

Library design and cloning

Custom DNA libraries were designed to perform mutational tiling in 86 target genes of the DNA damage response (Table S1). To this end, all possible 20-mer sequences upstream of NGG PAM sites where the BE3 editing window (13-18 nucleotides from the PAM (Zafra et al., 2018)) would totally or partially fall within a coding exon in any described gene isoform were annotated in the hg38 human genome assembly. Three types of negative controls were included: i) sgRNAs with targetable bases within the *AAVS1* locus, ii) non-targeting sgRNAs obtained from the GeCKO_v2 library, and iii) sgRNAs targeting common essential genes from (Hart et al., 2015) containing no editable base within an extended, 8-bp long BE3 window (aforementioned 6-bp window +1 nt at both 3′ and 5′ ends). As positive controls for cell lethality, stop codons in common essential genes were designed using the iSTOP search parameters (Billon et al., 2017), and randomly sampled to include at least 1,000 sgRNAs in the final libraries. Duplicated sgRNAs and sgRNAs containing BsmBI (CGTCTC) restriction sites were excluded. Two sublibraries were designed (sublibrary 1 and sublibrary 2) and combined in the current screen analyses (library) (Table S1). Oligonucleotides were designed including BsmBI restriction sites for cloning into the pLentiGuide-Puro expression plasmid, and primer sites were appended for specific library amplification within mixed oligonucleotide pools. Final oligonucleotides for sublibrary 1 were designed for Gibson assembly-based cloning and looked as follows: 5′[Forward primer]cgtctcgtcatgtggaaaggacgaaacaccg[sgRNA 20-mer]gttttagagctagaaatagctgcacgagacg[Reverse primer]3′. Oligonucleotides for sublibrary 2 were designed for Golden Gate-based cloning as two sgRNA concatemers separated by Esp3I sites allowing for the decoupling of individual sgRNAs. Oligonucleotide pools were ordered from Agilent Technologies (Santa Clara, CA). Libraries were amplified using 1 ng of starting oligonucleotide pool and Q5 high-fidelity DNA polymerase (New England Biolabs, Ipswich, MA) in a 25 μL reaction for every 5,000 sgRNAs and a maximum of 20 PCR cycles. Amplified custom designed sublibraries were cloned as follows. Sublibrary 1 was digested with Esp3I (New England Biolabs) prior to cloning into pLenti-Guide-Puro using a Gibson assembly strategy as per manufacturer instructions, and the protocol described in (Joung et al., 2017). Sublibrary 2 was cloned into pLenti-Guide-Puro using a modified version of the Golden Gate assembly protocol. A combined Esp3I (New England Biolabs) digestion-assembly step was carried out using the Golden Gate technology as per manufacturer's indications. Upon isopropanol precipitation, libraries were transformed into Endura electrocompetent cells (Lucigen). Plasmid DNA libraries were isolated using ZymoPure II Plasmid Maxiprep Kit (Zymo Research, Irvine, CA). Prior to use, libraries were sequenced on a NovaSeq 6000 sequencing platform (Illumina, San Diego, CA) by the Genome Sciences Facility at The Pennsylvania State College of Medicine.

BE3 base editing screens

All screens were performed in biological triplicates and library coverage of at least 1,000 cells per sgRNA was preserved at every step. MCF10A-BE3, MCF7-BE3 and HAP1-BE3 cells were transduced with three independent batches of custom lentiviral sublibraries 1 and 2 at a low MOI (< 0.4) for 24 h. Puromycin-containing medium was then added and selection maintained for 48 h at 2 μg/mL. Subsequently, cells were cultured in medium containing 0.5 μg/mL puromycin and 5 μg/mL blasticidin for the duration of the screen. At this stage, which was considered time 0 (T0), cells were subcultured and medium with and without drugs refreshed every 3 days with collection time established at day 18 (T18). Collection times were optimized performing smaller screens in biological triplicates using sublibrary 1 and assessing the behavior of iSTOP controls and negative controls. Drug concentrations were optimized to achieve a lethal dose 25 (LD25) by previously culturing sublibrary 1-transduced MCF10A-BE3 and MCF7-BE3 cells in 6-well plates for 18 days and passaging them into refreshed drug medium every 3 days. Drugs were used as follows: i) cisplatin (Sigma-Aldrich, #P4394) at 1 μM and 750 nM, ii) olaparib (Selleck Chemicals, Houston, TX, USA, #S1060) at 2 μM and 1 μM, iii) doxorubicin (TCI Chemicals, Tokyo, Japan, #D4193) at 2.5 nM and 5 nM, iv) camptothecin (Sigma-Aldrich, #C9911) at 5 nM and 2.5 nM, for MCF10A-BE3 and MCF7-BE3, respectively. Cell pellets were frozen at T0 and T18 and subsequent gDNA isolation was performed using Quick-DNA Midiprep Plus Kit (Zymo Research). Genome-integrated sgRNA sequences were amplified with Q5 high-fidelity DNA polymerase (New England Biolabs). Briefly, 10 μg of genomic DNA were amplified in a 100 μL reaction utilizing a 5-staggered forward primer mix and a unique reverse primer and including Q5 High GC Enhancer as per manufacturer indications (Table S7). sgRNA representation was preserved considering that the genome of 10^6 diploid cells weights ~6.6 μg. PCR amplification was carried out as follows: 30 s at 98°C; followed by 10 s at 98°C, 30 s at 58°C, 10 s at 72°C for 18 cycles; and 2 min at 72°C. Samples were

barcoded in a second PCR step using 5 μL of the first PCR, custom primers containing TruSeq Illumina dual indexes (UDI) and Q5 High GC Enhancer, as per manufacturer indications (Table S7). The following PCR conditions were used: 30 s at 98°C; followed by 10 s at 98°C, 30 s at 68°C, 15 s at 72°C for 10 cycles; and 2 min at 72°C. sgRNA representation was preserved by running a barcoding reaction for every 5,000 sgRNAs. Final products were run on 2% agarose gels, gel purified, multiplexed and sequenced on a NovaSeq 6000 sequencing platform (Illumina) with a 10% spike-in of PhiX by the Princeton University Genomics Core Facility.

Variant annotation

All the possible mutational outcomes for a given sgRNA were computed by permutation of the editable bases within the BE3 window (13-18 nucleotides from the PAM (Zafra et al., 2018)). Base substitutions were translated into the corresponding protein mutations in all protein isoforms using Variant Effect Predictor (Ensembl version 93), and categorized based on their outcome as "synonymous," "missense," "nonsense," or "splice" variants. sgRNAs targeting exonic regions in putative non-protein coding isoforms were uniquely categorized as "putative non-coding" regardless of the mutational outcome. Each sgRNA was placed in a unique category, defined by the most damaging mutation possibly made within its editable window in any of the protein isoforms (e.g., if a sgRNA makes a nonsense mutation in one isoform and a synonymous mutation in another isoform, it will be categorized as "nonsense"). Mutations in the first codon were classified as "nonsense." sgRNAs with no editable bases in the window were categorized as "empty-window." To determine the clinical relevance, mutations were parsed against the *ClinVar* database using ANNOVAR (version 2018Apr16) (Wang et al., 2010). For simplicity, three categories were established: i) benign/likely-benign, ii) VUS, including variants of uncertain significance and variants with conflicting interpretations, and iii) pathogenic/likely-pathogenic. When multiple variants were generated by the same sgRNA, the clinical interpretation of the most deleterious amino acid change was assigned (e.g., if a given sgRNA is predicted to insert a missense mutation listed in *ClinVar* as a VUS and a nonsense mutation absent from *ClinVar*, then the sgRNA would not be classified as clinically-relevant). When all predicted sgRNA mutational outcomes were equally deleterious at the functional level, the clinical relevance of the sgRNA was determined based on the following criteria applied to the mutational outcomes: benign/likely-benign < absent in *ClinVar* (NA) < VUS < pathogenic/likely-pathogenic. Predicted mutations were also mapped to The Cancer Genome Atlas (TCGA) database (version gdc-1.0.0). TCGA somatic mutation data of 33 cancer types were downloaded from GDC (https://portal.gdc.cancer.gov/). sgRNAs were called as present in TCGA when a mutation in the TCGA database coincided with any of the possible mutations that sgRNA could introduce, regardless of the presence of bystander mutations. Post-translational modification sites were mapped against the PhosphoSitePlus database utilizing predicted amino acid substitutions in the canonical isoform for each gene.

Two-color competitive growth assays

Cells were transduced with lentiviral particles of LentiGuide-NLS-mCherry-*AAVS1* or LentiGuide-NLS-GFP-sgRNA-of-interest (Hustedt et al., 2019; Noordermeer et al., 2018), at high MOI (> 1). Twenty-four h post-infection, puromycin was added at 2 μg/mL for 24 h in MCF10A populations expressing BE3, and for 48 h in both MCF10A-BE3 and MCF7-BE3 to select for transductants. GFP- and mCherry-expressing cells were subsequently mixed in a 1:1 ratio to seed on black, clear-bottom 96-well plates at a density of 5,000 cells per well for experiments using MCF10A cells and 8,000 cells per wells for experiments on MCF7 cells. At this stage, considered day 0, cells were treated with drugs at the concentrations specified in the screen or left untreated. Cells were subcultured and medium with and without drugs refreshed every 4 days. Cells were imaged on day 1 to evaluate plating ratios and on days 4, 8, 12, 16 and 20. Image acquisition was performed with an ImageXpress Nano Automated Imaging System microscope (Molecular Devices, San Jose, CA) with a 10X objective. Image analyses for segmentation and counting of GFP and mCherry positive cells was performed using the MetaXpress imaging software. Each experiment was performed in at least two biological replicates (independent transductions), and each condition assessed in technical duplicates. Unless otherwise indicated in the figure legends, data are represented as the sgRNA-of-interest-GFP/*AAVS1*-sgRNA-mCherry ratio normalized to T1 and to the corresponding *AAVS1*-sgRNA-GFP/*AAVS1*-sgRNA-mCherry ratio at each experimental point.

Cell line genotyping

Evaluation of the effective mutational outcome was performed by PCR amplification followed by Inference of CRISPR Editing (ICE) analyses. Briefly, cells were transduced with sgRNAs targeting the genomic loci of interest and the *AAVS1* locus control as described above, collected at T4, T12 and T20 or T4 only, and genomic DNA extracted using Quick-Extract (Lucigen, Middleton, WI, USA) as per manufacturer indications. Genomic loci containing the corresponding sgRNA-targeted sequences were PCR amplified using the primer pairs in Table S7 and Sanger sequenced. Genomic DNA from *AAVS1*-targeted cells was used to obtain WT reference sequences. Sequencing traces were enquired for predicted in-window and out-of-window single-base substitutions using the ICE online tool (Synthego, 2019, v2.0 (Hsiau et al., 2019)).

Immunoblotting

Cells were collected, washed and resuspended in 0.375 volumes of PBS and mixed with 0.375 volumes of sample buffer (0.1M Tris pH 6.8, 4% SDS, 12% β-mercaptoethanol). Subsequently, 4x NuPage LDS sample buffer (Thermo Fisher Scientific) was added and samples were mixed by vortexing for 15 s and boiled at 95°C for 10 min. A second pulse of vortexing and boiling was performed prior to subjecting equivalent protein amounts to gel electrophoresis and proteins were transferred onto nitrocellulose membranes.

Proteins were detected using the appropriate primary and HRP-conjugated secondary antibodies at a 1:10,000 dilution. Primary antibodies used in this study include mouse anti-FLAG (Sigma-Aldrich F1804, 1:2,000), mouse anti-HA (Sigma-Aldrich H3663, 1:5,000), rabbit anti-USP28 (Bethyl Laboratories A300-898A, 1:1,000), mouse anti-p53 (DO-1, Santa Cruz Biotechnology sc-126, 1:5,000), sheep anti-TRAIP (a gift from Niels Mailand, 1:500), rabbit anti-phosphoRPA2$^{S4/8}$ (Bethyl Laboratories A300-083A, 1:10,000), rabbit anti-RPA2 (Bethyl Laboratories A300-244A, 1:5,000) rabbit anti-γ-H2AX (Bethyl Laboratories A300-081A, 1:10,000), rabbit H2AX (Bethyl Laboratories A300-081A, 1:5,000), rabbit anti-ATM (Cell signaling #2873, 1:1,000), rabbit anti-phosphoCHK2^{T68} (Cell signaling #2197, 1:1,000), mouse anti-CHK2 (Cell signaling #3440, 1:500), and mouse anti-vinculin (Sigma-Aldrich V9131, 1:25,000).

Co-immunoprecipitation

HEK293T cells seeded in 10 cm dishes were transfected with 10 μg of pMSCV-FLAG-HA-53BP1 WT, mutant or a GFP control using TransIT-293 (Mirus Bio) as per manufacturer's indications. Three days after transfection, cells were harvested in PBS and resuspended in 750 μl of mammalian cell lysis buffer (MCLB; 50 mM Tris-HCl pH 7.5, 1% NP40) supplemented with 150 mM NaCl and protease and phosphatase inhibitor cocktails (GB-331 and GB-450, Goldbio, St Louis, MO). Following end-over-end rotation for 30 min at 4°C, cell lysates were cleared by centrifugation and the low-salt supernatant collected. Cell pellets were then resuspended in 250 μl of MCLB supplemented with 500 mM NaCl and protease and phosphatase inhibitor cocktails and incubated for 1 h at 4°C with end-over-end rotation. After centrifugation, the salt concentration of the high-salt supernatant was adjusted to 150 mM NaCl with MCLB and combined with the low-salt supernatant. The combined lysates were then incubated with anti-HA agarose beads (A2095, Sigma-Aldrich) for 4 h at 4°C with end-over-end rotation. Protein-bound beads were washed five times in buffer (MCLB supplemented with 150 mM NaCl) and bound proteins subsequently eluted by boiling in 1x LDS sample buffer (NP0007, Life Technologies) supplemented with 5% (v/v) β-mercaptoethanol.

RNA interference

Cells were subjected to reverse siRNA transfection utilizing of firefly (FF) siRNA or BRCA1 siRNA at 20 nM and lipofectamine RNAiMAX (Thermo Fisher Scientific) as per manufacturer's indications. Twenty-four h after siRNA transfection, MCF10A cells were seeded on black 96-well bottom-glass plates and incubated for 2 additional days prior to immunofluorescence analyses.

High-content imaging

Cells were seeded on black, clear-bottom 96-well plates at a density of 6,000 cells per well for overnight growth. Subsequently, cells were either mock-treated, treated with drugs as specified in every experiment, or treated with 5 Gy ionizing radiation for 4 h using a JL Shepherd Mark I cesium irradiator (JL Shepherd & Associate). Cells were fixed and permeabilized using a solution of 2% (v/v) paraformaldehyde and 0.5% (v/v) Triton X-100 for 10 min. After extensive PBS washing, cells were incubated in blocking buffer (3% (w/v) BSA in TBS- 0.1% (v/v) Tween20) for 1 h. Incubation with corresponding primary antibodies in blocking buffer, i.e., rabbit anti-53BP1 (Bethyl laboratories A300-272A, 1:2,000), rabbit anti-RAD51 (Bioacademia 70-002, 1:10,000), anti-BRCA1 (Santa Cruz Biotechnology sc-6954, 1:100), mouse anti-γ-H2AX (BioLegend #613402, 1:5,000) mouse anti-cyclin A (Santa Cruz Biotechnologies sc-271682, 1:1000) was performed overnight at 4°C. Cells were washed with TBS-T, and incubated with anti-rabbit and anti-mouse Alexa Fluor 488- and 594-labeled secondary antibody (Thermo Fisher Scientific) at a 1:1,000 dilution. Upon TBS-T washing, cell nuclei were counterstained with DAPI. Image acquisition was performed with an ImageXpress Nano Automated Imaging System microscope (Molecular Devices, San Jose, CA) with a 40X Plan APO objective. Image analyses for segmentation and counting 53BP1 foci, RAD51 foci, BRCA1 foci, 53BP1 NBs were performed using MetaXpress imaging software. Unless otherwise indicated, data represented include a minimum of 5,000 cells from two biological replicates. For nuclear aberrations/micronuclei analyses, at least 400 cells belonging to a minimum of 5 different DAPI images were blind-counted per biological replicate.

RT-PCR

RNA extractions were performed using the Quick-RNA microprep kit (Zymo Research) as per manufacturer's instructions. cDNA was generated using 1 μg starting RNA and random hexamers (New England Biolabs) and MMLV high performance reverse transcriptase (Lucigen). PCR amplifications were performed using Q5 High-fidelity DNA polymerase (New England Biolabs) and primer pairs obtained from IDT as described in (Harley et al., 2016).

Neutral comet assay

DNA double-strand breaks were evaluated by neutral comet assay. MCF10A cells were plated in 12-well plates at a density of 50,000 cells/well and the following day treated with camptothecin at 10 nM for 24 h. Cells were collected by trypsinization and kept on ice prior to mixing with molten LMAgarose and pipetted onto slides coated with a single layer of 1% NMAgarose. Slides were then incubated with a lysis solution (30 mM EDTA, 0.5% SDS) for 45 min, and cells migrated in an electrophoretic chamber (Fisher Biotech) at 20 V for 20 min in TBE. Slides were extensively washed with water, and fixed with ice-cold methanol for 5 min. Prior to image acquisition, slides were stained with a fluorescent dye (GelRed, Biotium, 1:1,000 in water). Images were acquired utilizing a Nikon Eclipse 50i microscope and the amount of DNA damage assessed by calculating comet tail moment values using CometScore Software Version 1.5. Apoptotic cells (small comet head and very large comet tail) were excluded from the analysis. A minimum total of 100 cells belonging to 3 independent experiments were analyzed for each experimental point.

Protein structural modeling and conservation analyses

Partial or total protein structures for 53BP1's tandem Tudor domain (TTD) (2IG0, (Botuyan et al., 2006)), ATM (5NP0, (Baretić et al., 2017)), CHK2 (3I6W, (Cai et al., 2009)), ATR (5YZ0, (Rao et al., 2018)) were retrieved from the Protein Data Bank (PDB). BRIP1 helicase domain structure was modeled using the Phyre2 web portal with default parameters (Kelley et al., 2015). The effect of missense variants in protein structure was predicted using Missense3D (Ittisoponpisan et al., 2019). Structure graphics were generated using PyMOL (Delano Scientific LLC). Amino acid conservation analyses on 53BP1's TTD were implemented on Consurf using the aforementioned structure and default multiple sequence alignment (MSA) parameters (Ashkenazy et al., 2016). In the case of TRAIP, a multisequence alignment including sequences from the organisms depicted in Figure S4E was performed using MAFFT (Madeira et al., 2019), and visualization generated with ESPript3.0 (Robert and Gouet, 2014).

QUANTIFICATION AND STATISTICAL ANALYSES

Screen quality control and analyses

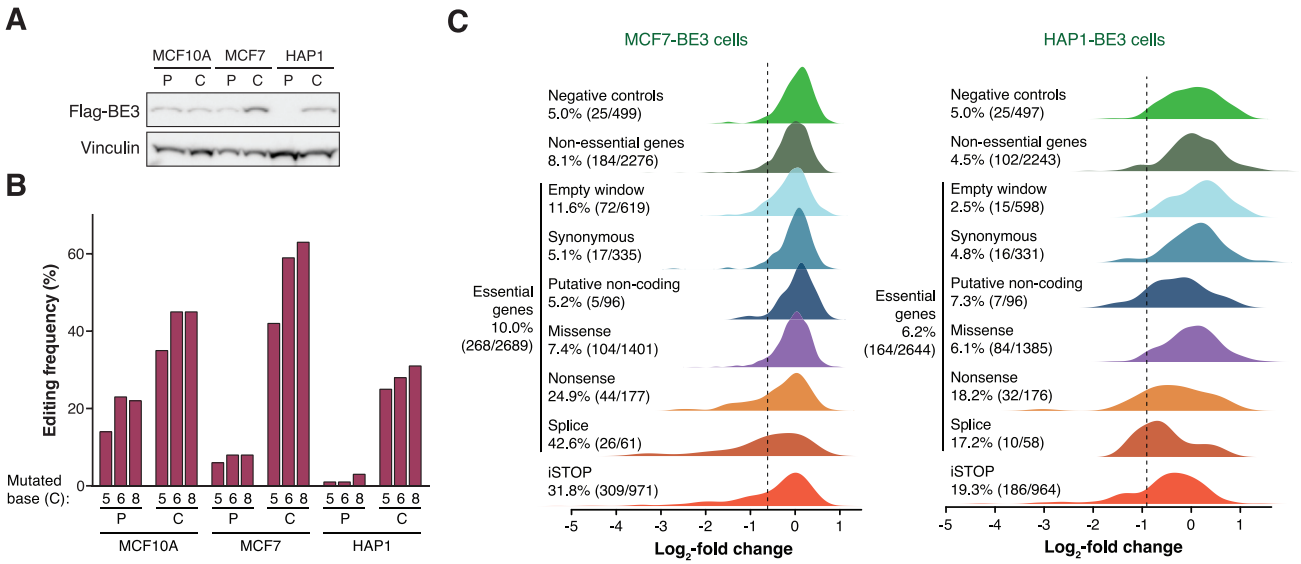
Pair-end reads were joined using PEAR to create a single sequencing read from the two paired-end fastq files (Zhang et al., 2014). After joining paired end reads, the single joined reads were trimmed at the 5′ and 3′ ends using cutadapt to remove the constant sequences flanking the sgRNA sequence (Martin, 2011). After trimming, the number of reads per sgRNA sequence was computed using the MAGeCK count command (Li et al., 2014). The MAGeCK standard count summary revealed that the percentage of mapped reads oscillated between 51% and 82.5% for all the samples. Gini indexes of read-count distribution for all conditions were on the interval between 0.035 and 0.076. Additionally, off-target determination was performed using the Genetic Perturbation Platform (GPP) portal suite for sgRNA design. 582 sgRNAs targeting more than 10 sequences with a Cutting Frequency Determination (CFD) score of 1 were eliminated from subsequent analyses. Read counts for the two sublibraries were independently normalized using the median ratio method. In order to avoid false positives, sgRNAs with low sequencing coverage (average normalized read count at T0 < 200) were eliminated from the analyses (Table S2). Comparisons of T0 versus T18 were performed using paired MAGeCK robust rank aggregation (RRA) using normalized values as input (e.g., mageck test –k Countfile.txt, -t T18_rep1, T18_rep2, T18_rep3, -c T0_rep1, T0_rep2, T0_rep3–paired–norm-method none -n T18–adjust-method fdr).

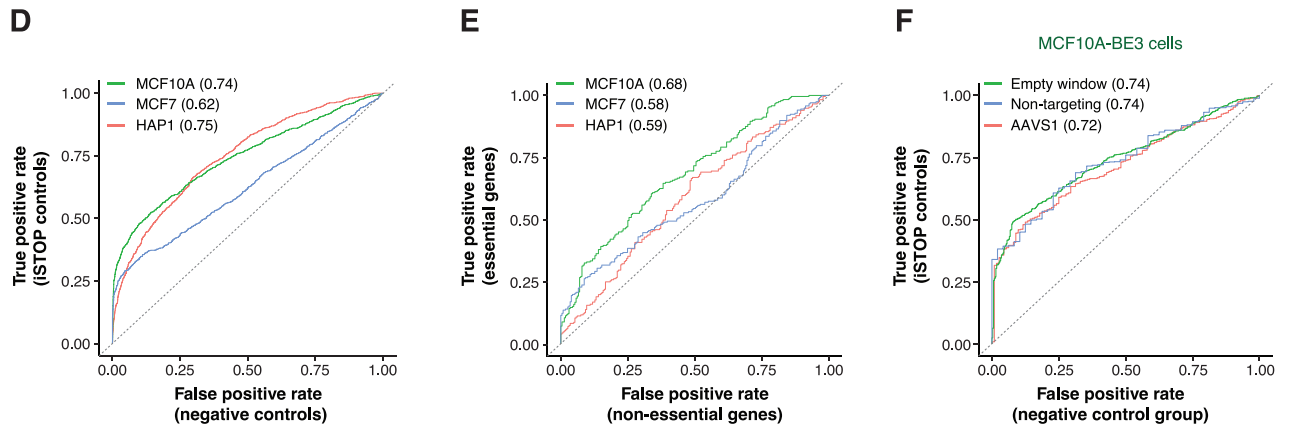
Data analyses and graphical representations

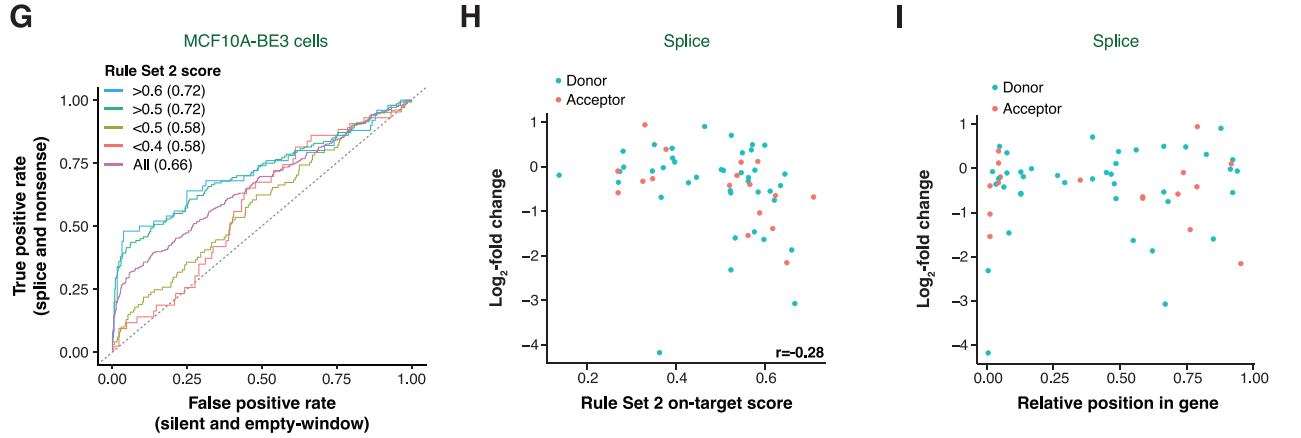
Density and rank plot analyses were performed in R version 3.5.0 (2018-04-23) and original plots were produced using the ggplot2 package. When required, sgRNA efficiency values were obtained from Rule Set 2 on-target scores using the GPP portal suite for sgRNA design (Table S3). Unless otherwise indicated, biological significance thresholds were set at the top and bottom 1% values of ranked LFC for the negative controls. For statistical significance, the threshold was set to one-sided p value < 0.01. sgRNAs meeting both criteria were considered hits, referred to as ''relevant sgRNAs'' and clustered into groups by the Ward.D algorithm using the R package pheatmap (Pretty Heatmaps v1.0.12) (parameters: clustering_distance_cols = 'euclidean', cutree_cols = 12, clustering_method = 'ward.D'). Enrichment analysis for relevant sgRNAs in each group was performed using the R function fisher.test (parameter: alternative = 'two.sided') for Fisher's exact test. Lollipop plots were performed in R version 3.5.0 (2018-04-23) using the ggplot2 package including exclusively values and annotation for sgRNAs targeting the canonical transcript. Other depicted plots and statistical analyses were generated using Prism v6 (Graphpad Software Inc, San Diego, CA, USA). Unless otherwise indicated, mean ± SD are represented, and details for statistical tests used (i.e., unpaired t test, one-way ANOVA, chi-square) are shown in the corresponding figure legends. Values given to n refer to independent biological replicates.

CellPress

Supplemental Figures







(legend on next page)

Figure S1. Quality controls for base editing screens, related to Figure 1

(A) Immunoblot showing Flag-BE3 expression in the MCF10A, MCF7 and HAP1 cells utilized for base editing screens. P: pool, C: clone.

(B) Analysis of editing frequency induced by an *AAVS1*-targeted control sgRNA at its target locus at day 4 post-selection in MCF10A, MCF7, and HAP1 cells, as determined by Sanger sequencing and ICE analyses. P: pool, C: clone.

(C) Density plots of LFC values for different categories of sgRNAs in MCF7-BE3 and HAP1-BE3, represented as in Figure 1C. Dotted line: MCF7-BE3, LFC = −0.61511, HAP1-BE3, LFC = −0.9068.

(D) ROC analyses of MAGeCK ranks for iSTOP sgRNA controls (true positives) versus negative control sgRNAs (false positives). AUC values are indicated in brackets.

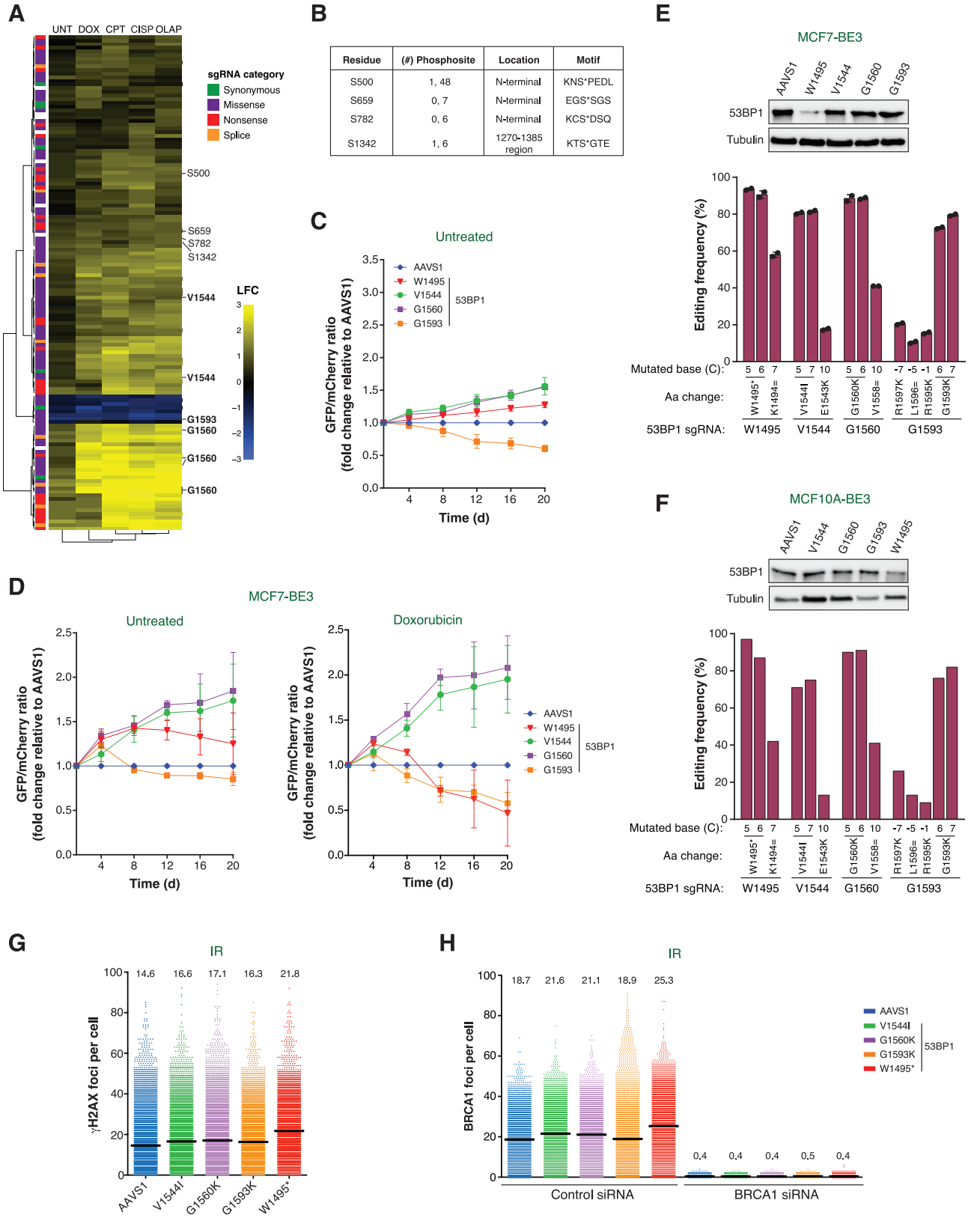
(E) ROC analyses as in (D) for sgRNAs introducing deleterious mutations in essential genes (true positives) versus non-essential genes (false positives).

(F) ROC analyses as in (D) for iSTOP sgRNA controls (true positives) versus sgRNAs of each individual negative control subgroup, i.e., empty-window, non-targeting, *AAVS1*-targeting (false positives).

(G) ROC analyses conducted as in Figure 1D after filtering deleterious and neutral sgRNAs targeting essential genes at increasing thresholds of Rule Set 2 on-target scores.

(H) Correlation of LFC values and Rule Set 2 on-target scores for sgRNAs introducing mutations in splice donor (blue) or acceptor (orange) sites in essential genes. Pearson correlation value is shown.

(I) Graphical representation of LFC values and the relative position in the gene of sgRNAs targeting splice donor (blue) and acceptor (orange) sites in essential genes.



(legend on next page)

Figure S2. Characterization of LOF and GOF mutations in 53BP1's tandem Tudor domain, related to Figure 3

(A) Heatmap of LFC values for relevant 53BP1 sgRNAs in MCF10A-BE3 cells, as shown in Figure 4A for ATM sgRNAs. sgRNAs uniquely targeting 53BP1 phosphorylation sites (regular font) and residues selected for validation (bold font) are highlighted.

(B) Table of 53BP1 phosphorylation sites uniquely targeted by relevant sgRNAs. Number of studies reporting phosphorylation at the indicated sites according to the PhosphoSite database (low-throughput, high-throughput), location in the canonical 53BP1 protein isoform and motifs surrounding the phosphosites are indicated.

(C) Competitive growth assays in untreated conditions conducted on a BE3-expressing MCF10A cell population transduced with lentiviral constructs expressing the indicated 53BP1 sgRNAs. Data are represented as in Figure 3B. Mean ± SD for n = 2.

(D) Competitive growth assays in untreated conditions and upon doxorubicin treatment (5 nM) conducted on MCF7-BE3 cells transduced with lentiviral constructs expressing the indicated 53BP1 sgRNAs. Data are represented as in Figure 3B. Mean ± SD for n = 4.

(E) Characterization of MCF7-BE3 cells utilized in (D) and in Figure 3E. Immunoblot showing 53BP1 expression (top) and graph of the corresponding editing frequencies, as determined by Sanger sequencing and ICE analyses (bottom) at day 4 after sgRNA selection. Mean ± SD for n = 2.

(F) Characterization of MCF10A-BE3 cells utilized in (G-H) and in Figures 3G–3I. Immunoblot showing 53BP1 expression (top) and graph of the corresponding editing frequencies, as determined by Sanger sequencing and ICE analyses (bottom) at day 4 after sgRNA selection.

(G) Dot plot of the number of IR-induced γH2AX foci per cell in MCF10A-BE3 cells from the experiments shown in Figure 3G. Data belong to two independent experiments. Mean values are shown.

(H) Dot plot of the number of IR-induced BRCA1 foci per cell in MCF10A-BE3 cells from the experiments shown in Figures 3H–3I. Data represent three independent experiments. Mean values are shown.

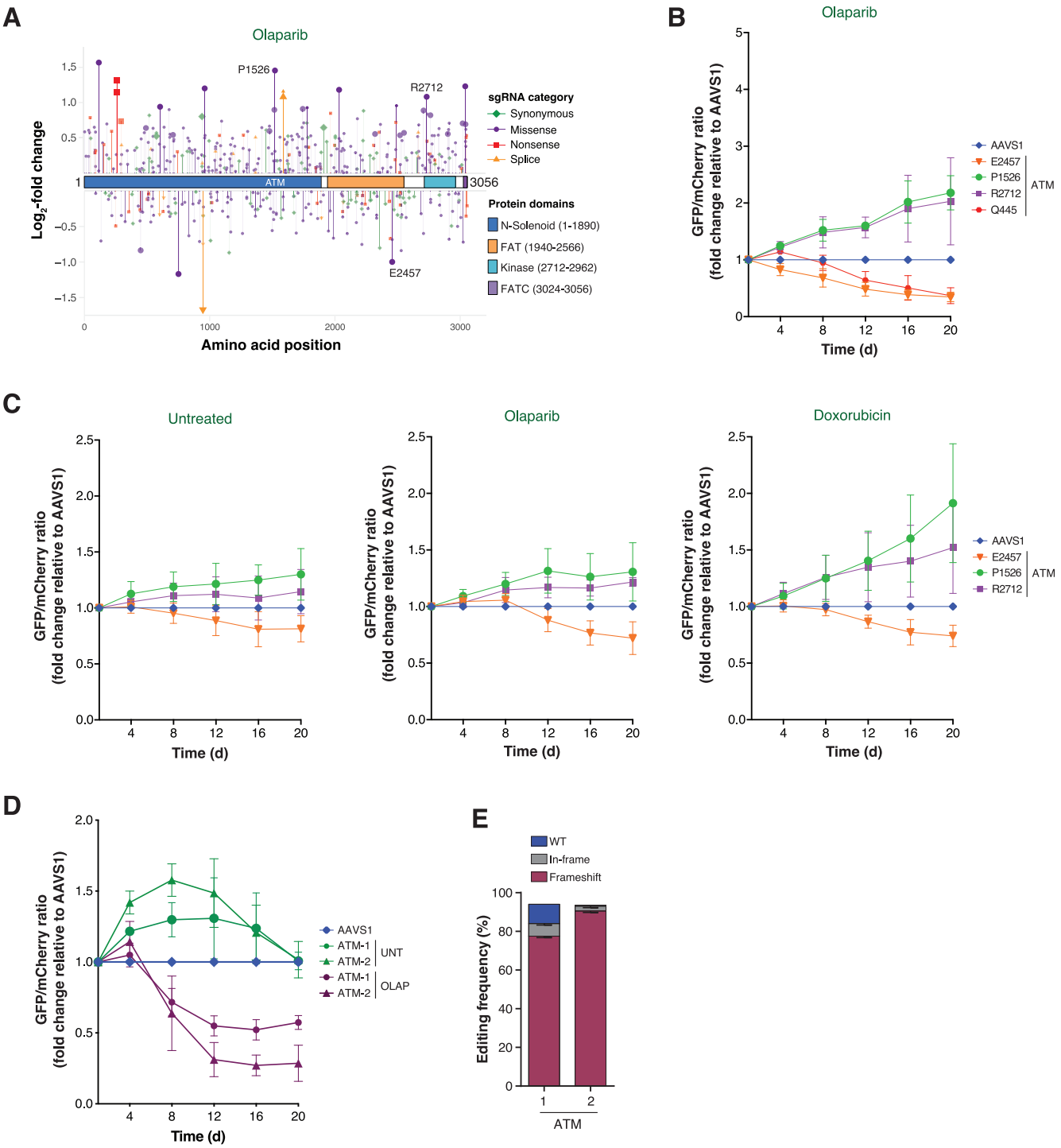


Figure S3. Phenotypic characterization of cells carrying ATM mutations, related to Figure 4

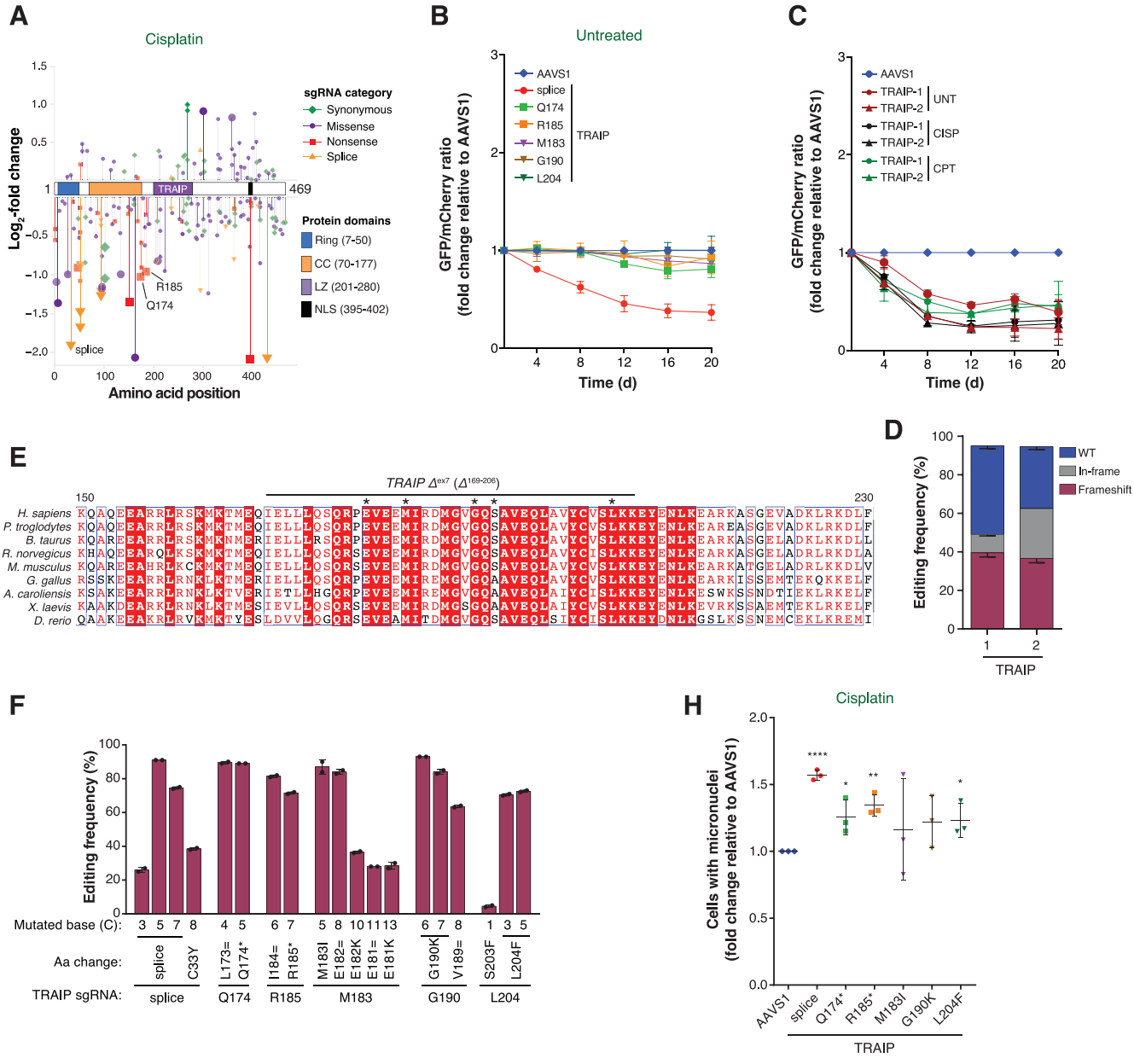
(A) Lollipop plot of ATM sgRNAs and their LFC values upon olaparib treatment mapped to the canonical ATM protein isoform, as shown in Figure 3A for 53BP1. Residues targeted by sgRNAs selected for validation are highlighted.

(B) Competitive growth assays upon olaparib (2 μM) treatment conducted on MCF10A-BE3 cells transduced with lentiviral constructs expressing the indicated ATM sgRNAs. Data are represented as in Figure 3B. Mean ± SD for n = 3.

(C) Competitive growth assays in untreated conditions and upon olaparib (2 μM) or doxorubicin (5 nM) treatment conducted on a BE3-expressing MCF10A cell population transduced with lentiviral constructs expressing the indicated ATM sgRNAs. Data are represented as in Figure 3B. Mean ± SD for n = 3.

(legend continued on next page)

(D) Competitive growth assays in untreated conditions and upon olaparib (2 μM) treatment conducted in a MCF10A cell population stably expressing SpCas9 and transduced with two independent ATM sgRNAs (1,2). Data are represented as in Figure 3B. Mean ± SD for n = 4.
(E) Analysis of editing frequency induced by the indicated ATM sgRNAs at their target loci at T4 in MCF10A cells expressing SpCas9, as determined by Sanger sequencing and ICE analyses. Mean ± SD for n = 2.



(legend on next page)

Figure S4. Characterization of the 174–204 amino acid region of the TRAIP ubiquitin ligase, related to Figure 5

(A) Lollipop plot of TRAIP sgRNAs and their LFC values under cisplatin treatment mapped to the canonical TRAIP protein isoform, as shown in Figure 3A for 53BP1. Nonsense and splice residues targeted by sgRNAs selected for validation are highlighted.

(B) Competitive growth assays in untreated conditions conducted on a BE3-expressing MCF10A cell population transduced with lentiviral constructs expressing the indicated TRAIP sgRNAs. Data are represented as in Figure 3B. Mean ± SD for n = 3.

(C) Competitive growth assays in untreated conditions and upon cisplatin (1 μM) and camptothecin (5 nM) treatment conducted on a MCF10A cell population stably expressing SpCas9 transduced with two independent TRAIP sgRNAs (1, 2). Data are represented as in Figure 3B. Mean ± SD for n = 2.

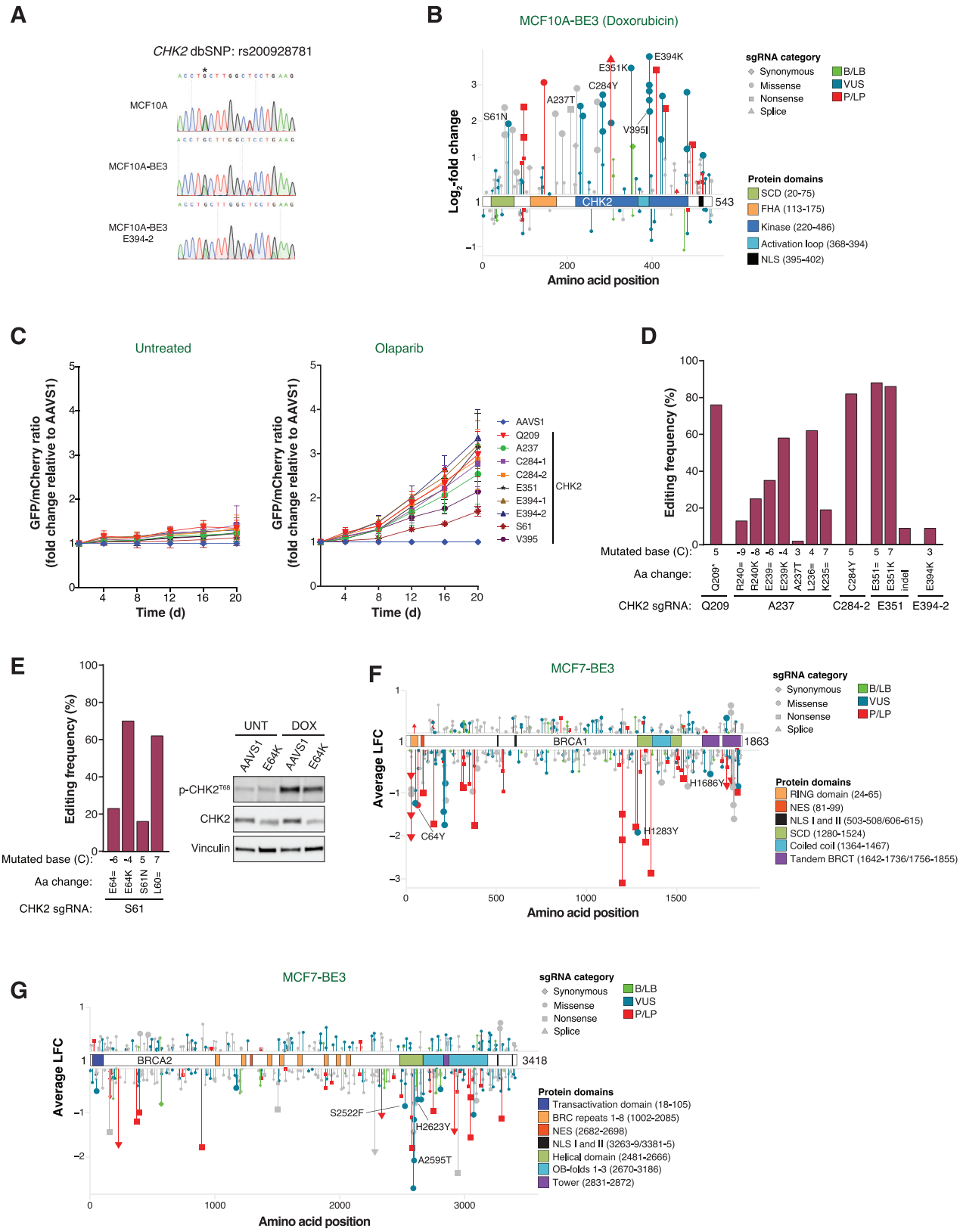
(D) Analysis of editing frequency induced by the indicated TRAIP sgRNAs at their target loci at T4 in MCF10A cells expressing SpCas9, as determined by Sanger sequencing and ICE analyses. Mean ± SD for n = 2.

(E) Alignment of the TRAIP protein sequences in the indicated species, highlighting the region encoded by exon 7. Residues targeted by relevant sgRNAs generating missense mutations are marked with asterisks, including the validated residues M183, and G190 and L204. Sequence alignments were conducted using MAFFT and visualization was performed on ESPript3.0.

(F) Analysis of editing frequency in MCF10A-BE3 cells induced by the indicated TRAIP sgRNAs at their target loci at day 4 post-selection prior to use in phenotypic characterization experiments, as determined by Sanger sequencing and ICE analyses. Mean ± SD for n = 2.

(G) Immunoblot showing TRAIP, RPA2 and H2AX expression, along with RPA2 S4/S8 phosphorylation and γH2AX levels in MCF10A-BE3 cells targeted with the indicated sgRNAs and subjected to UV treatment for 4 and 24 h (25 J/min). fl: full-length isoform; s: short isoform; *: non-specific band. Dotted line indicates a discontinuity on the blots.

(H) Analyses of micronuclei in MCF10A-BE3 cells targeted with the indicated sgRNAs upon camptothecin treatment (10 nM) for 48 h. Values are represented as fold change relative to the *AAVS1*-targeted control. Mean ± SD for n = 3. Statistical analysis for individual samples relative to the *AAVS1*-targeted control was performed using unpaired t test (*p value < 0.05, **p value < 0.01, ****p value < 0.0001).



(legend on next page)

Figure S5. Identification of variants of uncertain significance with pathogenic-like behavior in CHK2, BRCA1, and BRCA2, related to Figure 6

(A) Representative Sanger sequencing traces of the indicated *CHK2* locus from WT, BE3-expressing and sgRNA-edited MCF10A cells. The heterozygous mutation causing the Y390C AA substitution is marked with an asterisk.

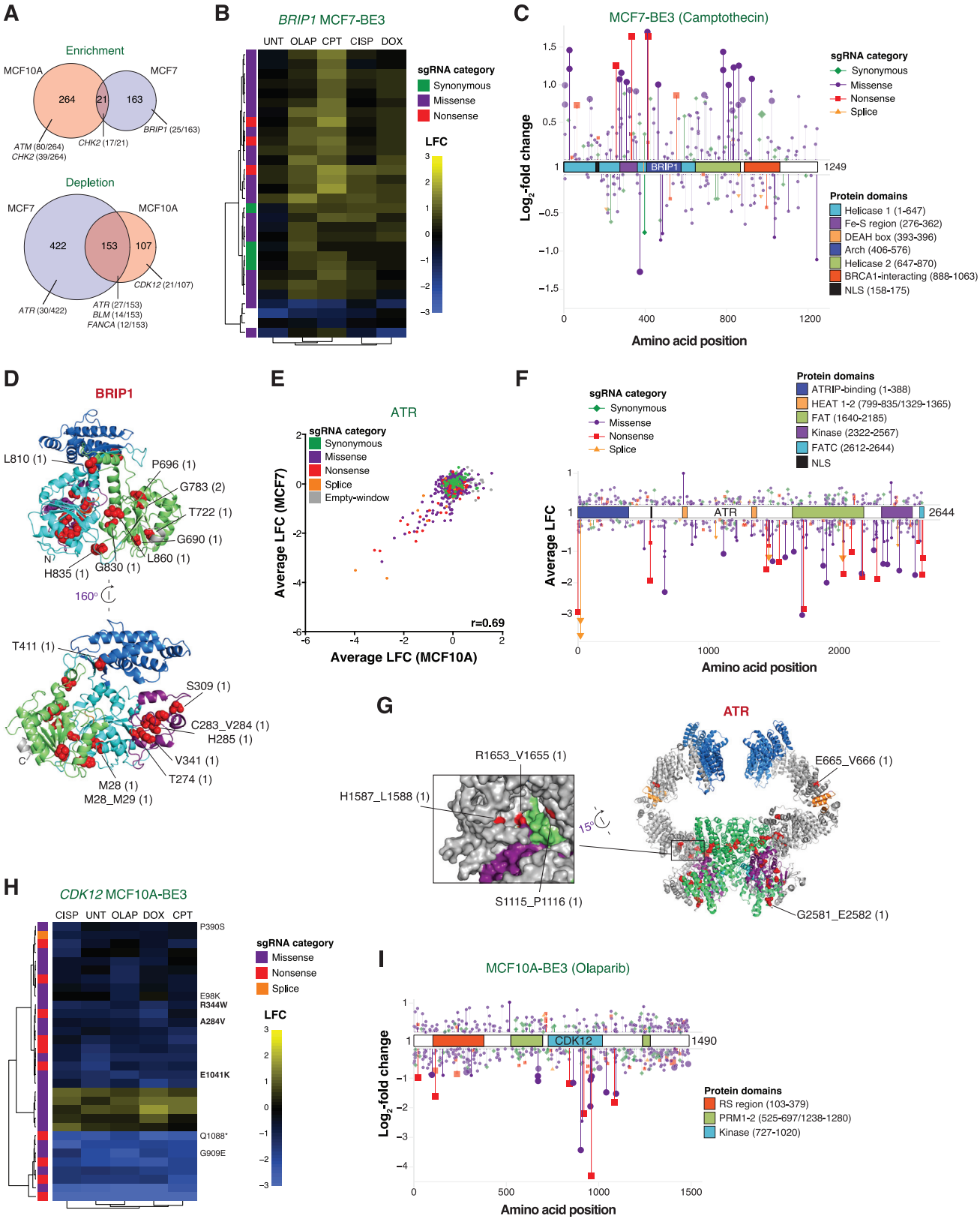
(B) Lollipop plot of CHK2 sgRNAs and their LFC values under doxorubicin treatment in MCF10A-BE3 cells mapped to the canonical CHK2 protein isoform, as in Figure 6J for BARD1. Colors indicate clinical relevance. sgRNAs that uniquely generate a VUS in the *CHK2* gene are indicated.

(C) Competitive growth assays in untreated conditions and upon olaparib treatment (2 μM) conducted on a BE3-expressing MCF10A cell population transduced with lentiviral constructs expressing the indicated CHK2 sgRNAs. Data are represented as in Figure 3B. Mean ± SD for n = 4.

(D) Analysis of editing frequency in MCF10A-BE3 cells induced by the indicated CHK2 sgRNAs at their target loci at day 4 post-selection, as determined by Sanger sequencing and ICE analyses.

(E) Characterization of MCF10A-BE3 cells targeted with the CHK2 S61 sgRNA. (Left) Editing frequencies induced by the CHK2 S61 sgRNA, as determined by Sanger sequencing and ICE analyses at T4 after sgRNA selection. (Right) Immunoblot showing CHK2 expression and its phosphorylation on T68 in MCF10A-BE3 cells, with or without doxorubicin treatment (100 nM, 8 h).

(F and G) Lollipop plots of BRCA1 (F) or BRCA2 (G) sgRNAs and their average LFC values under olaparib and cisplatin treatments in MCF7-BE3 cells mapped to the canonical BRCA1 or BRCA2 protein isoforms, as in Figure 6J for BARD1. Colors indicate clinical relevance. sgRNAs that uniquely generate a VUS with pathogenic-like behavior in the *BRCA1* or *BRCA2* gene are indicated.



(legend on next page)

Figure S6. Identification of mutational patterns of interest in other DDR genes, related to Figure 6

(A) Venn diagrams of relevant sgRNAs in MCF10A-BE3 and MCF7-BE3 screens. Diagrams were generated as in Figure 6C.

(B) Heatmap of LFC values for relevant BRIP1 sgRNAs in MCF7-BE3 cells, as shown in Figure 4A for ATM sgRNAs.

(C) Lollipop plot of BRIP1 sgRNAs and their LFC values in MCF7-BE3 cells upon camptothecin treatment mapped to the canonical BRIP1 protein isoform, as shown in Figure 3A for 53BP1.

(D) Cartoon of the Phyre2-predicted structure for the BRIP1 helicase region, with domains colored as in (C). Residues targeted by relevant sgRNAs are highlighted and depicted as red spheres in the structure. The number of sgRNAs targeting each residue is displayed in brackets.

(E) Correlation of averaged LFC values (CISP, OLAP) for relevant sgRNAs targeting the *ATR* gene in MCF10A-BE3 versus MCF7-BE3 cells. Pearson correlation value is shown. Dots are colored based on sgRNA category.

(F) Lollipop plot of ATR sgRNAs and their LFC values mapped to the canonical ATR protein isoform, as shown in Figure 3A for 53BP1. LFC values are calculated as the average of the LFCs in cisplatin and olaparib treatment in both MCF10A-BE3 and MCF7-BE3 cells. Threshold for biological relevance is set at the 1% of the averaged LFC values for negative controls. Statistical significance is considered for p values < 0.01 in at least one condition (CISP and/or OLAP) in both cell lines.

(G) Partial cartoon of the cryo-EM structure of an ATR dimer in the conformation bound to ATRIP (not shown) (PDB: 5YZ0 chains A and B; Rao et al. [2018]), with domains colored as in (F). Residues of interest are highlighted, and the number of relevant sgRNAs targeting each residue is shown in brackets.

(H) Heatmap of LFC values for relevant CDK12 sgRNAs in MCF10A-BE3 cells, as shown in Figure 4A for ATM sgRNAs. CDK12 sgRNAs generating TCGA mutations are highlighted, and sgRNAs generating unique mutations are shown in bold.

(I) Lollipop plot of CDK12 sgRNAs and their LFC values in MCF10A-BE3 cells upon olaparib treatment mapped to the canonical CDK12 protein isoform, as shown in Figure 3A for 53BP1.