Molecular Therapy
Original Article

enAsCas12a Enables CRISPR-Directed Evolution to Screen for Functional Drug Resistance Mutations in Sequences Inaccessible to SpCas9

Jasper Edgar Neggers,[1,4] Maarten Jacquemyn,[1] Tim Dierckx,[1] Benjamin Peter Kleinstiver,[2,3,5] Hendrik Jan Thibaut,[1] and Dirk Daelemans[1]

[1]KU Leuven Department of Microbiology, Immunology and Transplantation, Laboratory of Virology and Chemotherapy, Rega Institute for Medical Research, 3000 Leuven, Belgium; [2]Molecular Pathology Unit, Center for Cancer Research and Center for Integrative Biology, Massachusetts General Hospital, Charlestown, MA, USA; [3]Department of Pathology, Harvard Medical School, Boston, MA, USA

While drug resistance mutations provide the gold standard proof for drug target engagement, target deconvolution of inhibitors identified from a phenotypic screen remains challenging. Genetic screening for functional in-frame drug resistance mutations by tiling CRISPR-Cas nucleases across protein coding sequences is a method for identifying a drug's target and binding site. However, the applicability of this approach is constrained by the availability of nuclease target sites across genetic regions that mediate drug resistance upon mutation. In this study, we show that an enhanced AsCas12a variant (enAsCas12a), which harbors an expanded targeting range, facilitates screening for drug resistance mutations with increased activity and resolution in regions that are not accessible to other CRISPR nucleases, including the prototypical SpCas9. Utilizing enAsCas12a, we uncover new drug resistance mutations against inhibitors of NAMPT and KIF11. These findings demonstrate that enAsCas12a is a promising new addition to the CRISPR screening toolbox and allows targeting sites not readily accessible to SpCas9.

INTRODUCTION

One of the bottlenecks of (phenotypic) drug discovery is the subsequent identification and validation of the cellular target of candidate hit molecules.[1] Knowledge on the cellular target of identified hits is critical for predicting side effects and the validity and tractability of interfering with the target in the disease context, and it supports hit-to-lead optimization and biomarker development.[2] However, target deconvolution and target confirmation for hit molecules remain challenging. There is no clear go-to, off-the-shelf methodology. Nevertheless, genetic approaches provide particularly powerful tools for the drug mechanism of action studies considering that the discovery of mutations conferring drug resistance within a cellular context is viewed as the gold standard proof for target engagement.[3,4]

Recently, we have developed a CRISPR-directed evolution approach that allows for target deconvolution based on targeted and rapid generation of functional drug resistance mutations in protein coding regions (CRISPR-induced resistance in essential genes [CRISPRres]).[5] In human cells, CRISPR-Cas-induced DNA double strand breaks (DSBs) are mainly repaired by non-homologous end joining (NHEJ), generating insertion and deletion mutations (indels).[6] Within protein coding sequences, these indels lead to a variety of both out-of-frame and in-frame mutations.[7–10] Out-of-frame (frameshift) indels severely disrupt protein synthesis and are commonly used to generate loss-of-function alleles.[11,12] CRISPRres, however, leverages in-frame mutations to derive gain-of-function protein variants that confer drug resistance. Drug-resistant cells generated through this approach typically contain a variety of genetic mutations around the DSB, resulting in amino acid deletions, insertions, and substitutions that keep the protein's translational frame intact. Indeed, DNA DSBs introduced by single guide RNA (sgRNA)-guided SpCas9 targeted to resistance hotspots allowed for rapid selection of drug-resistant cells across multiple independent studies.[5,13–19] An important strength of CRISPRres is that the sgRNA sequences used to target the CRISPR nuclease directly annotate the genomic sequence containing the drug resistance-conferring mutations, allowing a single readout based on a simple targeted amplicon sequencing reaction of the sgRNA cassette. In addition, CRISPRres not only allows for identification of a drug's target protein, but it also provides information on the drug-protein binding interface with amino acid resolution.[5,13] We previously showed an SpCas9 sgRNA library tiling 64 genes that correctly identified PSMB5 as the target of bortezomib.[5] It also identified the bortezomib binding interface with amino acid resolution and uncovered novel resistance mutations. Moreover,

Received 14 May 2020; accepted 15 September 2020;
https://doi.org/10.1016/j.ymthe.2020.09.025.

[4]Present address: Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA.

[5]Present addresses: Center for Genomic Medicine and Department of Pathology, Massachusetts General Hospital, Boston, MA, USA; and Department of Pathology, Harvard Medical School, Boston, MA, USA.

Correspondence: Dirk Daelemans, KU Leuven Department of Microbiology, Immunology and Transplantation, Laboratory of Virology and Chemotherapy, Rega Institute for Medical Research, 3000 Leuven, Belgium.
E-mail: dirk.daelemans@kuleuven.be

tiling of another 75 genes with SpCas9 sgRNAs identified a binding site within the enzymatic pocket of NAMPT as the molecular target for the antineoplastic compound KPT-9274.

Target binding and enzymatic activity of the CRISPR-SpCas9 endonuclease requires recognition of the 3-nt protospacer-adjacent motif (PAM) 5′-NGG.[20,21] Tiling with SpCas9 is thus limited by the availability of NGG motifs across the targeted coding sequences. If no NGG PAM sites near resistance hotspots for the investigated compound are available, it is unlikely the approach will generate resistance mutations. The sparsity of PAMs further limits the approach when the sequence context of the genetic target regions favor only certain major mutational outcomes.[7,8] To partially mitigate these constraints, we previously showed that CRISPRres-based target deconvolution is compatible with the class 2 type V CRISPR-Cas AsCas12a,[5] which utilizes the relatively rare TTTV PAM.[22] However, the combination of AsCas12a with canonical SpCas9 is still insufficient to reach targeting saturation of many protein coding sequences, highlighting the need for additional CRISPR endonucleases utilizing new PAM recognition sites to identify resistance mutations in genomic regions that lack NGG and TTTV PAMs. In this study, we demonstrate that the recently reported enhanced AsCas12a variant (enAsCas12a)[23] is compatible with CRISPRres and allows for identification of new drug resistance mutations in previously inaccessible genomic regions, thanks to its substantially expanded targeting range.

RESULTS

Engineered CRISPR Nucleases Complement the Targetable Genetic Space of SpCas9 and AsCas12a

To determine the fraction of protein coding sequences that are accessible for mutagenesis by the canonical SpCas9 CRISPR endonuclease, we set out to examine the outcome of DNA repair of SpCas9-induced DNA DSBs. We utilized CrispRVariants[24] to analyze existing reference-aligned targeted amplicon next-generation sequencing reads to determine the mutational spectra and outcome at 71∼78 CRISPR-SpCas9 cut sites in human HCT116, HEK293T and K562 cells 11 days after lentiviral transduction of the CRISPR machinery.[9] Across all 223 samples (4,000–20,000 sequencing reads/sample), we detected a total of 127,186 unique mutated sequence variants (Table S1), indicating robust editing. However, the fraction of unaltered sequences varied between cell lines: HCT116, 3.5% ± 12.6%; HEK293T, 39.0% ± 20.5%; K562, 8.1% ± 12.8% (mean ± SD). Most altered sequences consisted of 1- to 30-bp deletions or a single nucleotide insertion, although some larger variants were also observed (Figure 1A). Regardless of sequence context, we categorized deletions and insertions as in-frame or out-of-frame mutations and included nucleotide variants as in-frame mutations. Although the fraction of in-frame and out-of-frame variants varied heavily between sgRNAs, on average 35.7% ± 16.5% (mean ± SD) of variants at each target site consisted of in-frame mutations (Figure 1B). To spatially examine the variants at the nucleotide level, we utilized the DeepSNV package.[25] We examined a region of 46 nt surrounding the genomic sgRNA target sequence for each sgRNA independently across the three cell lines. For each sample, we normalized all DeepSNV nucleotide calls (from all in-frame, out-of-frame, and wild-type reads) to the positive orientation and normalized call positions relative to the SpCas9 cut site. We then summed the results to obtain a cumulative average of variant calls. Interestingly, a skewed normal distribution was observed for each cell type, with deletions making up most of the variants (Figure 1C; insertions incorporated as mutations). Nucleotides upstream of the PAM sequence appeared altered more often than did downstream nucleotides (Figure 1C). Of note, when examining individual target sites in individual samples, the mutational spectrum varied wildly, indicating that the sequence context and the function of each genomic target site strongly influence the DNA repair outcome (Table S2), in line with previous observations.[7,8] When nucleotide calls for only in-frame variants where examined, the overall mutational spectrum remained similar (compare Figure S1A with Figure 1C), while the spectrum in individual samples differed substantially and was characterized by more nucleotide variants (compare Table S2 with Table S3), as might be expected.

Based on these observations, we set an effective window of mutagenesis after NHEJ repair to 14 bp around the SpCas9 cut site, including the PAM site, and extrapolated this window to AsCas12a (Figure 2A). We then calculated the targetable genetic space within entire protein coding sequences for 190 human anticancer drug target genes (Table S4) as a fraction using CRISPR RNAs (crRNAs) designed for the canonical 5′-NGG (SpCas9) and 5′-TTTV (AsCas12a) PAMs (Table S5). SpCas9 was able to cover 83.0% ± 9.5% (mean ± SD) of the examined coding sequences with 65,841 unique sgRNAs, while AsCas12a only covered 29.7% ± 13.6% (mean ± SD) with 10,498 crRNAs (Figure 2B). Combining SpCas9 and AsCas12a only increased coverage marginally, indicating strong overlap and highlighting the need for additional CRISPR endonucleases with different PAM recognition to expand the targetable genetic space. In recent years, new CRISPR nucleases have been engineered for increased activity and relaxed PAM recognition. Applying the same 14-bp window to calculate the mutable window for two SpCas9 variants, xCas9-3.7 (PAM: 5′-NG, 5′-GAT, 5′-GAA, or 5′-CAA)[26] and SpCas9-NG (PAM: 5′-NG),[27] highlighted that these two variants cover close to 100% of coding sequences, albeit with a steep increase in the amount of sgRNAs (xCas9-3.7: 99.8% ± 0.4% (mean ± SD), 317,915 sgRNAs; SpCas9-NG: 99.7% ± 0.5% (mean ± SD), 220,102 sgRNAs) (Figure 2B; Table S5). Utilizing the top 23 different PAMs associated with enAsCas12a, an engineered variant of AsCas12a with increased activity and relaxed PAM recognition,[23] showed that a total of 73,802 crRNAs cover 80.9% ± 9.1% (mean ± SD) of the coding sequences (Figure 2B; Table S5). Interestingly, when combining SpCas9 and enAsCas12a, 97.4% ± 1.1% (mean ± SD) of the coding sequences was covered (Figure 2B), indicating that not all SpCas9 and enAsCas12a target spaces are mutually accessible. Analysis of the targetable space in individual genes, such as *NAMPT*, further illustrated that the targetable space of a given coding sequence can vary between Cas effector endonucleases (Figure 2C; Table S6). As the broadened PAM compatibility of xCas9-3.7, SpCas9-NG, and enAsCas12a allow targeting of previously inaccessible sites within the genome, we set out

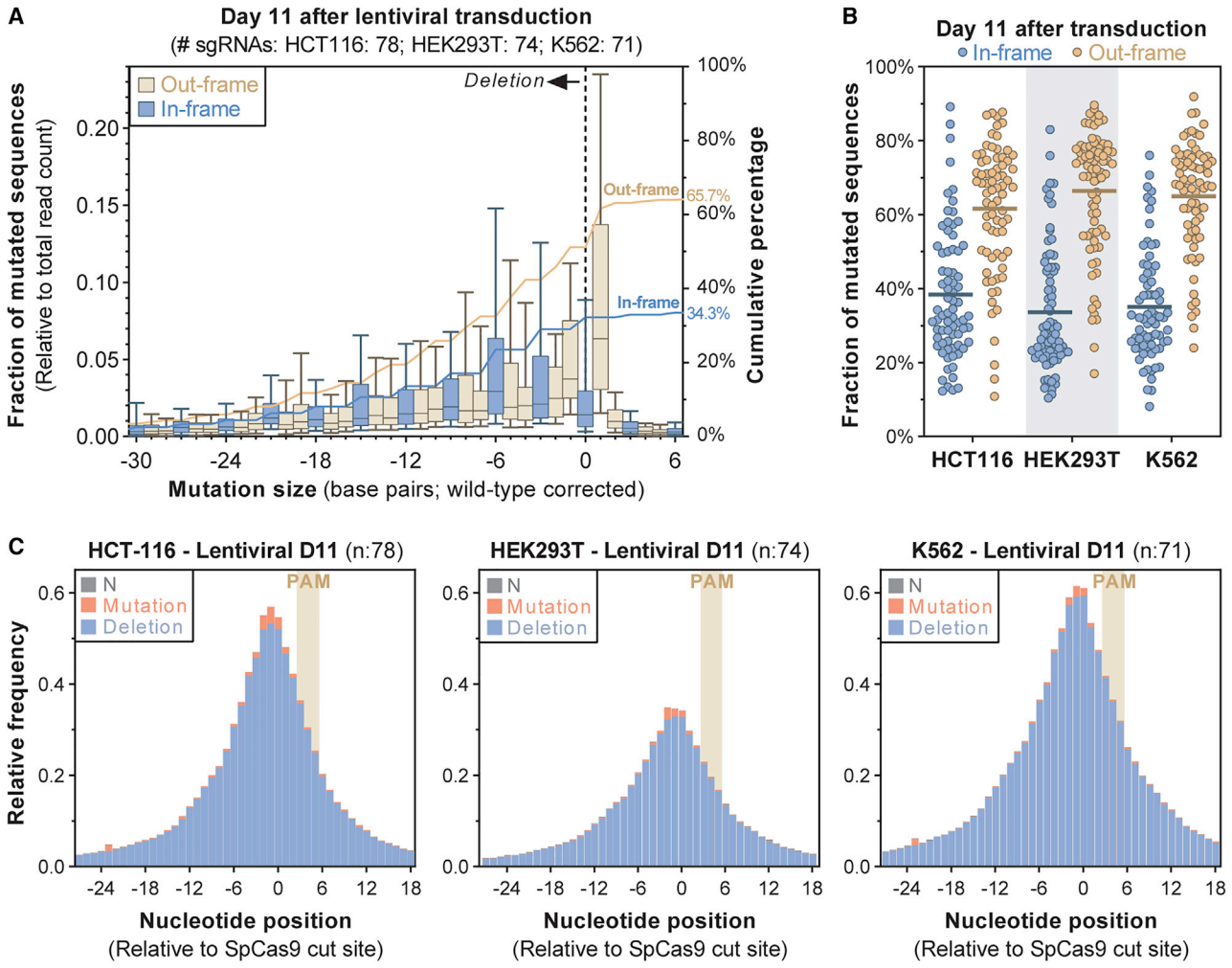


Figure 1. SpCas9-Induced NHEJ Repair Leads to In-Frame and Out-of-Frame Mutations Localized in Front of the 5′-NGG PAM

(A) Boxplots showing the fraction of all mutated genetic variants (left y axis) of a certain indel size (x axis) as determined by CrispRVariants analysis of targeted amplicon sequencing of 223 SpCas9 sites across three human cancer cell lines. Samples were analyzed on day 11 after lentiviral transduction of the SpCas9 machinery. Boxes indicate the 25th to 75th percentiles with the median (horizontal black line), and whiskers indicate the 10th to 90th percentiles. Indels that are multiples of 3 nt are indicated as in-frame (blue). Deletions are indicated by negative values, while insertions are represented by positive values (x axis). Mutations with a size of 0 indicate single or multiple nucleotide variants. The two overlaying lines indicate the cumulative fraction of out-of-frame (orange) and in-frame (blue) mutations (right y axis). All values are wild-type corrected, meaning that they are relative to only the total number of variants excluding wild-type alleles. (B) Scatterplots showing the fraction of in-frame (blue) and out-of-frame (orange) genetic variants detected at the cut site of individual SpCas9 sgRNAs (y axis) grouped by cell line (x axis). Horizontal lines indicate means. Each dot represents a different cut site at a specific SpCas9 sgRNA. Fractions are wild-type corrected and only include sequences that are different from the reference (e.g., mutated). N: 78 (HCT116), 74 (HEK293T), and 71 (K562) sgRNAs. (C) Barplots showing single-nucleotide variant occurrence at the sgRNA target site (y axis) at each nucleotide position relative to the SpCas9 cut site (x axis) observed across all target sequences grouped by mutation type for each of three cell lines. Values represent means. Mutation types include N (gray; unclear), mutation (pink; SNVs, multiple-nucleotide variants [MNVs], and insertions) and deletions (blue). Values were obtained by DeepSNV analysis of targeted amplicon sequencing of SpCas9 target sites. PAM, protospacer-adjacent motif. See also Figure S1 and Tables S1, S2, and S3.

to determine whether these CRISPR endonucleases would permit the discovery of new drug resistance mutations previously unidentified with SpCas9.

Mutagenesis Scanning with SpCas9-NG Identifies New, but Few, Drug Resistance Hotspots

We first wanted to investigate the ability of alternative SpCas9 nucleases to generate resistance mutations in *NAMPT*, the target gene of the antineoplastic agents FK866 (Figure S1B)[28,29] and KPT-9274 (Figure S1C).[5,30–34] Using SpCas9 and a large, multi-gene tiling sgRNA library at a coverage of 5,000 cells/sgRNA, we previously identified three coding regions in *NAMPT* that confer KPT-9274 and FK866 resistance when mutated: I11 to Y18 (exon 1), P236 to Y240 (exon 6), and G381 to G384 (exon 9).[5] FK866 and KPT-9274 bind the same enzymatic pocket of NAMPT, although KPT-9274 is branched and binds two different patches, whereas FK866 only occupies a single

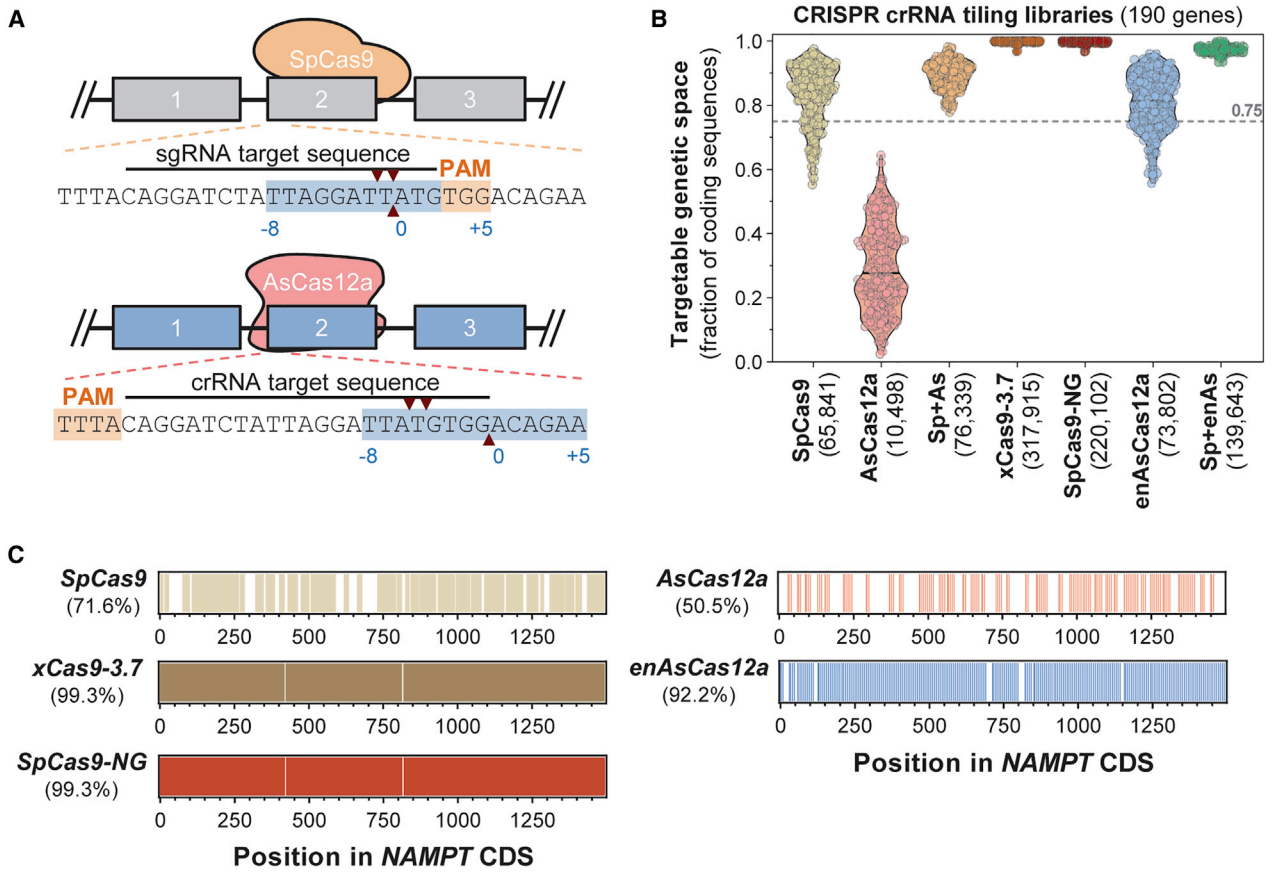


Figure 2. Engineered CRISPR Nucleases Complement the Targetable Genetic Space of SpCas9 and AsCas12a

(A) Graphical illustration of the theoretical mutable window in which in-frame mutations are still expected at a reasonable frequency for a given crRNA. Red arrowheads highlight cut sites of the shown CRISPR endonucleases. (B) Violin plot showing the theoretical targetable genetic space (y axis) of various CRISPR guide tiling libraries (x axis) covering the coding sequences of 190 human anticancer drug target genes. Values were calculated using the 14-bp mutable range defined in (A). Horizontal lines represent the median, and each dot represents a gene. The number of crRNAs for each tiling library is indicated in brackets (x axis). (C) Estimates of the mutational coverage for SpCas9 (beige), xCas9-3.7 (brown), SpCas9-NG (red), AsCas12a (pink), and enAsCas12a (blue) using a 14-bp mutational window within the coding sequence of *NAMPT*, the target gene of the anticancer agents KPT-9274 and FK866. Colored regions indicate the region of the coding sequence that falls within a mutable range, while white areas indicate regions that fall outside. See also Tables S4, S5, and S6.

site.[5] FK866 failed to progress to phase 2 clinical trials due to on-target dose-limiting toxicity (thrombocytopenia) and lack of efficacy.[35] KPT-9274 is less potent than FK866 and can be dosed more frequently. It is currently entering clinical trials but still elicits gastrointestinal toxicities and non-clinically observed cardiac and retinal toxicities. KPT-9274 is also reported to bind PAK4, although the clinical relevance of this activity remains unclear[30,36].

Recent reports demonstrated that xCas9-3.7 shows strongly reduced DNA cleavage activity at canonical 5′-NGG PAMs and even at some non-canonical PAMs (5′-NGA, 5′-SAG, 5′-NGA) when compared to SpCas9.[27,37–39] This observation was supported by colleagues (unpublished data), and hence we did not further explore xCas9-3.7. Instead, we focused on SpCas9-NG, which has been shown to have higher DNA cleavage activity than xCas9-3.7 at most PAMs, but it is still known to have lower activity at canonical 5′-NGG PAMs than does SpCas9.[27,37–39] To determine whether SpCas9-NG allows screening for drug resistance against KPT-9274, we generated a library of 5′-NG PAM sgRNAs tiling the entire coding sequence of *NAMPT* (Table S7). In addition, we also designed a 5′-NG PAM sgRNA library tiling *KIF11* to scan for resistance against the antineoplastic agent ispinesib (Table S7; Figure S1D). Ispinesib is a small molecule antineoplastic agent that binds an allosteric site in the N-terminal motor domain of the essential protein kinesin-5 (KIF11).[40,41] Binding of this site inhibits KIF11 function, which leads to impaired mitotic spindle formation and eventually to cell death. Cellular resistance against ispinesib can be conferred by mutations in the drug's binding pocket in KIF11,[41] and we previously identified four sgRNAs that confer ispinesib resistance mutations in *KIF11* exons 4 and 5, which encode the majority of the ispinesib binding pocket, by applying a SpCas9 mutagenesis scan.[5]

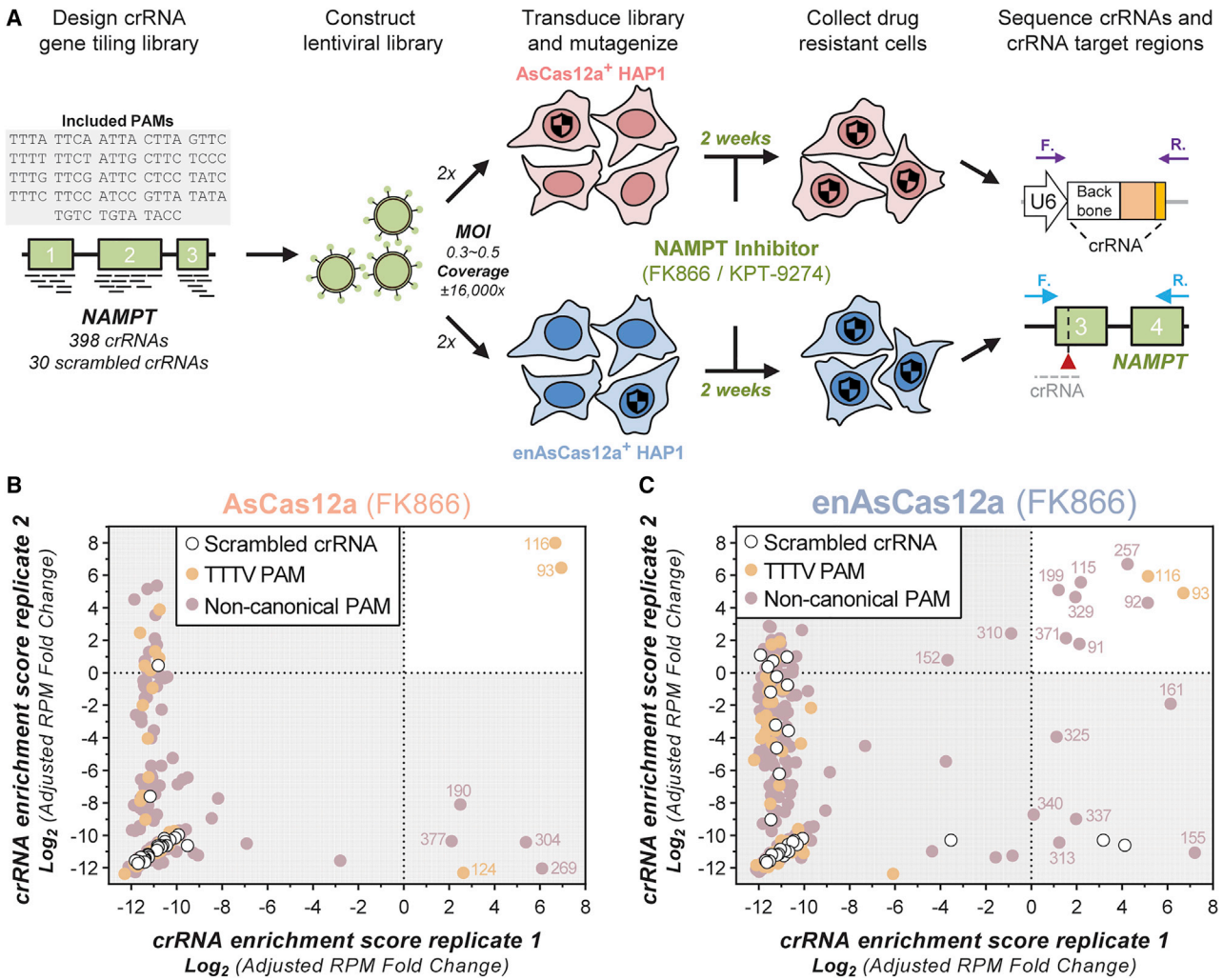


Figure 3. enAsCas12a Enables Efficient Mutagenesis Scanning for Resistance against FK866

(A) Workflow of the chemical resistance mutagenesis screening experiment to compare AsCas12a with enAsCas12a using the NAMPT-KPT-9274/FK866 drug-target interactions. A lentiviral crRNA tiling library based on the indicated PAMs was generated to cover the coding sequence of human *NAMPT*. AsCas12a- or enAsCas12a-stable HAP1 cells were transduced with the lentiviral library and allowed to generate mutations for a period of 3 days under puromycin selection. Cells were then treated with 300 nM KPT-9274 or 10 nM FK866 for 2 weeks. Following treatment, the gDNA from resistant colonies was harvested and subjected to next-generation sequencing. The experiment was repeated once. (B) Enrichment scores (adjusted fold change over day 0 representation) for crRNAs present in transduced AsCas12a$^+$ HAP1 cells after treatment with 10 nM of the NAMPT inhibitor FK866. Fold change was determined by taking the $\log_2$ of the reads per million (RPM) after treatment added by 1. This value was then divided by the +1 adjusted RPM before treatment. Both replicate screens are shown (x and y axis). Read counts were determined using EdgeR. Each dot represents a different crRNA, which are divided into three categories, including scrambled control (white), canonical TTTV PAM (orange), and non-canonical PAM (pink). (C) Enrichment scores (adjusted fold change over day 0) for crRNAs present in transduced enAsCas12a$^+$ HAP1 cells after treatment with FK866, visualized as in (B). See also Figure S3 and Tables S10 and S11.

We transduced both libraries, in duplicate, at a depth-coverage of 20,000 (KPT-9274) or 10,000 (ispinesib) cells per sgRNA into HAP1 cells stably expressing SpCas9-NG to screen for drug resistance against KPT-9274 and ispinesib (Figure S2A). Our original results with SpCas9 indicated that canonical SpCas9 was more efficient at generating drug resistance mutations across *NAMPT* and *KIF11* than was SpCas9-NG, even though we originally only used 5,000 cells per sgRNA for SpCas9.[5] In more detail, across both replicate screens with SpCas9-NG, three sgRNAs were enriched after KPT-9274 treatment (Figure S2B; Table S8), while we previously identified six sgRNAs conferring KPT-9274 resistance with SpCas9.[5] Interestingly, the three sgRNAs utilized the non-canonical NGA/NGT PAMs and targeted codons (K84, Y87, F91) at a new resistance hotspot site in exon 3. Unexpectedly, none of the sgRNAs associated with an NGG PAM that we previously identified was enriched with SpCas9-NG, although these sgRNAs were present in the library (Table S8). Similarly, for ispinesib, only two sgRNAs were enriched (Figure S2C; Table S9). One utilized the non-canonical 5′-GGA PAM (KIF11_1153)

and the other utilized a 5′-GGG PAM (KIF11_1154). Both sgRNAs mapped to codon L132 at the beginning of *KIF11* exon 5. Mutations in this exon, and in particular around codons D130–A133, are well known to confer ispinesib resistance, and we previously identified resistance mutations at this site with SpCas9.[5,40,41] In conclusion, SpCas9-NG showed less overall mutational activity than did canonical SpCas9 in our experimental setup and identified less KPT-9274 and ispinesib resistance hotspots than did SpCas9, although for KPT-9274 it did identify a new site in exon 3. This highlights that SpCas9-NG supports CRISPR-directed evolution and allows screening mutations in regions inaccessible to SpCas9, albeit at an overall lower efficiency.

enAsCas12a Enables Efficient Screening For Resistance against NAMPT Inhibitors

We next turned to the class 2 type V CRISPR-Cas nuclease AsCas12a (AsCpf1) and an engineered variant, enAsCas12a.[22,23] We cloned a crRNA library targeting the top 23 enAsCas12a compatible PAMs across the *NAMPT* coding sequence and screened in duplicate for FK866 and KPT-9274 resistance mutations using HAP1 cells stably expressing AsCas12a or enAsCas12a (Figure 3A). This crRNA library contained 398 targeted crRNAs covering 92% of *NAMPT* the coding sequence (Figure 2C: enAsCas12a), and 30 scrambled control crRNAs.

For FK866, only two crRNAs (NAMPT_93 and NAMPT_116) were enriched in both replicate screens with AsCas12a (Figure 3B; Table S9), and both associated with a canonical TTTV PAM. For enAsCas12a, nine crRNAs were enriched in both replicate screens (Figure 3C; Table S10). Of these, seven targeted a non-canonical PAM, while the other two were the ones that were also enriched with AsCas12a. To validate the screening hits, all nine enriched guides were cloned into plasmids individually and co-transfected transiently into parental HAP1 cells with a plasmid encoding AsCas12a or enAsCas12a. Using AsCas12a, only NAMPT_116 and NAMPT_93 clearly conferred FK866 drug resistance (Figure S3A), although a marginal number of resistant colonies appeared with the TTTT PAM-associated crRNAs NAMPT_92 and NAMPT_115 (Figure S3A). These crRNAs target similar sites as NAMPT_93 and NAMPT_116, indicating that AsCas12a showed residual activity for TTTT PAMs, which is a known phenomenon.[22] Transfections with enAsCas12a validated that all enriched guides were able to induce drug resistance against FK866 (Figure S3B). The enriched guides targeted *NAMPT* either in exon 3 (NAMPT_115, 116, 199, 257) or in exon 6 (NAMPT_91, 92, 93, 329, 371). Although both regions were uncovered in the AsCas12a screen with a single crRNA, screening with enAsCas12a strengthened this finding by three or four additional crRNAs, highlighting that the increased targeting space of enAsCas12a allows for more robust identification of drug resistance hotspots.

The site in exon 3 was not identified in our previous SpCas9-mutagenesis screen and harbors the D93 residue, previously reported to confer resistance to FK866 and other NAMPT inhibitors upon mutation.[42] This site was probably not identified in the SpCas9 screen because it is only covered by a single sgRNA targeting its boundary. Additionally, the abundance of this sgRNA in our original screen was low, possibly due to a technical reason. The site at exon 6 spans the G239 to P243 amino acid stretch, which we previously identified as a KPT-9274 resistance hotspot, and a new site at A245. Although we were unable to sequence exon 1 and 2 of *NAMPT* due to high GC content, we uncovered 61 unique in-frame mutations and 9 out-of-frame mutations in the other exons across all four samples (Table S11). Although out-of-frame mutations were not abundant, *NAMPT* is essential for HAP1 survival, suggesting that some cells had turned diploid during the experiment, which is known to occur spontaneously in HAP1 cells.[43] As expected, most in-frame mutations localized to exon 3 and 6 and mainly consisted of deletions in the D85 to D93 (exon 3) and G239 to E246 (exon 6) amino acid stretches (Table S11). These results show that genetic screening with (en)AsCas12a enables identification of resistance mutations that were unachievable with SpCas9.

When performing the screen with KPT-9274, 19-fold more KPT-9274-resistant colonies were derived from enAsCas12a$^+$ cells than from AsCas12a$^+$ cells (Figure 4A). Both populations were highly resistant to treatment with KPT-9274, showing a 350- to 500-fold increase in the 50% effective concentration (EC_{50}) (Figure S4). In AsCas12a$^+$ cells, six crRNAs were enriched in both replicate experiments (Figure 4B; Table S12). Of these, three crRNAs utilized a canonical PAM while the remaining three crRNAs required a non-canonical PAM, TTTT (NAMPT_115), TTCT (NAMPT_202), or TACC (NAMPT_369). As NAMPT_202 and NAMPT_369 were only enriched slightly and were not enriched in enAsCas12a$^+$ cells (Figure 4C; Table S12), these are likely false positives. In contrast, 22 crRNAs were enriched in both replicate experiments using enAsCas12a$^+$ cells (Figure 4C; Table S12). Many of these crRNAs required a non-canonical PAM, demonstrating the advantages of the broadened PAM compatibility of enAsCas12a in a lentiviral tiling screening setting using human cells.

Individual validation transfections with AsCas12a demonstrated that only 2 of 24 crRNAs (NAMPT_93 and 116) were able to confer clear resistance to KPT-9274 (Figure S5). Not surprisingly, these two crRNAs utilized a canonical TTTV PAM and were also the most enriched screening hits in AsCas12a$^+$ cells (Figure 4B). We did observe outgrowth of a minor number of resistant colonies with other guides, including guides associated with non-canonical PAMs (Figure S5; Table S12). However, it remains unclear whether these arose from incomplete selection, background-resistant cells, or from bona fide resistant mutations induced by the transfected guide. Individual validation experiments with enAsCas12a showed that 13 of 24 crRNAs (NAMPT_53, 91, 92, 93, 115, 116, 164, 199, 224, 253, 257, 310, and 371) clearly conferred KPT-9274 resistance (Figure S6). These crRNAs were also the ones that were enriched the most in the screens (Figure 4C; Table S12), with top enriched guides generating a high number of resistant colonies (Figure S6). The crRNAs that did not validate mainly included very poorly enriched guides and scrambled

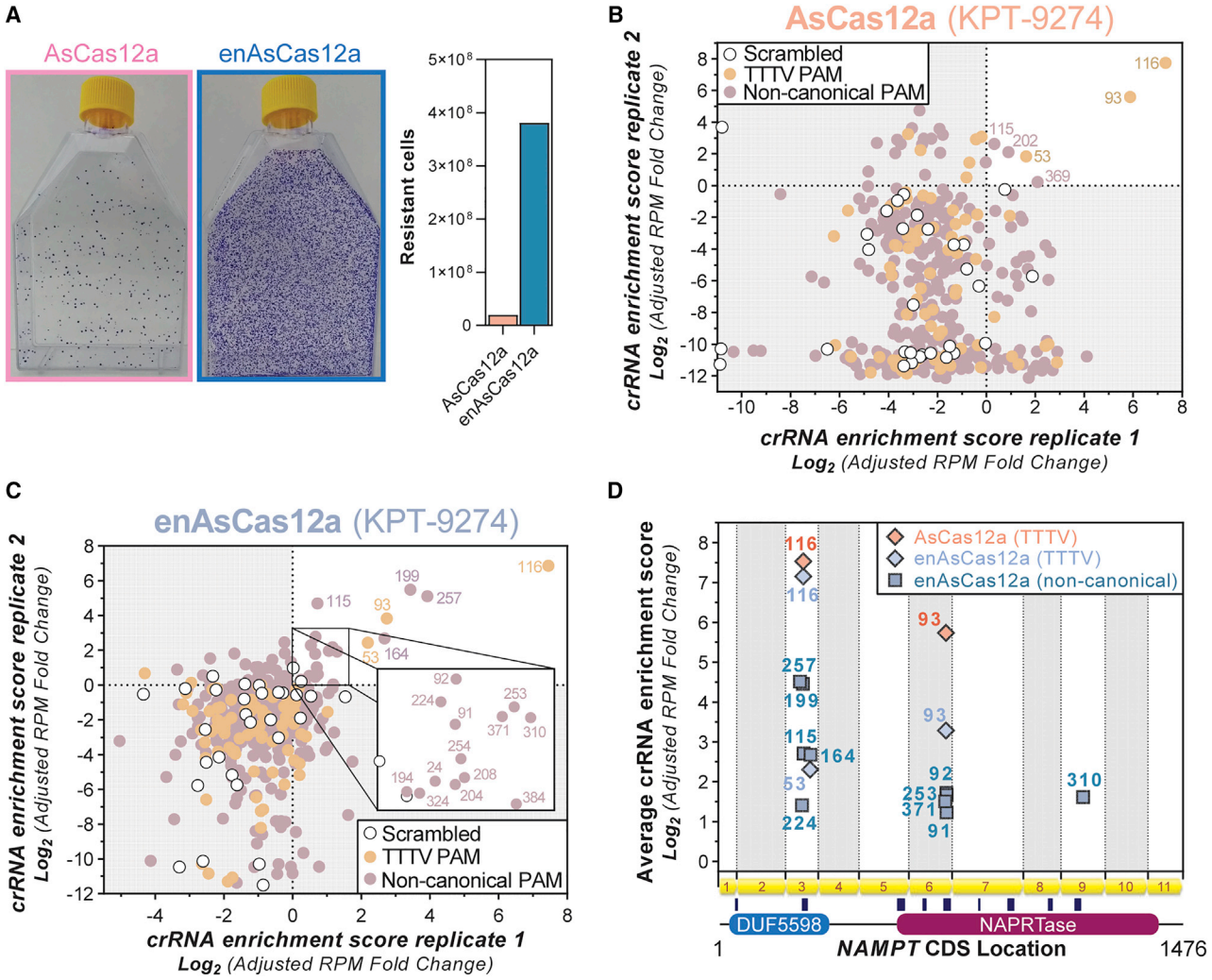


Figure 4. enAsCas12a Enables Efficient Mutagenesis Scanning for Resistance against KPT-9274

(A) Visualization of KPT-9274-resistant colonies obtained during the screen after 12-day treatment with KPT-9274 using 0.5% gentian violet. The total amount of resistant cells harvested from replicate 1 from both the AsCas12a (pink) and enAsCas12a (blue) mutagenesis scanning experiments for KPT-9274 resistance are shown on the right. A 19-fold increase in resistant cells was obtained with enAsCas12a. Cells were counted using trypan blue exclusion and a Luna II automated cell counter. (B) Enrichment scores for crRNAs (log_2-adjusted fold change over day 0 representation) present in transduced AsCas12a$^+$ HAP1 cells after treatment with 300 nM KPT-9274. Fold change was determined by taking the log_2 of the RPM after treatment added by 1. This value is then divided by the +1 adjusted RPM before treatment. Both replicate screens are shown (x and y axis). Read counts were determined using EdgeR, and each dot represents a different crRNA, which are divided into three categories, including scrambled control (white), canonical TTTV PAM (orange), and non-canonical PAM (pink). (C) Enrichment scores for crRNAs present in transduced enAsCas12a$^+$ HAP1 cells after treatment with KPT-9274, visualized as in (B). The blowup highlights the cluster of poorly enriched crRNA guides. (D) Genetic target locations for the validated crRNAs enriched in AsCas12a$^+$ (pink) and enAsCas12a$^+$ (blue) HAP1 cells after treatment with KPT-9274. The x axis denotes the location within the *NAMPT* coding sequence with exons indicated by numbered yellow arrows. A schematic showing the overall structure of NAMPT is shown below the exons. NAMPT contains two annotated protein domains (PFAM v33.1; N-terminal: domain of unknown function 5598 (DUF5598); C-terminal: nicotinate phosphoribosyltransferase family [NAPRTase]). The residues that surround the active and drug binding site in the NAMPT enzyme are highlighted by dark blue boxes. See also Figures S4–S7 and Tables S12 and S13.

controls (Figure 4C, enlarged box; Figure S6). Consistent with the expanded PAM preference of enAsCas12a, several crRNAs designed for a TTTT PAM (NAMPT_91, 92 and 115) were validated in enAsCas12a$^+$ cells but not in AsCas12a$^+$ cells, highlighting the increased compatibility of enAsCas12a with TTTT PAMs (Figures S5 and S6; Table S12). Finally, as the NAMPT_329 crRNA was able to induce drug resistance against FK866 (Figure 3C) but did not enrich with KPT-9274, we also tested this guide against KPT-9274 and found that it induced robust drug resistance when coupled to enAsCas12a (Figure S3C).

Mapping of the 13 validated crRNAs to the *NAMPT* coding sequence revealed two major target sites at exon 3 and exon 6 (Figure 4D), similar to the results with FK866. These sites span the residues E89

to D93 and G239 to P243 in the KPT-9274 binding pocket. Both sites were only identified by a single crRNA in AsCas12a$^+$ cells, while seven and five crRNAs targeting these regions were enriched and validated in enAsCas12a$^+$ cells. One crRNA targeting exon 9 (NAMPT_310) was also enriched and validated only in enAsCas12a$^+$ cells (Figure 4D). This crRNA targets the L387 codon, which is located just behind the G383 to G385 stretch we previously identified with SpCas9, further strengthening the finding that mutations in this region can confer KPT-9274 resistance. These results highlight the superiority of enAsCas12a over AsCas12a for identifying resistance hotspots with high precision and efficiency. Of note, however, both (en)AsCas12a tiling screens for FK866 and KPT-9274 resistance failed to nominate exon 1, although with SpCas9 we previously identified this site as a NAMPT inhibitor resistance hotspot.[5] Examination of the mutational window across *NAMPT* of the (en)AsCas12a and SpCas9 tiling libraries revealed that AsCas12a does not cover any of the first 29 codons of *NAMPT* (exon 1 and parts of exon 2) and that enAsCas12a only covers the first 10 codons (start of exon 1), but then shows a gap in mutational coverage until codon 30 (halfway exon 2) (Figure 2C; Table S6). In contrast, SpCas9 sgRNAs cover the first 13 codons, then show a small gap for codons 14–17, and continue covering codon 18 onward. Mutation of the Y18 codon in exon 1 is likely to generate KPT-9274/FK866 drug resistance due to its role in PI stacking of the pyridyl groups of KPT-9274/FK866,[5] explaining why SpCas9, but not (en)AsCas12a, identified this resistance site.

Deep sequencing of *NAMPT* uncovered a total of 81 in-frame mutations and 8 out-of-frame mutations across replicates, with most mutations consisting of deletions of a single to a few amino acids (Table S13). Most in-frame mutations localized to exon 3 and 6, right next to the cut sites of the enriched crRNAs (Table S13). enAsCas12a$^+$ cells also carried a wider spectrum of mutations at exon 3. The most common mutations were similar as observed for FK866 (Table S11), and they consisted of deletions in the E89 to D93 (exon 3) and G239 to P243 (exon 6) amino acid stretches, which both map to the KPT-9274 binding site (Figure S7). Moreover, a L387del mutation at exon 9 was identified in enAsCas12a$^+$ cells. We also identified a few single point mutations in exons 4 (T133K, G148V), 5 (A173D), and 8 (P346H), and a silent SNP (R429R) in exon 10 (Table S13), although their relevance remains uncertain, as these were only enriched in single replicates and are not clearly related to the target site of any of the enriched crRNAs.

An enAsCas12a Multi-gene Tiling Library Enables Efficient Target Deconvolution

To investigate whether enAsCas12a can be utilized in a larger deconvolution screening setting, we applied a multigene targeting crRNA library to scan for resistance against the antineoplastic agent ispinesib.[5,40,41] For this purpose, HAP1 cells stably expressing enAsCas12a were transduced with a crRNA library containing 3,838 different crRNAs tiling 10 genes (Figure 5A; Table S14). After selection for transduction, cells were pulsed twice with 15 nM ispinesib during a period of 14 days and the experiment was carried out in duplicate. Many colonies rapidly formed, and these cells were highly resistant to ispinesib treatment (Figure S8A). Sequencing revealed that 10 crRNAs were enriched in both replicate experiments (Figure 5B; Table S14). Of these, nine crRNAs targeted *KIF11* and one targeted *RPS3A* (Table S14). All *KIF11*-targeting crRNAs except KIF11_192 could be validated to confer ispinesib resistance (Figure S8B). The *RPS3A* crRNA (RPS3A_98) failed validation, suggesting that this crRNA may have been derived from inherently resistant cells or co-enriched with the one of the other crRNAs, a phenomenon we previously observed.[5]

The validated crRNAs mapped to exons 4, 5, and 6 of *KIF11* (Figure 5C; Table S14). These exons encode the majority of the ispinesib binding pocket within KIF11, which mainly consists of α helices 2 and 3 and loop 5 (Figure S8C).[44] The site in exon 4 encompasses loop 5 and was identified by four crRNAs (KIF11_172 [TTTA], 286 [TTCC], 753 [TCCC], and 802 [TACC]) targeting codons for S120, P121, and N122, highlighting the broadened PAM compatibility of enAsCas12a allowed for efficient identification of this resistance hotspot. The second resistance hotspot localized around codon A133 at the end of loop 5 encoded by exon 5 and was identified by two crRNAs (KIF11_463 [ATTA] and KIF11_658 [TATA]) (Table S14). Mutations of A133 are well known to confer ispinesib resistance.[5,40,41] One enriched crRNA (KIF11_706 [CTTC]) targeted a region of loop 6 encoded by codons N173 to E180 in the second half of exon 5. Loop 6 is next to α helix 3, and mutations in loop 6 might disrupt the orientation of α helix 3, which could disturb the conformation of the ispinesib binding site (Figure S8C). A resistance hotspot surrounding codon R221 in exon 6 was also identified by one guide (KIF11_190 [TTTA]). R221 localizes to the center of α helix 3 and is in proximity with ispinesib, again suggesting that mutations in this region are likely to disrupt binding of ispinesib (Figure S8C). Finally, the last enriched crRNA (KIF11_192 [TTTC]), which failed validation, targeted codon F239 in exon 7 (Table S14). F239 maps to a β sheet closing the ispinesib binding pocket at the core of the motor domain and its side chain is in proximity (±2 Å) to ispinesib (Figure S8C). Thus, it remains unclear why KIF11_192 failed validation. We next sequenced the underlying hotspots and uncovered various mutations around residues 117 to 124, 132 and 133, 173 to 180, and 217 to 232 (Table S15). Only in a single replicate did we uncover a large multi-nucleotide variation consisting of S237I, V238C, F239Y, S240N, and V241I. Most other mutations consisted of single or multiple amino acid deletions and often affected residues within the ispinesib binding site. We also uncovered two mutations, G117V and D149Y, which did not associate with any enriched crRNA sequence. Interestingly, the G117V mutation localizes to the beginning of loop 5 and is likely to confer ispinesib resistance. In contrast, the D149Y mutation maps to the end of the second α helix, away from the ispinesib binding site.

DISCUSSION

In humans, DNA DSBs are often repaired by non-homologous or microhomology-mediated end joining, which are regarded as error-prone processes that lead to nucleotide indels.[6,45] In coding sequences, these indels may result in a frameshift of the reading frame, causing loss of protein function due to premature stop codons and

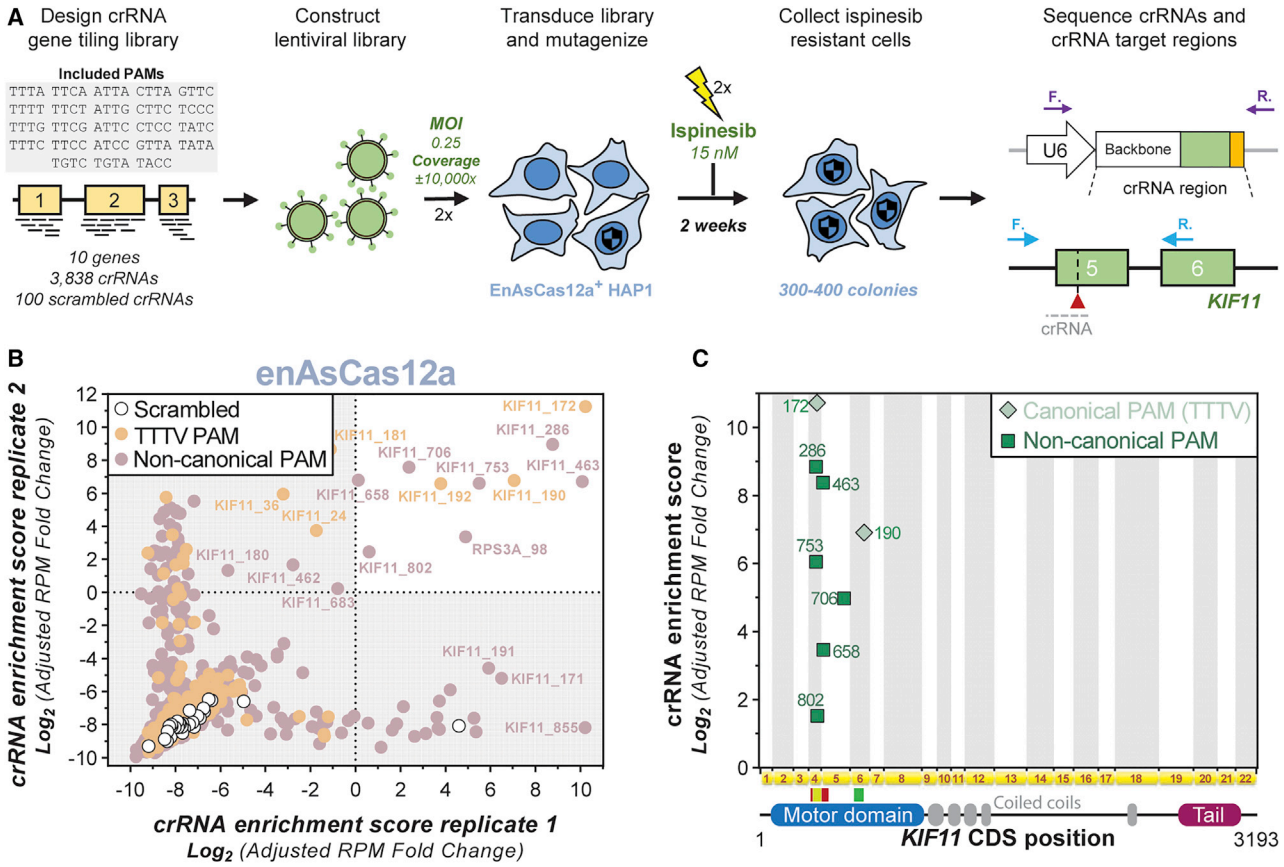


Figure 5. An enAsCas12a Multi-gene Tiling Library Enables Deconvolution of the Target of Ispinesib

(A) Workflow of the chemical resistance mutagenesis screening experiment to determine whether enAsCas12a can correctly identify the target of ispinesib using the known KIF11-ispinesib drug-target interaction from a set of 10 genes. A lentiviral crRNA tiling library based on the indicated PAMs targeting the coding sequence of 10 different human genes was generated. The lentiviral library was then transduced into HAP1 cells expressing enAsCas12a. Cells were selected with puromycin for a period of 3 days and then pulsed twice with 15 nM ispinesib during a period of 2 weeks. Following treatment, the DNA from resistant colonies was harvested and subjected to next-generation sequencing. The experiment was repeated once. (B) Enrichment scores for crRNAs present in transduced enAsCas12a^{+} HAP1 cells after treatment with ispinesib from both replicate screens (x and y axis). Fold change was determined by taking the $\log_2$ of the RPM after treatment added by 1. This value was then divided by the +1 adjusted RPM before treatment. Read counts were determined using EdgeR, and each dot represents a different crRNA, divided into three categories, including control (white), canonical PAM (orange), and non-canonical PAM (pink). (C) Genetic target locations for the validated crRNAs enriched in enAsCas12a^{+} HAP1 cells after treatment with ispinesib. The x axis denotes the location within the *KIF11* coding sequence with exons indicated by yellow arrows and red numbers. The overall structure of KIF11 is shown as a schematic below the exons. KIF11 contains three annotated protein domains (PFAM v33.1; the N-terminal kinesin motor domain, the linear coiled-coil region, and the C-terminal kinesin-associated microtubule-binding domain). The residues that encompass the ispinesib drug binding site in the N-terminal domain are highlighted (red, α helix 2; yellow, loop 5; green, α helix 3). See also Figure S8 and Tables S14 and S15.

nonsense-mediated decay. Based on this concept, targeted endonucleases have been used extensively to generate loss-of-function alleles of a desired genetic locus.[46] With the establishment of RNA-guided CRISPR endonucleases, such approaches have become mainstream.[47,48] However, indels may also resolve in in-frame mutations, changing only a few amino acids within the protein sequence. By analyzing the repair outcome at 223 canonical SpCas9 CRISPR endonuclease target sites 11 days after lentiviral transduction,[9] we observed that such in-frame mutations on average comprise about one-third of the introduced variants. This might have been expected considering the triplet nature of the genetic code. For individual target sites, however, the results are unique and much more variable. In-frame variants account for 8% of variants at some sites, while they make up 90% of variants at other sites. The outcome of DNA repair at a given CRISPR-SpCas9 DSB target site has been shown to be reproducible and can be predicted based on the local DNA sequence context and occurrence of microhomology regions.[7–10,49,50] Although not examined in this study, CRISPR sites in protein coding sequences might be predisposed to the selection of in-frame mutations, especially when the targeted sequence encodes for essential genes.

Exon-intron boundaries might further limit the diversity of sequence variants detected at such loci. Indeed, various studies have demonstrated that the likelihood to observe triplet-conserving variants in

protein coding sequences is slightly higher than in non-coding regions.[7,8] Interestingly, more than 80% of the variants generated at the 223 examined target sites consisted of deletions below 30 bp, while insertion of a single base pair (often A or T) on the SpCas9 cut site was by far the most observed insertion.

When in-frame amino acid substitutions occur within a binding pocket of a small molecule inhibitor, these mutations can confer resistance against this molecule by preventing the inhibitor from engaging the target protein. Previously, we demonstrated that this principle can be harnessed for target deconvolution of antineoplastic drugs. We showed that tiling of targeted CRISPR-Cas guide RNAs across coding sequences of a set of protein targets allows for rapid generation of functional drug-resistant protein variants, allowing us to uncover the target and mechanism of action of small molecule inhibitors.[5] This approach allows mutagenesis of endogenous alleles in their native context, and because RNA-guided nucleases introduce a DNA DSB at a specific genomic sequence defined by the guide RNA and the presence of a PAM, the location of gain-of-function mutations can be easily retrieved based on the target sequence of the crRNA. However, such CRISPR mutagenesis approaches require PAMs inside genetic loci of drug resistance hotspots. If the required PAM is unavailable near the resistance hotspot, CRISPR-based mutagenesis will be unable to generate the required resistance mutations, preventing the identification of the molecule's protein target.

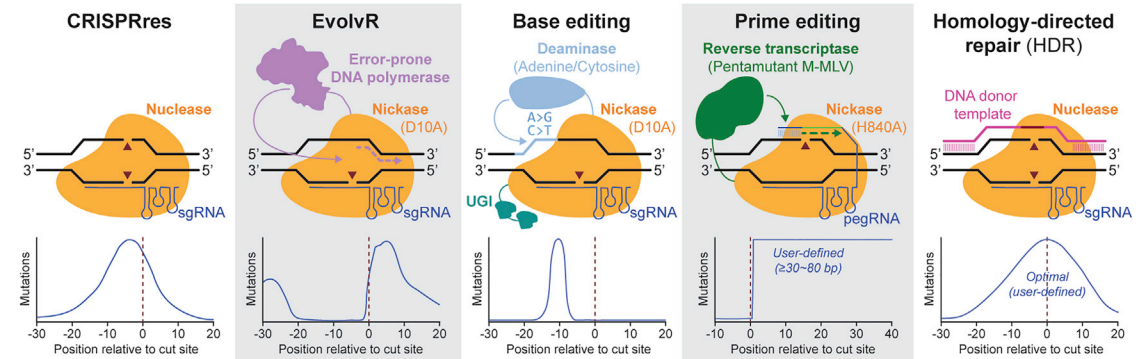
In this study, we calculated that the canonical and widely used CRISPR-SpCas9 endonuclease, which utilizes the NGG PAM, covers between 55% and 95% of a set of 190 human protein coding sequences. The AsCas12a endonuclease, which uses the less prevalent TTTV PAM,[22] only covers between a few percent and 65% of the same protein coding sequences. Combining AsCas12a and SpCas9 increases the coverage to 75%–97%, but on average 12% of the coding sequences remains unreachable. enAsCas12a, an engineered endonuclease with expanded PAM recognition,[23] covered a similar percentage of protein coding sequences as SpCas9. However, combining SpCas9 with enAsCas12a allowed targeting of 95%–100% of coding sequences, finally reaching saturation. This saturation can also be achieved by newly developed SpCas9 endonucleases that are being engineered or evolved at a rapid speed for relaxed PAM requirements.[26,27,39,51] In particular, we observed that xCas9-3.7 and SpCas9-NG[26,27] allow covering of 99%–100% of the examined coding sequences. However, recent reports indicated that these variants can show decreased DNA cleavage activity (especially xCas9-3.7) and increased off-target activity when compared to SpCas9.[37–39,52]

When we applied SpCas9-NG to develop drug resistance against the antineoplastic agent KPT-9274, which targets the NAD^+-generating enzyme NAMPT,[31,33,34] and ispinesib, which inhibits the mitotic function of KIF11 (ispinesib),[44,53] we observed that SpCas9-NG was less efficient than canonical SpCas9 at generating drug resistance mutations for these two drugs. However, SpCas9-NG, through sgRNAs associated with a non-canonical PAM, did identify a new KPT-9274 resistance site in exon 3 of *NAMPT*. This highlights that SpCas9-NG supports CRISPR-directed evolution and allows screening mutations in regions inaccessible to SpCas9, albeit at an overall lower efficiency. Together with recent reports, this highlights that new SpCas9 variants can be utilized for their additional targeting space, but their utility should be examined on a case-by-case basis.

In a search for additional alternatives, we then experimentally demonstrated that enAsCas12a allowed for the detection of FK866, another NAMPT inhibitor,[28,29,35] KPT-9274, and ispinesib resistance hotspots that were not accessible to canonical SpCas9.[28,29,31,33–35,44,53] In line with our previous observations with SpCas9, most in-frame resistance mutations generated by enAsCas12a consisted of deletions of one or more amino acids inside or close to the drug binding pocket. For both NAMPT inhibitors, we identified an additional amino acid stretch around the D85 to D93 residues that can drive drug resistance upon mutation. Due to the limited availability of NGG PAM sites in this region, this site was previously not identified with SpCas9.[5] Of note however, this site was the only one identified by SpCas9-NG. When using enAsCas12a, we also observed a striking increase in drug-resistant colonies when compared to AsCas12a. This difference could be attributed to the increased accessibility of the enAsCas12a endonuclease to drug resistance hotspots due to increased compatibility with non-canonical PAMs. We note that our results with KPT-9274 were less clear than with FK866. This can be explained by the observation that KPT-9274 is a weaker inhibitor than FK866. Whereas a low nanomolar dose of FK866 killed most cells within 2 days, KPT-9274 required a 10-fold higher dose to kill most cells during 3–5 days. Moreover, KPT-9274 was originally identified as a PAK4 binder and is labeled as a dual inhibitor of PAK4 and NAMPT,[31,34] although the relevance of KPT-9274's proposed activity against PAK4 remains unclear in our context.

Mutagenesis scanning of 10 different genes by enAsCas12a correctly led to identification of KIF11 as the target of ispinesib. In addition, four different hotspots in KIF11 relevant for ispinesib resistance were identified. These included residues surrounding S120, A133, N173 to E180, and R221. Using SpCas9, we previously only identified three hotspot sites for ispinesib resistance, namely residues around T126, A133, and S175, demonstrating that enAsCas12a was able to uncover more sites than SpCas9.[5] Notably, both endonucleases identified unique sites and, as such, they are complementary to one another. Moreover, although various studies have previously reported drug resistance mutations against ispinesib,[41,53] some of the detected mutations using enAsCas12a are new and locate to different domains in the protein. Therefore, CRISPR tiling screens enable the identification of novel drug resistance mutations located farther away from the direct drug binding pocket.

Although CRISPR-directed evolution using DNA DSBs is very feasible, other CRISPR systems are available to perform targeted mutagenesis at endogenous loci (Figure 6). Base editors use nuclease-deficient CRISPR endonucleases to target cytosine and adenosine deaminases to genetic sequences and allow for site-specific mutagenesis without introducing a DSB.[54–57] Originally, these



	CRISPRres	EvolvR	Base editing	Prime editing	Homology-directed repair (HDR)
Components	Any CRISPR nuclease	Nickase SpCas9 (D10A) fused to *E. coli* PolI (3M/5M)	Nickase CRISPR nucleases fused to adenine or cytosine base editors	Nickase SpCas9 fused to reverse transcriptase +donor sgRNA template	Any CRISPR nuclease +donor DNA template
Predominant mutation type	Deletions and insertions (1~30bp)	Single nucleotide variants	Single/multi nucleotide variants	Small variants or indels *(template design)*	User determined *(template design)*
Mutational diversity	Diverse, but predictable and not unlimited	Mainly single nucleotide transitions	Mainly A>G, C>T transitions	User determined *(template design)*	User determined *(template design)*
Mutational window	-15 to +6bp *5' / 3' of cut site*	0 to +20bp *3' of nickase site*	-17 to -10bp *5' of nickase site*	+1 to +33~80bp *3' of nickase site*	Theoretically kilobases *5' / 3' of target site*
Demonstrated scalability	Large, thousands of crRNAs, multi-gene	Small, few crRNAs, few sites	Medium, some crRNAs, some sites/full genes	Small, few crRNAs, few sites	Small, few crRNAs, few sites
Continous evolution	Unlikely and difficult	Possible	Somewhat possible	Unclear / Unlikely	Unlikely and difficult
Multiplexing	Risk of translocations, cleavage toxicity	Relatively easy	Relatively easy	Unlikely, especially with double nicking strategy	Difficult
Applied organisms	Mammals, plants, yeast	Bacteria, yeast	Bacteria, yeast, mammals, plants	Mammals, plants	Dividing mammal, plant and yeast cells

Figure 6. Overview of Currently Available Approaches for CRISPR-Directed Evolution

Five classes of CRISPR-based genome editing systems are highlighted. A schematic is shown for each class to highlight the basic principle underlying each method, using canonical SpCas9 as the example CRISPR endonuclease. For each method, a generalized plot is also shown below the schematic to provide a sense of the mutational window for effective edits. sgRNA, single-guide RNA; pegRNA, prime editing guide RNA; HDR, homology-directed repair; UGI, uracil glycosylase inhibitor; M-MLV, Moloney murine leukemia virus.

utilized a nuclease-dead version of the canonical SpCas9, but they have since expanded to different nuclease-dead and nickase endonucleases with increased targeting range, including enAsCas12a.[23,58] Different base editors have now also been successfully combined in a single experiment.[51,59–61] However, the experimental setup of these deaminase fusions is different from the NHEJ-based mutagenesis approach described herein. Base-editor approaches are logically not suited to derive deletions or insertions. We note that in-frame indels seem to provide a major mechanism for drug resistance, even in functionally important protein domains. In addition, base editors are generally optimized for therapeutic use and often generate single, specific cytosine or adenine transition point mutations in a narrow window, limiting their potential to generate genetic variation. A high genetic variation is essential to uncover new drug resistance mutations, as it is often unknown which mutations will confer drug resistance or even which gene should be targeted. For this purpose, hyperactive base editors have been developed to successfully screen for drug resistance mutations, although these studies were limited to screening mutations on single well-known drug target genes.[62–64] Base-editing approaches might also be less probable to recover recessive drug resistance mutations, as this requires the same mutation to be present in both alleles. Nevertheless, we think that the development of base editors for use of rapid mutagenesis and directed evolution will prove useful and complementary. Besides base editors, it is also possible to utilize homology-directed repair, or more recently, prime editing, to insert desired mutations at a given target site using targeted endonucleases.[65–68] These approaches, however, require up-front development of a template sequence carrying the desired mutation of interest. Although useful for validation studies, targeted mutations, and single-site saturation,[69] it remains difficult to envision the successful use of

such approaches in a large-scale directed evolution screening setting. In this light, the coupling of DNA polymerases to targeted endonucleases shows promise, as it has been shown to enable diversification of all nucleotides in a predefined window at single sites.[70] However, so far, this approach remains limited to bacteria and yeasts.[70,71]

In summary, our findings highlight the efficiency and increased resolution of mutagenesis screening with enAsCas12a, demonstrating that enAsCas12a forms a great addition to the CRISPR toolbox. We did not assess off-target activity of enAsCas12a in this study, but enAsCas12a previously showed a slightly increased off-target activity when compared with AsCas12a, and a high-fidelity version of enAsCas12a was generated to counteract this constraint.[23] Although we did not perform any explicit measurements, we did not encounter any problems with DNA cleavage activity for enAsCas12a. Furthermore, enAsCas12a was recently utilized successfully for large-scale loss-of-function screening.[72] Based on our observations, enAsCas12a shows all of the required features to become a standard tool for CRISPR-based mutagenesis screens.

Conclusions

In this work, we demonstrate that the recently engineered CRISPR-Cas enAsCas12a endonuclease allows for rapid and straightforward genetic screening for gold standard drug resistance mutations and can be applied in a target deconvolution setting. The increased PAM recognition of enAsCas12a enables it to uncover new resistance mutations in previously undetected loci against the antineoplastic agents FK866, KPT-9274, and ispinesib. These uncovered mutations provide additional insight into the drug-target interaction profile of these compounds. Our findings further highlight that the expanded PAM recognition of enAsCas12a opens up new possibilities and applications for CRISPR genetic screening by allowing efficient targeting of genetic regions that are difficult to reach with canonical CRISPR-Cas endonucleases. Based on this, we envision that enAsCas12a will provide a strong standard for CRISPR-based mutagenesis scanning, in which target saturation of large genetic regions with crRNAs is critical.

MATERIALS AND METHODS

SpCas9 Mutational Outcome Analysis

To improve our understanding of the occurrence of in-frame mutations after NHEJ repair of CRISPR-SpCas9 DNA DSBs in human cells, we obtained reference-aligned targeted amplicon sequencing results from 223 different single sgRNA-SpCas9 mutagenesis experiments across three different human cell lines (HCT116, HEK293T, K562) from a previous study uploaded to the NCBI Sequence Read Archive (SRA) (project SRP076796).[9] Specifically, we obtained sequencing data for all three cell lines 11 days after they were infected with the SpCas9 machinery by lentiviral transduction. Previously, these sequences were aligned to the human reference genome hg19, and pre-aligned BAM files (not SRA files) were downloaded through direct URL links to the NCBI SRA project using the Google Chrome web browser (v80.0.3987.132) equipped with the Simple mass downloader plugin (by George Prec, v0.831) on a personal computer running Windows 10. BAM files were then sorted and indexed using the Rsamtools Bioconductor package (v2.2.3) in RStudio with R (v3.6.2). These sorted and indexed BAM files were then analyzed using a custom R script utilizing the R package CrispRVariants (v1.14).[24] Briefly, we used the sequence "CIGAR" string output from CrispRVariants to determine the size of deletions, insertions, and nucleotide variants for each sgRNA sample and saved these results together with the occurrence of each variant into a table (Table S1). This table was then analyzed in Microsoft Excel 365 by hand to determine the frequencies of different variant types, e.g., in-frame or out-of-frame for each sgRNA sample. For these analyses, the "no variant" sequences were excluded, and the results were visualized in GraphPad Prism v8.3.0. Following determination of the in-frame and out-of-frame occurrences, we next applied a custom R script using the DeepSNV package[25] on the BAM files to determine single-nucleotide occurrences around the SpCas9 target sites. We initially performed this analysis on all sequencing reads for each sgRNA sample independently (Table S2) by setting the orientation in the positive direction for each sample, and then normalizing the results relative to the position of the SpCas9 cut site (3 nt before the PAM) to obtain relative normalized values. We repeated the analysis using only in-frame sequencing reads (Table S3) by, prior to BAM sorting and indexing, filtering the associated BAM files for in-frame reads using a custom Python script. After these individual analyses, we then accumulated all nucleotide occurrences for each position across samples into a single value grouped by cell line and averaged these out by dividing by the number of samples (Figure 1C; Figure S1A). For this purpose, we categorized all nucleotide variants and insertions into one class called mutations, while we kept deletions as a separate class. These results were then saved into a table and visualized using GraphPad Prism v8.3.0.

Cut Window Analysis

Based on our SpCas9 mutational outcome analysis, we set a window of 14 bp around the SpCas9 and AsCas12a cut sites (Figure 1D), in which we expect, on average, frequent occurrence of in-frame mutations. We then applied a custom R script to design SpCas9, xCas9-3.7, and SpCas9-NG sgRNA and (en)AsCas12a crRNA libraries targeting their associated PAMs across the protein coding sequences of 190 well-known human anticancer target genes. We used the web-based Ensembl BioMart datamining tool (https://www.ensembl.org/biomart/martview) to obtain the genomic location of each coding exon of a list of 190 genes (Table S4). For this purpose, we used the human genes genome reference Ch38.p13 dataset and applied the BioMart filter tool to select genes of interest using the "Input External References List – gene names" function. For the BioMart output attributes, we selected "Structures" with, under the "Gene" subheading, the "Gene name," "Chromosome/scaffold name," and "Strand" attributes and, under the "Exon" subheading, the "Genomic coding start," "Genomic coding end," "CDS start," and "CDS end" attributes. The BioMart "Results" output was then downloaded as a Microsoft Excel .xls file with the "unique results" checkbox selected. The resulting .xls file was cleaned up by removing duplicate entries, empty rows, rows with emtpy "CDS start"/"CDS end" values, and rows with gene

mappings to alternative haplotypes (e.g., the major histocompatibility complex [MHC] region of chromosome 6). The file was then sorted by the "Genomic coding start" column followed by the "Gene name" column using Microsoft Excel 365 and saved as an .xlsx file. This cleaned .xlsx file was then used in a custom R script (created in RStudio with R 4.0.0) to design the CRISPR sgRNAs/crRNAs. For each sgRNA and crRNA in our libraries we saved the genomic coordinates of the endonuclease cut site and extended these with the 14-bp mutational window to obtain mutational genomic coordinate ranges. These extended coordinate ranges were then compared to the genomic coordinate ranges for each exon of the protein coding sequences of the target genes to determine the fraction of the protein coding sequences that fall into the mutational window of the sgRNA and crRNA libraries. These values were then saved, extracted, and visualized using GraphPad Prism v8.3.0.

Cell Lines

Human parental HAP1 cells were obtained from Horizon Discovery, which validated the parental cell line in-house. HEK293T cells were obtained from ATCC (CRL-3216). Cells were grown at 37°C and 5% CO_2 and passaged every 2–3 days in Iscove's modified Dulbecco's medium (IMDM) (HAP1) or DMEM (HEK293T) supplemented with 10% fetal bovine serum and 20 μg/mL gentamicin. The cells were regularly checked for mycoplasma contamination using the Venor GeM OneStep PCR kit (Minerva Biolabs).

Compounds

KPT-9274 was provided by Karyopharm Therapeutics (Newton, MA, USA). FK866 and ispinesib were obtained from Selleck Chemicals. All compounds were dissolved in DMSO (Sigma-Aldrich).

Generation of AsCas12a and enAsCas12a Stable HAP1 Cell Lines

HAP1 cells stably expressing SpCas9-NG, AsCas12a, or enAsCas12a were generated using the CRISPaint principle.[73] Briefly, the *AAVS1* locus (19q13) was targeted for insertion by transient transfection of HAP1 cells plated into a six-well plate (TPP, 150,000 cells/well) using TurboFectin (OriGene) according to the manufacturer's instructions. Transfection was done with a plasmid encoding for an sgRNA under an hU6 promotor together with active SpCas9 under a cytomegalovirus (CMV) promotor and coupled to mCherry through a T2A (1 μg, pU6-sgRNA-AAVS1-CMV-1× FLAG-mCherry-T2A-SpCas9-NLS_{SV40}) in addition to a donor plasmid containing a sgRNA target site followed by a CMV promotor sequence with the SpCas9-NG or (en)AsCas12a coding sequence coupled to a P2A-hygromycin resistance gene (1 μg, pGL4-AAVS1-CMV-SpCas9-NG/[en]AsCas12a-P2A-HygroR) and a sgRNA plasmid with an hU6-expressed sgRNA targeting the donor plasmid (500 ng, pFrame0 GGG sgRNA donor). Following transfection, cells were incubated for 2 days and then selected over a period of 2 weeks with 300 μg/mL hygromycin B. Resistant cells from all wells were harvested and pooled and then plated as small pools at low density (~100 cells) in 48-well plates to screen for functional SpCas9-NG/(en)AsCas12a clones using green fluorescence on an Incucyte ZOOM after transfection with an in-house-developed SpCas9/AsCas12a traffic-light reporter containing a CMV promotor followed by a mRFP cassette separated from an out-of-frame mNeonGreen cassette by a tandem mirrored SpCas9 or AsCas12a crRNA target sequence (pGL4_CMV-mRFP-2 × $AsCas12a_{target\ site}$-mNeonGreen-T2A-PuroR) and a hU6-coupled crRNA-expressing plasmid targeting this traffic light reporter. Cutting of this site by active endonuclease brings mNeonGreen back into frame, resulting in green fluorescent cells.

Design and Cloning of the Lentiviral Tiling crRNA Libraries

To obtain the *KIF11*, *NAMPT*, and 10-gene crRNA tiling libraries, the coordinates for the *NAMPT*, *ABL1*, *ACTB*, *ERCC3*, *H2BFM2*, *KIF11*, *PAK4*, *RPS3A*, *TP53*, *TUBB*, and *XPO1* coding sequences were extracted from Ensembl BioMart (https://www.ensembl.org/biomart/martview, retrieved December 2017). Duplicate coordinates associated with different gene isoforms were removed and each coordinate was shifted by addition or subtraction of 9–20 additional bases to allow for the design of crRNAs spanning intron-exon boundaries. For SpCas9-NG, these extended coordinates were used to design 20-base-long crRNAs on both the forward and reverse strand using a custom R script utilizing the following PAM sites: AG, TG, CG, GG. In addition, using these extended coordinates, 23-base-long crRNAs for use with (en)AsCas12a were designed on both the forward and reverse strand utilizing the following PAM sites: TTTA, TTTT, TTTG, TTTC, TTCA, TTCT, TTCG, TTCC, TGTC, ATTA, ATTG, ATTC, ATCC, TGTA, CTTA, CTTC, CTCC, GTTA, TACC, GTTC, TCCC, TATC, TATA. crRNA sequences containing a BfuAI recognition site (ACCTGC) or TTTTT repeat were removed from all lists, 30 scrambled crRNAs were added as negative controls, and the remaining crRNAs were then appended, when needed, with the AsCas12a crRNA backbone and additional adaptor sequences to facilitate PCR and BfuAI-mediated cloning (total oligonucleotide length 85 bases [SpCas9-NG] or 129 bases [(en)AsCas12a]). All sequence lists (SpCas9-NG *KIF11*, 1,034 sgRNAs; *NAMPT*, 565 sgRNAs; [en]AsCas12a *NAMPT*, 428 crRNAs; and 10-gene library, 3,838 crRNAs) of appended crRNAs were then sent to CustomArray (Bothell, WA, USA) to synthesize the oligonucleotide pools separately on a 12K chip. The oligonucleotide pools were then amplified in five parallel PCR reactions using 5 ng per reaction in 25-μL reactions with the Phusion high-fidelity PCR master mix with HF buffer (NEB). PCR products were purified with the QIAquick nucleotide removal kit (QIAGEN). Purified PCR products were then subjected to restriction digestion with BfuAI (NEB) overnight at 50°C. Following digestion, 33 ng of the crRNA products was ligated into 500 ng of BfuAI and NsiI predigested pLCKO (Addgene 73311) overnight at 16°C with T4 DNA ligase (NEB). The ligated mixtures were purified with the QIAquick nucleotide removal kit (QIAGEN) and then electroporated into Endura competent cells (Lucigen) with a Gene Pulser System (Bio-Rad) using recommended settings. Electroporated cells were incubated for 1 h in recovery medium (Lucigen) and then plated in 15-cm diameter Petri dishes with prewarmed Luria-Bertani (LB) agar containing 100 μg/mL ampicillin. Cells were grown overnight at 32°C and colonies were counted to estimate the fold representation. All colonies were collected and pooled to perform plasmid extraction

with the PureLink HiPure plasmid maxiprep kit (Invitrogen). The purified pLCKO-U6-crRNA plasmid libraries were then transfected into HEK293T cells together with the lentiviral packaging plasmids pMD2.G (Addgene 12259) and psPAX2 (Addgene 12260) to generate lentiviral particles coated with the vesicular stomatitis virus glycoprotein G (VSV-G) protein. HEK293T cells were plated in 40 mL of supplemented DMEM in T150 (TPP) flasks at 45% confluency and incubated overnight. The following afternoon cells were transfected using X-TremeGENE 9 (Roche) and incubated overnight. The medium was replaced with DMEM supplemented with serum-free BSA growth media (DMEM + 1.1 g/100 mL BSA and 20 μg/mL gentamicin) the next morning to facilitate efficient harvest of the lentiviral particles the following day. Lentiviral stocks were frozen and kept at −80°C. Lentiviral stocks were titrated on wild-type HAP1 cells using the puromycin resistance marker to estimate the multiplicity of infection (MOI).

Mutagenesis Screens to Obtain Drug Resistance

HAP1 cells stably expressing SpCas9-NG, AsCas12a, or enAsCas12a were passaged and cultured in T150 flasks under 300 μg/mL hygromycin B selection prior to lentiviral transduction. Upon the day of transduction, HAP1 cells were collected by trypsin detachment followed by centrifugation at 400 × *g*, resuspended in supplemented IMDM, and then counted using a Luna II automated cell counter (Westburg). For both AsCas12a and enAsCas12a cells, 24×10^6 cells (KPT-9274 and FK866) or 128×10^6 cells (10-gene ispinesib screen) were diluted to 1×10^6 cells/mL in IMDM supplemented with 10% fetal bovine serum (FBS), 20 μg/mL gentamicin, and 8 μg/mL Polybrene. The lentiviral libraries were then transduced at an estimated MOI of 0.3 by 2-h spinfection at 2,000 rpm, 37°C in 12-well tissue culture plates with 2×10^6 cells/well (coverage of ±16,000× per crRNA for KPT-9274/FK866, ±10,000× for ispinesib). After spinfection, cells were cultured overnight at 37°C and 5% CO_2. The next morning, transduced cells were collected by trypsin detachment and plated in supplemented IMDM with 1 μg/mL puromycin in five tissue-cultured TPP T150 flasks with 7.5×10^6 cells/flask (KPT-9274 and FK866) or in a single (FK866/KPT-9274 replicate experiments) or four (ispinesib, both replicates) tissue-cultured 500-cm^2 Nunc square bioassay dishes (Thermo Fisher Scientific) at 37.5×10^6 cells/dish. Cells were allowed to develop mutations for a period of 4–5 days under puromycin selection. Following this period, cells were collected by trypsin detachment and 2×10^6 cells (KPT-9274 or FK866) or 10×10^6 cells (ispinesib) were harvested for genomic DNA (gDNA) extraction at the early time point (day 0). The remaining $AsCas12a^+$ and $enAsCas12a^+$ cells were plated at even amounts while maintaining the desired coverage (10,000–16,000 cells per crRNA) in T150 flasks or 500-cm^2 Nunc square bioassay dishes in supplemented IMDM containing 300 nM KPT-9274, 10 nM FK866, or 15 nM ispinesib. Drug treatment was maintained for 2 weeks, and the medium was refreshed twice to refresh the drug and to remove dead cells. Following this period, surviving colonies were harvested by trypsin detachment and counted. Cells were then pelleted by centrifugation at 400 × *g* and the pellet was reserved for gDNA extraction. Two flasks containing resistant AsCas12a or enAsCas12a colonies obtained during the initial KPT-9274 resistance screen were washed with PBS and the surviving colonies were stained for 1 h with 0.5% (w/v) gentian violet in 50% (v/v) methanol in water. Stained colonies were extensively washed with demineralized water and then visualized using a smartphone-integrated photo camera.

gDNA Extraction and PCR Amplification

gDNA from drug-resistant cells was extracted using the QIAamp DNA mini kit (QIAGEN). To amplify enriched crRNAs, 1,400 ng of DNA was subjected to a 100-μL 24-cycle PCR reaction using Phusion high-fidelity PCR master mix with HF buffer (NEB) coupled with pLCKO primers carrying Nextera adaptor sequences. 5 μL of the PCR-amplified DNA was then amplified in a second PCR using the CloneAmp HiFi polymerase kit (Clontech) during 18 cycles with Nextera XT indexing primers (Illumina). The indexed PCR products were purified using magnetic Agencourt AMPure XP beads (Beckman Coulter) according to the manufacturer's instructions. For targeted amplicon next-generation sequencing, the region of interest was first amplified during 24 cycles in 25-μL PCR reactions containing 50 ng of gDNA with Phusion high-fidelity PCR master mix with HF buffer (NEB) and with custom primers containing adaptor regions for Nextera indexes (Integrated DNA Technologies). Amplified DNA was purified with the QIAquick PCR purification kit (QIAGEN), and 1.5–2 μL of this DNA was PCR amplified during 25 cycles with CloneAmp HiFi PCR premix (Clontech) using indexing primers containing P5 and P7 Illumina adapters in 25-μL reactions to index the samples. Indexed samples were purified using magnetic Agencourt AMPure XP beads (Beckman Coulter) and eluted in Tris-EDTA (TE) buffer. Samples were then diluted to 2–4 nM and pooled to form the initial library.

Three-Day Cell Viability Assay

3,000 HAP1 cells/well were plated in 96-well plates containing a $\log_{10}$ dilution series of KPT-9274, ispinesibm or DMSO in supplemented IMDM. Cells were incubated for 72h at 37°C and 5% CO_2. After incubation, cell viability was measured using the CellTiter 96 AQueous non-radioactive cell proliferation Assay (Promega) according to the manufacturer's instructions, and colorimetric signals were measured with a Safire2 (Tecan). The assay was performed in triplicate and each experiment was repeated at least once. Signals were adjusted by the background signal and normalized to the DMSO control. Relative data values were then visualized and analyzed using a log-based four-parameter model (GraphPad Prism v8.3.0).

Individual Validation of Enriched crRNAs

For individual validation of guides enriched during the drug resistance screens, each enriched crRNA was appended with the AsCas12a crRNA backbone and a terminator (TTTTTT) and adaptor sequences to facilitate cloning into the pLentiGuide plasmid (Addgene 52963) following the standard cloning protocol. HAP1 cells were plated into a six-well plate containing supplemented IMDM and incubated overnight at 37°C and 5% CO_2. The next day, cells were cotransfected transiently with TurboFectin (OriGene) with the individually cloned crRNAs and a plasmid expressing (en)AsCas12a from a CMV

promotor according to the manufacturer's instructions. After 3 days, cells were washed with PBS and then incubated with the desired drug (300 nM KPT-9274, 10 nM FK866, 15 nM ispinesib) in supplemented IMDM for a period of 5 days. Cells were then washed with PBS and resistant colonies were imaged with an IncuCyte S3. Images were extracted and converted to black-and-white images with increased contrast using the. GNU Image Manipulation Program (GIMP 2.8)

Visualization of the Drug Binding Site of KPT-9274 and Ispinesib

To visualize the drug-binding site and the uncovered resistance mutations within this site, the co-crystal structures of KPT-9274 with NAMPT (Protein Data Bank [PDB]: 5NSD) and ispinesib with KIF11 (PDB: 4AP0) were downloaded from the RCSB PDB (http://www.rcsb.org/).[5,44,74] The crystal structures were visualized manually using YASARA View[75] and exported as ray-traced images in the .png format.

Next-Generation Sequencing

The concentration of purified Nextera-indexed PCR samples was measured using a NanoPhotometer small-volume spectrophotometer (Westburg). Indexed samples were diluted to 4 nM. Compatible indexed PCR products were then pooled at 4 nM. This library was then denatured and diluted (12.5 pM) according to the instructions for paired-end sequencing on a MiSeq (Illumina) with a MiSeq v2, 500-cycle kit (Illumina) and 10% PhiX v3 (Illumina) spike-in. Sequencing reads of the genetic screens were obtained by next-generation sequencing on a MiSeq (Illumina) machine using 250-bp paired-end sequencing with Nextera XT indexes (Illumina). Raw Fastq files were trimmed for adapters using the MiSeq Reporter software (Illumina). The R package EdgeR was used to count crRNA sequences,[76,77] and the fold change of enriched sgRNAs was calculated as follows:

$$\mathrm{Log}_2\left(\frac{1 + (\mathrm{ReadCountperMillionReads})_{\mathrm{afterdrugtreatment}}}{1 + (\mathrm{ReadCountperMillionReads})_{\mathrm{afterpuromycinselection}}}\right).$$

Targeted exon sequencing was performed by mapping paired FastQ reads to the reference sequence with Geneious R11.4.[78] Mappings were extracted as BAM files and analyzed using the R package CrispRVariants to determine and count nucleotide variants.[24] Nucleotide variants (detection threshold 0.1%) were then imported into Geneious R11.4 to determine amino acid changes. Visualizations of all genetic screening results was done using GraphPad Prism v8.3.0.

Data Availability

The used R scripts (Data S1) and plasmid maps (Data S2) are available in Supplemental Information. Next-generation sequencing data files are available upon request.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.ymthe.2020.09.025.

AUTHOR CONTRIBUTIONS

Conceptualization and Design, J.E.N. and D.D.; Development of Methodology, J.E.N., M.J., and B.P.K.; Software, T.D. and J.E.N.; Formal Analysis: J.E.N.; Investigation and Validation, J.E.N. and M.J.; Resources, B.P.K. and M.J.; Data Curation, J.E.N. and M.J.; Writing – Original Draft, J.E.N., M.J., and D.D.; Writing – Review & Editing, J.E.N., M.J., T.D., B.P.K., H.J.T., and D.D.; Visualization, J.E.N., B.P.K, and M.J.; Supervision, D.D. and H.J.T.; Project Administration, D.D.; Funding Acquisition, D.D.

CONFLICTS OF INTEREST

J.E.N. and D.D. are inventors on a patent filed by the KU Leuven regarding the use of CRISPRres or CRISPR-directed evolution. B.P.K. is a consultant for Avectas and ElevateBio, an advisor to Acrigen Biosciences, and is an inventor on various patents and patent applications that describe genome engineering technologies.

ACKNOWLEDGMENTS

We thank Jae Keith Young for scientific input and technical support. We also thank Lotte Bral, Kristien Minner, and Nathalie Thys for excellent technical support. We are grateful to the KU Leuven Genomics Core and Sarah Gillemot for technical support with Illumina-based next-generation sequencing. J.E.N. is funded by a fellowship award from the Damon Runyon Foundation for Cancer Research (DRG: 2384-19). B.P.K. is supported by NIH R00-CA218870.

REFERENCES

1. Moffat, J.G., Vincent, F., Lee, J.A., Eder, J., and Prunotto, M. (2017). Opportunities and challenges in phenotypic drug discovery: an industry perspective. Nat. Rev. Drug Discov. *16*, 531–543.

2. Bunnage, M.E., Gilbert, A.M., Jones, L.H., and Hett, E.C. (2015). Know your target, know your molecule. Nat. Chem. Biol. *11*, 368–372.

3. Nijman, S.M. (2015). Functional genomics to uncover drug mechanism of action. Nat. Chem. Biol. *11*, 942–948.

4. Kapoor, T.M., and Miller, R.M. (2017). Leveraging chemotype-specific resistance for drug target identification and chemical biology. Trends Pharmacol. Sci. *38*, 1100–1109.

5. Neggers, J.E., Kwanten, B., Dierckx, T., Noguchi, H., Voet, A., Bral, L., Minner, K., Massant, B., Kint, N., Delforge, M., et al. (2018). Target identification of small molecules using large-scale CRISPR-Cas mutagenesis scanning of essential genes. Nat. Commun. *9*, 502.

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

7. Shen, M.W., Arbab, M., Hsu, J.Y., Worstell, D., Culbertson, S.J., Krabbe, O., Cassa, C.A., Liu, D.R., Gifford, D.K., and Sherwood, R.I. (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. Nature *563*, 646–651.

8. Allen, F., Crepaldi, L., Alsinet, C., Strong, A.J., Kleshchevnikov, V., De Angeli, P., Páleníková, P., Khodak, A., Kiselev, V., Kosicki, M., et al. (2018). Predicting the mutations generated by repair of Cas9-induced double-strand breaks. Nat. Biotechnol. *37*, 64–72.

9. van Overbeek, M., Capurso, D., Carter, M.M., Thompson, M.S., Frias, E., Russ, C., Reece-Hoyes, J.S., Nye, C., Gradia, S., Vidal, B., et al. (2016). DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. Mol. Cell *63*, 633–646.

10. Chuai, G., Yang, F., Yan, J., Chen, Y., Ma, Q., Zhou, C., Zhu, C., Gu, F., and Liu, Q. (2016). Deciphering relationship between microhomology and in-frame mutation occurrence in human CRISPR-based gene knockout. Mol. Ther. Nucleic Acids *5*, e323.

11. Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80–84.

12. Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87.

13. Vinyard, M.E., Su, C., Siegenfeld, A.P., Waterbury, A.L., Freedy, A.M., Gosavi, P.M., Park, Y., Kwan, E.E., Senzer, B.D., Doench, J.G., et al. (2019). CRISPR-suppressor scanning reveals a nonenzymatic role of LSD1 in AML. Nat. Chem. Biol. *15*, 529–539.

14. Ipsaro, J.J., Shen, C., Arai, E., Xu, Y., Kinney, J.B., Joshua-Tor, L., Vakoc, C.R., and Shi, J. (2017). Rapid generation of drug-resistance alleles at endogenous loci using CRISPR-Cas9 indel mutagenesis. PLoS ONE *12*, e0172177.

15. Donovan, K.F., Hegde, M., Sullender, M., Vaimberg, E.W., Johannessen, C.M., Root, D.E., and Doench, J.G. (2017). Creation of novel protein variants with CRISPR/Cas9-mediated mutagenesis: turning a screening by-product into a discovery tool. PLoS ONE *12*, e0170445.

16. Pettitt, S.J., and Lord, C.J. (2019). Dissecting PARP inhibitor resistance with functional genomics. Curr. Opin. Genet. Dev. *54*, 55–63.

17. Butt, H., Eid, A., Momin, A.A., Bazin, J., Crespi, M., Arold, S.T., and Mahfouz, M.M. (2019). CRISPR directed evolution of the spliceosome for resistance to splicing inhibitors. Genome Biol. *20*, 73.

18. Zhang, X., Yue, D., Wang, Y., Zhou, Y., Liu, Y., Qiu, Y., Tian, F., Yu, Y., Zhou, Z., and Wei, W. (2019). PASTMUS: mapping functional elements at single amino acid resolution in human cells. Genome Biol. *20*, 279.

19. Kahn, J.D., Miller, P.G., Silver, A.J., Sellar, R.S., Bhatt, S., Gibson, C., McConkey, M., Adams, D., Mar, B., Mertins, P., et al. (2018). *PPM1D*-truncating mutations confer resistance to chemotherapy and sensitivity to PPM1D inhibition in hematopoietic cells. Blood *132*, 1095–1105.

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

21. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

22. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759–771.

23. Kleinstiver, B.P., Sousa, A.A., Walton, R.T., Tak, Y.E., Hsu, J.Y., Clement, K., Welch, M.M., Horng, J.E., Malagon-Lopez, J., Scarfò, I., et al. (2019). Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. Nat. Biotechnol. *37*, 276–282.

24. Lindsay, H., Burger, A., Biyong, B., Felker, A., Hess, C., Zaugg, J., Chiavacci, E., Anders, C., Jinek, M., Mosimann, C., and Robinson, M.D. (2016). CrispRVariants charts the mutation spectrum of genome engineering experiments. Nat. Biotechnol. *34*, 701–702.

25. Gerstung, M., Beisel, C., Rechsteiner, M., Wild, P., Schraml, P., Moch, H., and Beerenwinkel, N. (2012). Reliable detection of subclonal single-nucleotide variants in tumour cell populations. Nat. Commun. *3*, 811.

26. 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.

27. 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.

28. Hasmann, M., and Schemainda, I. (2003). FK866, a highly specific noncompetitive inhibitor of nicotinamide phosphoribosyltransferase, represents a novel mechanism for induction of tumor cell apoptosis. Cancer Res. *63*, 7436–7442.

29. Khan, J.A., Tao, X., and Tong, L. (2006). Molecular basis for the inhibition of human NMPRTase, a novel target for anticancer agents. Nat. Struct. Mol. Biol. *13*, 582–588.

30. Abu Aboud, O., Chen, C.H., Senapedis, W., Baloglu, E., Argueta, C., and Weiss, R.H. (2016). Dual and specific inhibition of NAMPT and PAK4 by KPT-9274 decreases kidney cancer growth. Mol. Cancer Ther. *15*, 2119–2129.

31. Aboukameel, A., Muqbil, I., Senapedis, W., Baloglu, E., Landesman, Y., Shacham, S., Kauffman, M., Philip, P.A., Mohammad, R.M., and Azmi, A.S. (2017). Novel p21-activated kinase 4 (PAK4) allosteric modulators overcome drug resistance and stemness in pancreatic ductal adenocarcinoma. Mol. Cancer Ther. *16*, 76–87.

32. Takao, S., Chien, W., Madan, V., Lin, D.C., Ding, L.W., Sun, Q.Y., Mayakonda, A., Sudo, M., Xu, L., Chen, Y., et al. (2018). Targeting the vulnerability to NAD^+ depletion in B-cell acute lymphoblastic leukemia. Leukemia *32*, 616–625.

33. Mitchell, S.R., Larkin, K., Grieselhuber, N.R., Lai, T.H., Cannon, M., Orwick, S., Sharma, P., Asemelash, Y., Zhang, P., Goettl, V.M., et al. (2019). Selective targeting of NAMPT by KPT-9274 in acute myeloid leukemia. Blood Adv. *3*, 242–255.

34. Li, N., Lopez, M.A., Linares, M., Kumar, S., Oliva, S., Martinez-Lopez, J., Xu, L., Xu, Y., Perini, T., Senapedis, W., et al. (2019). Dual PAK4-NAMPT inhibition impacts growth and survival, and increases sensitivity to DNA-damaging agents in Waldenström macroglobulinemia. Clin. Cancer Res. *25*, 369–377.

35. Hovstadius, P., Larsson, R., Jonsson, E., Skov, T., Kissmeyer, A.M., Krasilnikoff, K., Bergh, J., Karlsson, M.O., Lönnebo, A., and Ahlgren, J. (2002). A phase I study of CHS 828 in patients with solid tumor malignancy. Clin. Cancer Res. *8*, 2843–2850.

36. Fulciniti, M., Martinez-Lopez, J., Senapedis, W., Oliva, S., Lakshmi Bandi, R., Amodio, N., Xu, Y., Szalat, R., Gulla, A., Samur, M.K., et al. (2017). Functional role and therapeutic targeting of p21-activated kinase 4 in multiple myeloma. Blood *129*, 2233–2245.

37. Kim, H.K., Lee, S., Kim, Y., Park, J., Min, S., Choi, J.W., Huang, T.P., Yoon, S., Liu, D.R., and Kim, H.H. (2020). High-throughput analysis of the activities of xCas9, SpCas9-NG and SpCas9 at matched and mismatched target sequences in human cells. Nat. Biomed. Eng. *4*, 111–124.

38. Legut, M., Daniloski, Z., Xue, X., McKenzie, D., Guo, X., Wessels, H.H., and Sanjana, N.E. (2020). High-throughput screens of PAM-flexible Cas9 variants for gene knockout and transcriptional modulation. Cell Rep. *30*, 2859–2868.e5.

39. 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.

40. Rath, O., and Kozielski, F. (2012). Kinesins and cancer. Nat. Rev. Cancer *12*, 527–539.

41. Kasap, C., Elemento, O., and Kapoor, T.M. (2014). DrugTargetSeqR: a genomics- and CRISPR-Cas9-based method to analyze drug targets. Nat. Chem. Biol. *10*, 626–628.

42. Wang, W., Elkins, K., Oh, A., Ho, Y.C., Wu, J., Li, H., Xiao, Y., Kwong, M., Coons, M., Brillantes, B., et al. (2014). Structural basis for resistance to diverse classes of NAMPT inhibitors. PLoS ONE *9*, e109366.

43. Essletzbichler, P., Konopka, T., Santoro, F., Chen, D., Gapp, B.V., Kralovics, R., Brummelkamp, T.R., Nijman, S.M., and Bürckstümmer, T. (2014). Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. Genome Res. *24*, 2059–2065.

44. Talapatra, S.K., Schüttelkopf, A.W., and Kozielski, F. (2012). The structure of the ternary Eg5-ADP-ispinesib complex. Acta Crystallogr. D Biol. Crystallogr. *68*, 1311–1319.

45. Chiruvella, K.K., Liang, Z., and Wilson, T.E. (2013). Repair of double-strand breaks by end joining. Cold Spring Harb. Perspect. Biol. *5*, a012757.

46. Silva, G., Poirot, L., Galetto, R., Smith, J., Montoya, G., Duchateau, P., and Pâques, F. (2011). Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. Curr. Gene Ther. *11*, 11–27.

47. Wright, A.V., Nuñez, J.K., and Doudna, J.A. (2016). Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. Cell *164*, 29–44.

48. Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. *32*, 347–355.

49. Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera, Mdel.C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat. Biotechnol. *32*, 267–273.

50. Bae, S., Kweon, J., Kim, H.S., and Kim, J.S. (2014). Microhomology-based choice of Cas9 nuclease target sites. Nat. Methods *11*, 705–706.

51. Miller, S.M., Wang, T., Randolph, P.B., Arbab, M., Shen, M.W., Huang, T.P., Matuszek, Z., Newby, G.A., Rees, H.A., and Liu, D.R. (2020). Continuous evolution of SpCas9 variants compatible with non-G PAMs. Nat. Biotechnol. *38*, 471–481.

52. Wang, J., Meng, X., Hu, X., Sun, T., Li, J., Wang, K., and Yu, H. (2019). xCas9 expands the scope of genome editing with reduced efficiency in rice. Plant Biotechnol. J. *17*, 709–711.

53. Brier, S., Lemaire, D., DeBonis, S., Forest, E., and Kozielski, F. (2006). Molecular dissection of the inhibitor binding pocket of mitotic kinesin Eg5 reveals mutants that confer resistance to antimitotic agents. J. Mol. Biol. *360*, 360–376.

54. 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.

55. 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.

56. Rees, H.A., and Liu, D.R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. Nat. Rev. Genet. *19*, 770–788.

57. Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science *353*, aaf8729.

58. Huang, T.P., Zhao, K.T., Miller, S.M., Gaudelli, N.M., Oakes, B.L., Fellmann, C., Savage, D.F., and Liu, D.R. (2019). Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. Nat. Biotechnol. *37*, 626–631.

59. Sakata, R.C., Ishiguro, S., Mori, H., Tanaka, M., Tatsuno, K., Ueda, H., Yamamoto, S., Seki, M., Masuyama, N., Nishida, K., et al. (2020). Base editors for simultaneous introduction of C-to-T and A-to-G mutations. Nat. Biotechnol. *38*, 865–869.

60. Grünewald, J., Zhou, R., Lareau, C.A., Garcia, S.P., Iyer, S., Miller, B.R., Langner, L.M., Hsu, J.Y., Aryee, M.J., and Joung, J.K. (2020). A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing. Nat. Biotechnol. *38*, 861–864.

61. Zhang, X., Zhu, B., Chen, L., Xie, L., Yu, W., Wang, Y., Li, L., Yin, S., Yang, L., Hu, H., et al. (2020). Dual base editor catalyzes both cytosine and adenine base conversions in human cells. Nat. Biotechnol. *38*, 856–860.

62. Hess, G.T., Frésard, L., Han, K., Lee, C.H., Li, A., Cimprich, K.A., Montgomery, S.B., and Bassik, M.C. (2016). Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat. Methods *13*, 1036–1042.

63. Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat. Methods *13*, 1029–1035.

64. Chen, H., Liu, S., Padula, S., Lesman, D., Griswold, K., Lin, A., Zhao, T., Marshall, J.L., and Chen, F. (2020). Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor. Nat. Biotechnol. *38*, 165–168.

65. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., and Liu, D.R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature *576*, 149–157.

66. Neggers, J.E., Vercruysse, T., Jacquemyn, M., Vanstreels, E., Baloglu, E., Shacham, S., Crochiere, M., Landesman, Y., and Daelemans, D. (2015). Identifying drug-target selectivity of small-molecule CRM1/XPO1 inhibitors by CRISPR/Cas9 genome editing. Chem. Biol. *22*, 107–116.

67. Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., Wang, Y., Anzalone, A.V., Raguram, A., Doman, J.L., et al. (2020). Prime genome editing in rice and wheat. Nat. Biotechnol. *38*, 582–585.

68. Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K.M., Gregg, A., Noggle, S., and Tessier-Lavigne, M. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature *533*, 125–129.

69. Erdogan, M., Fabritius, A., Basquin, J., and Griesbeck, O. (2020). Targeted *in situ* protein diversification and intra-organelle validation in mammalian cells. Cell Chem. Biol. *27*, 610–621.e5.

70. Halperin, S.O., Tou, C.J., Wong, E.B., Modavi, C., Schaffer, D.V., and Dueber, J.E. (2018). CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. Nature *560*, 248–252.

71. Tou, C.J., Schaffer, D.V., and Dueber, J.E. (2020). Targeted diversification in the *S. cerevisiae* genome with CRISPR-guided DNA polymerase I. ACS Synth. Biol. *9*, 1911–1916.

72. DeWeirdt, P.C., Sanson, K.R., Sangree, A.K., Hegde, M., Hanna, R.E., Feeley, M.N., Griffith, A.L., Teng, T., Borys, S.M., Strand, C., et al. (2020). Optimization of AsCas12a for combinatorial genetic screens in human cells. Nat. Biotechnol. Published online July 13, 2020. https://doi.org/10.1038/s41587-020-0600-6.

73. Schmid-Burgk, J.L., Höning, K., Ebert, T.S., and Hornung, V. (2016). CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. Nat. Commun. *7*, 12338.

74. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242.

75. Krieger, E., Koraimann, G., and Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. Proteins *47*, 393–402.

76. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

77. Dai, Z., Sheridan, J.M., Gearing, L.J., Moore, D.L., Su, S., Wormald, S., Wilcox, S., O'Connor, L., Dickins, R.A., Blewitt, M.E., and Ritchie, M.E. (2014). edgeR: a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens. F1000Res. *3*, 95.

78. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.