Optimization of AsCas12a for combinatorial genetic screens in human cells

Peter C DeWeirdt[1,*], **Kendall R Sanson**[1,*], **Annabel K Sangree**[1,*], **Mudra Hegde**[1,*], **Ruth E Hanna**[1], **Marissa N Feeley**[1], **Audrey L Griffith**[1], **Teng Teng**[2], **Samantha M Borys**[1], **Christine Strand**[1], **J Keith Joung**[3,4,5,6], **Benjamin P Kleinstiver**[6,7,8], **Xuewen Pan**[2], **Alan Huang**[2], **John G Doench**[1,9]

[1]Genetic Perturbation Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA

[2]Tango Therapeutics, 100 Binney St., Cambridge, MA, USA

[3]Molecular Pathology Unit, Massachusetts General Hospital, Charlestown, MA, USA

[4]Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA, USA

[5]Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, MA, USA

[6]Department of Pathology, Harvard Medical School, Boston, MA, USA

[7]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

[8]Department of Pathology, Massachusetts General Hospital, Boston, MA, USA

Abstract

Cas12a RNA-guided endonucleases are promising tools for multiplexed genetic perturbations because they can process multiple guide RNAs expressed as a single transcript and subsequently

cleave target DNA. However, their widespread adoption has lagged behind Cas9-based strategies due to low activity and the lack of a well-validated pooled screening toolkit. Here we describe the optimization of enhanced AsCas12a (enAsCas12a) for pooled, combinatorial genetic screens in human cells. By assaying the activity of thousands of guides, we refine on-target design rules and develop a comprehensive set of off-target rules to predict and exclude promiscuous guides. We also identify 38 direct repeat variants that can substitute for the wild-type sequence. We validate our optimized AsCas12a toolkit by screening for synthetic lethalities in OVCAR8 and A375 cancer cells, discovering an interaction between *MARCH5* and *WSB2*. Finally, we show that enAsCas12a delivers similar performance to Cas9 in genome-wide dropout screens but at greatly reduced library size, which will facilitate screens in challenging models.

Editorial summary

Improved Cas12a variants and sgRNA design rules enhance genome-wide screens.

CRISPR technology has injected new life into the field of functional genomics, with its robust on-target activity, acceptable off-target profile, and a myriad of derivations that allow manipulation of the genome beyond genetic knockout[1]. The Cas9 enzyme from *Streptococcus pyogenes* (SpCas9) is the tool of choice for most screening experiments, but delivery of multiple guide RNAs for pooled screening remains a challenge. The need for individual expression cassettes for each additional guide often requires multi-step cloning and customized sequencing readouts[2]; further, such arrangements are prone to recombination[3] and uncoupling[4,5] when made into lentivirus and retrieved by PCR. Thus, there is a need for a simpler system for perturbing multiple genes simultaneously.

The Cas12a family of CRISPR enzymes (previously known as Cpf1) potentially offers a better approach, because an array of guides can be expressed from a single transcript, separated only by a 20 nucleotide direct repeat (DR) sequence, and Cas12a can both process individual guides and execute target DNA cleavage[6]. The compact design of AsCas12a arrays confers significant advantages when synthesizing and sequencing DNA: for Cas12a, a second guide requires only 43 additional nucleotides, whereas a second Cas9 guide cassette is 346 nucleotides (Figure 1a). Several Cas12a orthologs have shown activity in human cells, including variants derived from *Acidaminococcus* (AsCas12a) and *Lachnospiraceae* (LbCas12a)[6], and the nucleotide preferences for both of these orthologs have been broadly assessed[7,8]. Further, LbCas12a has been developed for CRISPRa approaches[9] and base editing[10], whereas AsCas12a has undergone protein engineering to increase efficacy and is thus another option for CRISPRa and base editing approaches[11]. Other Cas12a orthologs have also seen continued development[12,13].

To date, the demonstrated multiplexing ability of Cas12a has been underwhelming. Even when delivered by transient transfection, which is generally a poor surrogate for the single copy lentivirus needed for pooled screens[14], proof-of-concept multiplexing assays have demonstrated low rates of editing[15,16]. One recent combinatorial genetic screen relied on stringent positive selection[17], which makes assessment of efficiency difficult. More encouragingly, Liu and colleagues[18] created a genome-wide library for dropout screens with

multiple guides per gene on the same vector, to reduce the number of cells required for such screens, although performance still fell short of optimized Cas9 libraries.

Here we optimize AsCas12a for pooled genetic screens in human cells. We tested several variants of AsCas12a to select a highly efficient construct, validated on-target sequence rules for selecting guides, and derived rules for minimizing off-target effects. Additionally, we developed alternative direct repeat sequences for use with multiplexed arrays, and demonstrate triple knockout using these variant sequences. We conducted a genetic interaction screen of genes implicated in apoptosis, revealing both known synthetic lethal interactions as well as a previously-uncharacterized interaction between *MARCH5* and *WSB2*, two genes implicated in protein degradation. Finally, we built a compact genome-wide library that leverages the multiplexing ability of Cas12a to reduce the number of unique library elements while maintaining multiple guides per gene, and demonstrate comparable performance to optimized Cas9 libraries, but with substantially fewer cells. In sum, we present a complete set of experimental and computational tools to enable the effective use of AsCas12a for pooled screening.

RESULTS

Optimization of the AsCas12a protein

To assess the potential of AsCas12a for pooled screens, we acquired a previously-described lentiviral vector[7], which has a single nuclear localization sequence (NLS), referred to here as 1xNLS-Cas12a (Figure 1b). We also constructed a second lentiviral vector with two NLSs (2xNLS-Cas12a). This served as a template for a third vector, from which we generated enCas12a, a recently-described variant of AsCas12a with increased activity and expanded PAM preferences[11].

We synthesized a library of all possible guides ranging between 20 and 23 nucleotides in length targeting TTTN PAMs across 43 genes. To test guide activity, we targeted pan-essential genes[19], cell-specific lethal genes[20], and genes with well-characterized small molecule interactions[21]; cell surface genes served as negative controls. Additionally, all guides regardless of PAM were included for the essential gene *EEF2*, for a total of 12,472 guides (Supplementary Table 1). We also synthesized an analogous version of this library for use with SpCas9 (NGG PAM), to enable direct comparison. The AsCas12a library was cloned into a modified version of lentiGuide (pRDA_052). Viability and drug resistance screens were conducted in duplicate (Figure 1c, Supplementary Table 2, Supplementary Data 1), as done previously to assess activity of SaCas9[22] and SpCas9[21].

We first examined the viability data from A375 cells. For each AsCas12a variant, we observed similar performance across all lengths of guides (Supplementary Figure 1a). Further, we saw the highest activity at TTTA PAM sites followed closely by TTTC and TTTG, and the lowest activity levels at TTTT sites (Supplementary Figure 1b), as observed previously[7]. Defining the 5th percentile of negative control guides as an activity cutoff, 57% of enCas12a guides targeting a TTTT site were active compared with 7% of 2xNLS-Cas12a and 7% of 1xNLS-Cas12a, demonstrating the ability of enCas12a to target previously

inaccessible sites[7,11]. Guides effective in one cell line tended to be effective in another (Supplementary Figure 1c).

To compare the efficacy of AsCas12a to SpCas9, we perfored ROC-AUC analysis, defining guides targeting essential genes as true positives (Figure 1d). SpCas9 performed well, with an AUC of 0.93, whereas the 1xNLS-Cas12a construct performed the worst (AUC = 0.61). An additional NLS site improved performance (2xNLS-Cas12a AUC = 0.84), consistent with recent observations[23]. Finally, enCas12a performed comparably to SpCas9 (AUC = 0.96). Two additional cell lines, MELJUSO and HT29, trended similarly (Figure 1e). We observed similar levels of 2xNLS-Cas12a and enCas12a expression (Supplementary Figure 2), suggesting that protein activity, rather than stability, explains the increased performance of enCas12a.

Evaluation of on-target scoring criteria

We used tiling screen data to evaluate a published deep learning model, Seq-DeepCpf1[8], selecting the top half of essential and control guides by predicted on-target activity. The ROC-AUC between essential and negative control guides increased substantially for 2xNLS-Cas12a across all cell lines – 0.96 (+0.09), 0.94 (+0.10) and 0.94 (+0.09) for MELJUSO, A375 and HT29 cell lines respectively – and to a lesser extent for enCas12a - 0.98 (+0.01), 0.97 (+0.01) and 0.97 (+0.01) (Figure 2a, Supplementary Figure 3). We binned 2xNLS-Cas12a guides by their Seq-DeepCpf1 score and observed that 95%, 92%, and 89% of guides with a score greater than 60 were active in MELJUSO, A375 and HT29, respectively (Figure 2b). Thus, applying Seq-DeepCpf1 scores to select guides for 2xNLS-Cas12a can substantially increase activity.

One major advantage of enCas12a is the broader range of PAM sites; the PAM for AsCas12a, TTTV, will occur on average once every 42.67 nts. The additional enCas12a PAMs were originally classified into three tiers[11], with tier 1 being most active. To evaluate these alternative PAMs we used the set of *EEF2* guides targeting non-TTTN PAMs, which covered 249 out of 264 possible PAMs with at least one guide. In agreement with previous results, we saw a relationship between the assigned tier and the measured activity (Supplementary Figure 4a). Notably, the guides targeting TTTN (62% active) showed similar efficacy to the other tier 1 PAMs (68% active). We used these data to develop a preliminary on-target predictor for enCas12a, Seq-DeepCpf1_mod1 (Supplementary Note 1).

Examination of enCas12a PAM preferences

To better quantify the PAM preferences of enCas12a under screening conditions, we designed a library tiling essential genes utilizing all PAMs assigned to tier 1, 2, or 3, as well as 10 randomly-selected inactive (no tier) PAMs (Supplementary Data 2). Using the fifth percentile of negative control guides targeting non-essential genes as an activity cutoff, 63% of guides with a tier 1 PAM scored as active, in contrast to 96% with a TTTV PAM (Figure 2c). Notably, guides targeting a TTCC PAM showed the highest activity for non-TTTV PAMs (88% active). We also identified two PAMs, ATTA and GTTA, originally classified as tier 2, which were the fifth and sixth most active PAM sites, with 81% and 80% active

respectively, demonstrating the importance of using large-scale readouts to characterize PAM activity with many unique guide sequences. Importantly, only 2% of no tier PAMs were active, which is below the 5% non-essential cutoff we used, demonstrating that enCas12a maintains PAM specificity and avoids genome-wide toxicity.

Using the PAM tiling data, we tested new models for predicting enCas12a activity. We tested both a convolutional neural network[8] and a gradient boosted model[21] on a hold out set of guides tiling across 9 genes (Figure 2d). Both models trained with the PAM tiling data had a higher spearman correlation than the original and mod1 versions of Seq-DeepCpf1 (Figure 2e). Mod1 predicted the unseen PAM sequences better than the original version, validating the *in silico* substitution we used for our interim model. Interestingly, a gradient boosted model trained using the 15,000 indel frequencies from Kim et al. outperformed Seq-DeepCpf1 on the holdout test set, potentially indicating that this model is less prone to overfitting. Thus while the gradient boosted model and convolutional network performed similarly when trained on the PAM tiling data, we decided to move forward with the gradient boosted model ("enPAM+GB") for better generalizability (Supplementary Note 1, Supplementary Fig. 5).

With 2xNLS-Cas12a, 27.1% of genes in the human genome have 0 – 5 guides with a Seq-DeepCpf1 score of >0.6 (Figure 2b), whereas only 0.5% of genes fall into that category for enCas12a using the enPAM+GB model, at a threshold of 0.7 (Figure 2f). At the same on-target scoring thresholds, 90.5% of genes have 20 or more guides for enCas12a, compared to only 26.0% for Cas12a. Thus, the combination of increased activity of enCas12a at canonical TTTV PAMs, along with the expansion of PAM sites, leads to much greater flexibility in guide selection.

Off-target predictions for Cas12a

In order to assess the off-target tolerance of AsCas12a, we constructed a library of guides intentionally mismatched to their target. We selected 300 of the most active guides from the viability, vemurafenib, and 6-thioguanine screens with the on-target tiling library screened with 2xNLS-Cas12a. We then designed every possible single and double nucleotide mismatch for these guides, selecting a random subset of the latter, resulting in a library with 300 perfect match guides, 19,512 single mismatch guides, 20,000 double mismatch guides, and 217 guides targeting cell surface genes as a set of negative controls (Figure 3a). We performed screens in duplicate in A375 cells with dropout, vemurafenib, and 6-thioguanine arms in cells stably expressing 2xNLS-Cas12a and enCas12a (Supplementary Data 3). In order to determine the fraction of guides that were active in this library, we mapped the distribution of guides targeting essential genes in dropout assays and set an activity cutoff at the 5th percentile of controls (Figure 3b). As expected, we found that perfect match guides were highly active with both constructs: 94% and 91% with 2xNLS-Cas12a and enCas12a, respectively. EnCas12a showed a higher tolerance for mismatches, with 54% and 16% of guides active for single and double mismatch guides respectively, compared with 28% and 9% for 2xNLS-Cas12a. Thus, enCas12a shows more propensity for off-target activity, as initially described[11].

We then calculated the fraction of active guides for each mismatch type and position to generate a cutting frequency determination (CFD) matrix, as done previously with SpCas9[21]. These matrices were similar across experimental conditions (Supplementary Fig 6a), so we merged the data to create a single CFD matrix for each Cas12a (Figure 3c). When we compared the CFD values for 2x-NLS-Cas12a and enCas12a, we saw a monotonic relationship (Supplementary Fig 6b), indicating similar specificity preferences. We consistently observed a higher tolerance for mismatches at the PAM distal end of the guide, as well as for rG:dT mismatches, which are two trends observed previously with other techniques to examine off-target activity of Cas12a enzymes[24,25], and that have also been seen with SpCas9[21,26].

To compare CFD matrices between the two AsCas12a constructs and SpCas9, we used data from our previously published SpCas9 CFD matrix[21]. Since SpCas9 guides are designed as 20mers, we focused on the 20 most PAM-proximal nucleotides of AsCas12a guides. We then ranked each mismatch and position by activity and saw that SpCas9 and enCas12a had similar activity levels across their profiles, whereas 2xNLS-Cas12a was the least promiscuous (Figure 3d), consistent with previous examinations of the specificity of Cas12a enzymes by orthogonal techniques[24]. Thus, although enCas12a is more promiscuous than 2xNLS-Cas12a, its specificity is comparable to that of SpCas9, suggesting it is suitable for genetic screens. Although not evaluated here, a high fidelity version of enCas12a has been developed[11], which may be useful when off-target effects are a substantial concern.

Finally, we used the CFD matrix to predict the activity of double mismatch guides. We calculated the CFD score as the product of the activities of each individual mismatch, an approach that has been validated by others to identify problematic off-target sites with more than one mismatch for SpCas9[27,28]. To evaluate this model, we binned double mismatch guides targeting essential genes into predicted quartiles and plotted the distribution of measured log2-fold-change values in the dropout screens (Figure 3e). We saw more activity from guides with higher CFD scores, indicating that this model can help identify problematic multi-mismatch off-target sites when designing guides. Notably, the largest portion of double mismatch guides fell into the lowest quartile of CFD scores (Figure 3e), which indicates that consideration of off-target activities does not overly restrict the set of guide RNAs available to target a gene of interest.

Development of variant direct repeat sequences

Multiplexing with Cas12a requires delivery of multiple direct repeats (DRs), separated by only the 20 – 23 nt guide sequence. To minimize homology in the array, we sought to find positions in the DR sequence that could be modified without reducing activity. We designed a library of 35,682 alternative DRs with up to 3 variable basepairs in the stem and 3 variable nucleotides in the single-stranded or loop regions. To assay the efficacy of these DRs in a negative selection screen, we targeted *BCL2L1* and *MCL1*, a known synthetic lethal pair[2,22,29], such that only effective use of both guide RNAs in the same cell should lead to cell death. We cloned the library of DR sequences into two vectors, with both orientations of the *MCL1* and *BCL2L1* guides (Figure 4a) and screened in MELJUSO cells expressing 2xNLS-Cas12a (Supplementary Data 4, Supplementary Figure 7).

This approach identified 38 sequences with wildtype-like activity across the two constructs (Figure 4b). Several positions were intolerant of any changes, whereas others tolerated certain nucleotide substitutions (Figure 4c). For example, position 1 tolerated A or G but not C, whereas active constructs were observed with all nucleotides at position 12 and 14. Interestingly, this observation in the 12th position of the loop agrees with alignment of direct repeat sequences across Cas12a orthologs[30]. All recovered active sequences maintained basepairing in the stem, but basepairing alone was insufficient for activity, as the nucleotide sequence proved important. At the base of the stem, a T-A basepair could be replaced with a C-G basepair, but no other orientations, indicating a preference for a pyrimidine on one side and a purine on the other. Thus, we identified a consensus sequence for active variant DRs (Figure 4d), and numerous examples thereof, which can be used to minimize repetitive sequences in multiplexed arrays.

Triple knockout with AsCas12a

Encouraged by the identification of variant DRs for higher order multiplexing, we assessed the capacity of AsCas12a to target three endogenous loci simultaneously. We designed guides targeting the cell surface markers *CD47, CD63* and *B2M*, which are highly expressed in A375 cells and do not show a viability effect upon knockout in the DepMap[31]. We generated 6 multiplexed arrays with all possible guide orientations (Figure 4e), separated by the wildtype DR followed by variants 1 and 3. We also created versions of constructs 2 and 6 utilizing only the wild-type DR. As controls, we included 3 single guide constructs and one empty vector, and used flow cytometry (Supplementary Figure 8) to quantify the fraction of edited cells. Irrespective of guide position, approximately 70% of cells showed editing with enCas12a at all three loci at the three timepoints assayed (Figure 4f, Supplementary Figure 9a); previous attempts at multiplexed editing observed less than 5% triplex editing in individual cells[16]. Further, we saw slightly diminished activity for array 6 when using three wildtype DRs (Figure 4g), recommending the use of variant DRs for higher-order multiplexing.

As the *B2M* single guide did not perform as well as the *CD47* or *CD63* guides, we were curious how the double knockout efficacy of *CD47* and *CD63* would be affected if we first gated by the B2M negative population (Supplementary Figure 9b). We found that *B2M*-negative cells were much more likely to also be double knockout for *CD47* and *CD63* (94% double knockout, array 3) than the cells that were *B2M*-positive (58%, array 3). This suggests a useful strategy, in which one of the guides targets a selectable marker gene such as *HPRT1*[32] or *ATP1A1*[33], to enrich for edited cells.

To assess potential DNA damage induced toxicity of simultaneously targeting three genes, we performed a competition assay. EGFP labels cells without enCas12a and we measured the fraction of enCas12a-positive, EGFP-negative cells over time by flow cytometry (Supplementary Figure 10a). The fraction of EGFP-negative cells decreased by 18 – 41% over time with single guide constructs relative to an empty vector control (Supplementary Figure 10b, c), evidence of the cutting toxicity inherent to using Cas enzymes[34,35]; triple guide constructs showed a 31 – 55% decrease. Thus, the impact of additional dsDNA breaks in minor in this cell type, although should be monitored when using any nuclease, especially

in combinatorial applications. We note that others have reported no cutting-based toxicity withCas12a[18], although the competition assay used here may be more sensitive than prior assessments. Nuclease-deactivated versions of AsCas12a fused to KRAB have been described in mammalian cells[36], as have transcriptional activators[9,11,16,36], while will be useful in models where toxicity due to numerous dsDNA breaks is overly confounding.

Multiplexing to assay synthetic lethal interactions

To assess the effectiveness of multiplexing in a large scale screen, we targeted 11 gene pairs previously identified as synthetic lethal[2,22,37]. For each gene in a given pair, we randomly selected up to 20 guides, 3/4 and 1/4 of which targeted TTTV and TTTT PAMs, respectively (Figure 5a). We synthesized guide pairs in both orientations, and separated each with the top three variant DRs described above; each gene pair was thus assessed with up to 2,400 unique constructs (20 guides x 20 guides x 2 orientations x 3 DRs). To account for single guide effects we also paired each guide with 25 guides targeting olfactory receptors as controls.

Following PCR and sequencing, we calculated the log2-fold-change (LFC) relative to pDNA and saw good correlation between replicates in both cell lines (Supplementary Figure 11a, b, Supplementary Data 5). To evaluate how position affects guide activity, we compared the average LFCs of the target-control constructs with the reverse orientation control-target constructs. The LFCs of target guides in position 1 were well correlated with the LFCs in position 2 for all three DR variants in both OVCAR8 and A375 (Figure 5b, Supplementary Figure 11c), indicating that guides perform similarly regardless of position in the array and the DR sequence.

To quantitate synergies between guide pairs, we first calculated an expected phenotype by determining the average LFC of each targeting guide when paired with controls (Supplementary Figure 12a), and then summing the resulting LFC values for the two guides. We then fit a line between the expected and observed LFCs using constructs with one control and one target guide (Figure 5c). Finally, we calculated the residual from the fit line for all constructs, where a negative residual indicates a synthetic lethal pair. We saw a large fraction of constructs with residuals two standard deviations below the mean of control-target pairs in both OVCAR8 and A375 cells, indicating synthetic lethal interactions. To assess the efficacy of on-target rules in this experimental setting, we filtered each gene for the top half of guides by enPAM+GB score. We saw an increase in the fraction of synthetic lethal constructs for a majority of gene pairs (Supplementary Figure 12b, c, d), confirming the utility of this algorithm. Additionally, enCas12a identified a higher fraction of synthetic lethal guide pairs than 2xNLS-Cas12a (Supplementary Figure 12e).

For some gene pairs we saw a very high fraction of guide pairs score as synthetic lethal. For example, over 75% of guide pairs for *STAG1 - STAG2* and *HDAC1 - HDAC2* scored in both OVCAR8 (Figure 5d) and A375 cells (Supplementary Figure 12b). To compare enCas12a to previous synthetic lethal screens conducted with the "Big Papi" approach using SaCas9 and SpCas9[22], we first reanalyzed the latter with the same analytical methodology to account for differences in library design strategy (see Methods). For *BCL2L1 - MCL1*, a gene pair we previously validated with small molecule inhibitors, the magnitude of the residuals is substantially greater with enCas12a in OVCAR8 cells (Figure 5e) and comparable in A375

cells (Supplementary Figure 13a). For *MAPK1 - MAPK3*, 98% of guide pairs scored as synthetic lethal in OVCAR8 cells with enCas12a, with a greater magnitude of residuals than with the Cas9 approach (Supplementary Figure 13a). In contrast, we do not observe synthetic lethality in A375 cells for this gene pair, likely due to the strong viability effect caused by loss of *MAPK1* alone in A375 cells (Supplementary Figure 12a). The correlation between the residuals from both approaches, albeit from a limited set of gene pairs, suggests that cell line-specific synergies largely reflect biological differences (Supplementary Figure 13b). Overall, these results show that enCas12a is at least comparable, and in some circumstances substantially more potent, than a Cas9-based approach. Given the considerable ease of generating and sequencing Cas12a combinatorial libraries, we suspect that this will become the technology of choice for such screens.

Combinatorial screen of apoptotic genes

Encouraged by this initial screen, we designed a library in which we assayed all combinations of 22 genes previously identified as synthetic lethal with the anti-apoptotic genes *BCL2L1* or *MCL1*[38] or deeply implicated in apoptosis via prior literature (Figure 6a). We designed this library to include up to 10 guides per gene in dual orientation, each paired with 20 guides targeting olfactory receptors to serve as negative controls, for a total of 26,082 unique constructs. We screened these combinations in OVCAR8 and A375 cells expressing enCas12a (Supplementary Data 6) and fit a linear model to determine residuals (Figure 6b).

Z-scores were well correlated across cell lines and negative controls were centered around 0, establishing a low false positive rate for detecting interactors (Figure 6c), with *BCL2L1 - MCL1* emerging as a top synthetic lethal hit. Most of the buffering interactions were observed with constructs targeting the same gene, as others have reported[2]; the interactions between *BCL2L1* and both *BAX* and *PMAIP1* have been observed previously[22]. Furthermore, we saw excellent agreement with prior small molecule and knockout screens focused on *BCL2L1* and *MCL1* conducted with genome-wide Cas9 libraries[38], indicating a high true positive and low false negative rate (Supplementary Figure 14). When we generated a gene interaction network, drawing edges at an absolute Z-score cutoff of 5, *BCL2L1* was the most central node for both A375 and OVCAR8 conditions with 6 and 7 interactors respectively, followed by *MCL1* with 3 interactors in both contexts (Figure 6d). We identified a dense set of interactors including *BCL2L1* and *MCL1* along with the *MARCH5,* encoding an E3 ubiquitin ligase, and *WSB2*, encoding a SOCS box-containing protein. Notably, in co-essentiality networks, *MCL1* is the top correlate of *MARCH5* in the Dependency Map (R = 0.68) and *MCL1* is the fourth correlate of *WSB2* (R = 0.25), highlighting their strong functional relationship to apoptosis[39–41]. The interactions between *MCL1 - BCL2L1, MCL1 - WSB2,* and *BCL2L1 - MARCH5* have been observed previously using orthogonal approaches[22,38], while the *MARCH5 - WSB2* interaction is novel (Figure 6e). Together these results nominate *MARCH5* and *WSB2* as important genes for focused validation and mechanistic studies. More broadly, these results demonstrate the high sensitivity and specificity of enCas12a for combinatorial screening and demonstrate its utility for uncovering novel genetic interactions.

enCas12a genome-wide libraries

The ease of multiplexing with Cas12a enables the construction of genome-wide libraries with multiple guides per gene in a single construct to reduce library size[18]. We designed the Humagne library with four guides per gene, divided them into set A and set B, and included two guides per gene on each construct (Figure 7a). We performed viability screens in duplicate with these libraries in A375, MELJUSO, and HT29 cells expressing enCas12a (Figure 7b, Supplementary Data 7, 8). We also conducted screens with the Brunello Cas9 library, to enable direct comparison (Supplementary Data 9).

To benchmark library performance, we conducted precision-recall analysis using guides targeting well-validated sets of essential and non-essential genes[19,42]. At a threshold of 95% precision for the screens conducted in A375 cells, Humagne set A and set B recalled 88% and and 83% of essential genes, respectively, while Brunello recalled 81% (Figure 7c); performance was similar in HT29 and MELJUSO cells (Supplementary Figure 15a). Comparing constructs targeting the same gene across Humagne set A and set B, Pearson correlations ranged from 0.65 – 0.71 across the three cell lines (Supplementary Figure 15b). After averaging across all constructs targeting a gene for the Humagne and Brunello libraries, we observed that the two Cas enzymes were well correlated (Pearson r = 0.78 – 0.84) suggesting that there were no systematic biases introduced by either enzyme (Figure 7d, Supplementary Figure 16a). Likewise, Gene Set Enrichment Analysis showed highly similar depletion of KEGG gene sets (Pearson r = 0.90 – 0.92) (Figure 7e, Supplementary Figure 16b). We also compared Brunello and Humagne to other genome-wide Cas9[43,44] and Cas12a[18] libraries, and saw that they are top performers by this precision-recall metric (Figure 7f).

That sets A and B of the Humagne library are well-correlated, and that they perform as well as the best Cas9 libraries on a per construct basis, suggests that a particularly efficient screening strategy would be to conduct one replicate of each set. Such an approach, with a total of two reagents per gene each screened once, would require one-fourth the number of cells of a typical Cas9 screen, with four reagents per gene each screened in duplicate. To test this, we analyzed the above data by averaging individual constructs to arrive at a gene-level answer, and again conducted precision-recall analysis, using either one or two replicates of each library (Figure 7g, Supplementary Data 10). While the top performance was achieved with two replicates of the Brunello library (recall of 93.6%), one replicate each of Humagne set A and B was nearly as effective (90.7%). For comparison, two replicates of the Avana (4 guides per gene) and 4 replicates of GeCKOv2 libraries (6 per gene) screened in the DepMap showed an average recall of 79.2% and 76.8%, respectively. Such an approach will be of particular value when working with challenging models, such as *in vivo* screens, primary cells, and organoid cultures.

DISCUSSION

Here we present the development of AsCas12a for large scale genetic screens in human cells. We show that highly efficient gene disruption, under conditions of single-copy lentiviral integration, is obtainable with optimized expression constructs, direct repeat sequences, and guide selection rules. We identify variant direct repeat sequences that

maintain wildtype activity while minimizing vector homology, which may reduce recombination and shuffling that can occur during lentiviral production[3,5,22,45]. Using these tools, we demonstrate that triple knockouts can be achieved in a majority of targeted cells, and conduct proof-of-principle combinatorial screens, uncovering a synthetic lethal relationship between *MARCH5* and *WSB2*. Finally, we show that Cas12a-based genome-wide libraries offer comparable performance to best-in-class Cas9 libraries, but with a four-fold reduction in screening scale.

The use of enCas12a proved critical to these efforts, both for its improved on-target activity and increased density of available PAM sequences. We validate the Seq-DeepCpf1 on-target scoring algorithm, show that it can be extended to enCas12a, and further improve upon enCas12a guide design with a new on-target scoring approach, enPAM+GB. Additionally, we screen a library of tens of thousands of mismatched guides to develop off-target profiles for AsCas12a constructs, which will enable selection of specific guides. The combination of on- and off-target selection rules for AsCas12a are available via CRISPick (https://broad.io/crispick).

One clear use-case for an optimized AsCas12a toolkit is combinatorial screens to study genetic interactions. The largest GI map to date comprises pairwise interactions among 472 genes[46], representing 0.05% of all possible interactions of human protein-coding genes. The massive landscape to be explored recommends the simplicity of synthesizing and sequencing AsCas12a guide constructs compared to Cas9-based approaches, and suggests that the former will become the preferred tool for these studies. Further, higher-order multiplexing will be particularly useful to study gene paralogs, where targeting only one may not reveal a phenotype[47]. Likewise, targeting the same gene with multiple guides in the same construct is an attractive screening approach, especially when cell numbers are limiting. In sum, the results presented here help to establish AsCas12a as a top choice for many applications of genetic screens in human cells.

METHODS

Life Sciences Reporting Summary is hereby referenced.

Vectors

Lenti-AsCpf1-Blast: 1xNLS-Cas12a lentiviral expression construct; also known as pRDA_113; Addgene 84750

pRDA_112: 2xNLS-Cas12a lentiviral expression construct; also known as pTG_12; Addgene 136475

pRDA_174: enCas12a lentiviral expression construct; modified version of pRDA_112 by introduction of point mutations; Addgene 136476

pRDA_052: modified version of pLentiGuide for expression of AsCas12a guides; Addgene 136474

pRosetta: lentiviral construct for expression of EGFP, puromycin resistance, and blasticidin resistance; Addgene 59700

pRosetta_v2: modification of pRosetta to include a hygromycin resistance cassette; also known as pRDA_018; Addgene 136477

Guide Sequences

Individual guides used in this study are provided in Supplementary Table 3. Sequences for pooled libraries are found in Supplementary Data 1 - 9.

Library production—Oligonucleotide pools were synthesized by CustomArray and Twist. BsmBI recognition sites were appended to each guide RNA sequence (whether single guides or tandem guides) along with the appropriate overhang sequences (bold italic) for cloning into the plasmid pRDA_052, as well as primer sites to allow differential amplification of subsets from the same synthesis pool. The final oligonucleotide sequence was thus: 5'-[Forward Primer]CGTCTCA**AGAT**[guide RNA]TTTTTT***GAAT***CGAGACG[Reverse Primer].

Primers were used to amplify individual subpools using 25 μL 2x NEBnext PCR master mix (New England Biolabs), 2 μL of oligonucleotide pool (~40 ng), 5 μL of primer mix at a final concentration of 0.5 μM, and 18 μL water. PCR cycling conditions: 30 seconds at 98°C, 30 seconds at 53°C, 30 seconds at 72°C, for 24 cycles. For combinatorial libraries, the number of cycles should be reduced to 3, to prevent swapping. In cases where a library was divided into subsets unique primers could be used for amplification:

Primer Set; Forward Primer, 5' – 3'; Reverse Primer, 5' – 3'

1; AGGCACTTGCTCGTACGACG; ATGTGGGCCCGGCACCTTAA

2; GTGTAACCCGTAGGGCACCT; GTCGAGAGCAGTCCTTCGAC

3; CAGCGCCAATGGGCTTTCGA; AGCCGCTTAAGAGCCTGTCG

4; CTACAGGTACCGGTCCTGAG; GTACCTAGCGTGACGATCCG

5; CATGTTGCCCTGAGGCACAG; CCGTTAGGTCCCGAAAGGCT

6; GGTCGTCGCATCACAATGCG; TCTCGAGCGCCAATGTGACG

The resulting amplicons were PCR-purified (Qiagen) and cloned into the library vector via Golden Gate cloning with Esp3I (Fisher Scientific) and T7 ligase (Epizyme); the library vector was pre-digested with BsmBI (New England Biolabs). The ligation product was isopropanol precipitated and electroporated into Stbl4 electrocompetent cells (Life Technologies) and grown at 30°C for 16 hours on agar with 100 μg mL^{-1} carbenicillin. Colonies were scraped and plasmid DNA (pDNA) was prepared (HiSpeed Plasmid Maxi, Qiagen). To confirm library representation and distribution, the pDNA was sequenced.

Lentivirus production—For small-scale virus production, the following procedure was used: 24 h before transfection, HEK293T cells were seeded in 6-well dishes at a density of 1.5e6 cells per well in 2 mL of DMEM + 10% FBS. Transfection was performed using TransIT-LT1 (Mirus) transfection reagent according to the manufacturer's protocol. Briefly, one solution of Opti-MEM (Corning, 66.25 μL) and LT1 (8.75 μL) was combined with a DNA mixture of the packaging plasmid pCMV_VSVG (Addgene 8454, 250 ng), psPAX2 (Addgene 12260, 1,250 ng), and the transfer vector (e.g., pLentiGuide, 1,250 ng). The solutions were incubated at room temperature for 20–30 min, during which time fresh media was added to the HEK293T cells. After this incubation, the transfection mixture was added dropwise to the surface of the HEK293T cells, and the plates were centrifuged at 1,000g for 30 min at room temperature. Following centrifugation, plates were transferred to a 37 °C incubator for 6–8 h, after which the media was removed and replaced with DMEM + 10% FBS media supplemented with 1% BSA and 1% penicillin/streptomycin. Virus was harvested 36 h after this media change.

A large-scale procedure was used for pooled library production. 20–24 h before transfection, 1.8e7 HEK293T cells were seeded in a T-175 tissue culture flask. The transfection was performed similarly to the small-scale production; 6 mL of Opti-MEM, 305 μL of LT1, and a DNA mixture of pCMV_VSVG (5 μg), psPAX2 (50 μg), and 40 μg of the transfer vector were used per reaction. Flasks were transferred to a 37 °C incubator for 6–8 h; after this, the media was aspirated and replaced with BSA-supplemented media. Virus was harvested 36 h after this media change.

Cell culture

A375, OVCAR8, MELJUSO, 786O, HT29, and A549 cells were obtained from the Cancer Cell Line Encyclopedia. HEK293Ts were obtained from ATCC (CRL-3216).

All cell lines were routinely tested for mycoplasma contamination and were maintained without antibiotics except during screens, when the media was supplemented with 1% penicillin/streptomycin. Cell lines were kept in a 37 °C humidity-controlled incubator with 5.0% carbon dioxide and were maintained in exponential phase growth by passaging every 2–4 days. The following media conditions and doses of polybrene, puromycin, and blasticidin, respectively, were used:

A375: RPMI + 10% FBS; 1 μg mL^{-1}; 1 μg mL^{-1}; 5 μg mL^{-1}

HEK293T: DMEM + 10% FBS; N/A; N/A; N/A

HT29: DMEM + 10% FBS; 1 μg mL^{-1}; 2 μg mL^{-1}; 8 μg mL^{-1}

MELJUSO: RPMI + 10% FBS; 4 μg mL^{-1}; 1 μg mL^{-1}; 4 μg mL^{-1}

OVCAR8: RPMI + 10% FBS; 4 μg mL^{-1}; 1 μg mL^{-1}; 8 μg mL^{-1}

A549: DMEM + 10% FBS; 1 μg mL^{-1}; 1.5 μg mL^{-1}; 5 μg mL^{-1}

786O: RPMI + 10% FBS; 4 μg mL^{-1}; 1 μg mL^{-1}; 8 μg mL^{-1}

Vemurafenib (S1267) and talazoparib (BMN-673) were obtained from Selleckchem. 6-thioguanine was obtained from Sigma-Aldrich. Olaparib (10621) was obtained from Cayman Chemical Co. A-1331852 (A-6048) was obtained from Active Biochem. S63845 was a gift from Guo Wei.

Determination of antibiotic dose

In order to determine an appropriate antibiotic dose for each cell line, cells were transduced with the pRosetta or pRosetta_v2 lentivirus such that approximately 30% of cells were infected and therefore EGFP+. At least 1 day post-transduction, cells were seeded into 6-well dishes at a range of antibiotic doses (e.g. from 0 μg/mL to 8 μg/mL of puromycin). The rate of antibiotic selection at each dose was then monitored by performing flow cytometry for EGFP+ cells. For each cell line, the antibiotic dose was chosen to be the lowest dose that led to at least 95% EGFP+ cells after antibiotic treatment for 7 days (for puromycin) or 14 days (for blasticidin and hygromycin).

Determination of lentiviral titer

To determine lentiviral titer for transductions, cell lines were transduced in 12-well plates with a range of virus volumes (e.g. 0, 150, 300, 500, and 800 μL virus) with 3.0×10^6 cells per well in the presence of polybrene. The plates were centrifuged at 640 x g for 2 h and were then transferred to a 37 °C incubator for 4–6 h. Each well was then trypsinized, and an equal number of cells seeded into each of two wells of a 6-well dish. Two days post-transduction, puromycin was added to one well out of the pair. After 5 days, both wells were counted for viability. A viral dose resulting in 30–50% transduction efficiency, corresponding to an MOI of ~0.35–0.70, was used for subsequent library screening.

Staining for Cas12a expression

Cas12a protein expression on an individual cell basis was determined by detecting a C'terminal HA tag. To prepare samples, cells were fixed (ab185917) for 15 minutes at room temperature, washed with chilled PBS, permeabilized (ab185917) and stained with PE anti-HA (Biolegend #901518) for 30 minutes at room temperature, washed with chilled PBS twice to remove residual antibody, and resuspended in flow buffer (PBS, 2% FBS, 5μM EDTA). Cells were visualized by flow cytometry on the BDAccuri C6 Sampler. Gates were set off of the A375 stained parental population.

Pooled screens

Cell lines stably expressing Cas12a were transduced with guides cloned into the pRDA_052 vector in two cell culture replicates at a low MOI (~0.5). Transductions were performed with enough cells to achieve a representation of at least 500 cells per guide per replicate, taking into account a 30–50% transduction efficiency. Throughout the screen, cells were split at a density to maintain a representation of at least 500 cells per guide, and cell counts were taken at each passage to monitor growth. Puromycin selection was added 2 days post-transduction and was maintained for 5–7 days. After puromycin selection was complete, each replicate was divided into untreated (i.e. no drug / dropout arms) and small molecule treatment arms, each at a representation of at least 500 cells per guide. 14 days after the

initiation of small molecule treatment, cells were pelleted by centrifugation, resuspended in PBS, and frozen promptly for genomic DNA isolation.

Genomic DNA isolation and sequencing

Genomic DNA (gDNA) was isolated using either the KingFisher Flex Purification System with the Mag-Bind® Blood & Tissue DNA HDQ Kit (Omega Bio-Tek #M6399–01), or the Machery Nagel NucleoSpin Blood Maxi (2e7–1e8 cells), Midi (5e6–2e7 cells), or Mini (<5e6 cells) kits as per the manufacturer's instructions. The gDNA concentrations were quantitated by Qubit. For PCR amplification, gDNA was divided into 100 μL reactions such that each well had at most 10 μg of gDNA. Per 96 well plate, a master mix consisted of 150 μL ExTaq DNA Polymerase (Takara), 1 mL of 10x Ex Taq buffer, 800 μL of dNTP provided with the enzyme, 50 μL of P5 stagger primer mix (stock at 100 μM concentration), and 2 mL water. Each well consisted of 50 μL gDNA plus water, 40 μL PCR master mix, and 10 μL of a uniquely barcoded P7 primer (stock at 5 μM concentration). For future experiments, we recommend the use of Titanium Taq DNA Polymerase (Takara) and the addition of 5% DMSO per well, as we have found that these changes improve PCR efficiency.

PCR cycling conditions: an initial 1 min at 95 °C; followed by 30 s at 94 °C, 30 s at 52.5 °C, 30 s at 72 °C, for 28 cycles; and a final 10 min extension at 72 °C. PCR primers were synthesized at Integrated DNA Technologies (IDT). PCR products were purified with Agencourt AMPure XP SPRI beads according to manufacturer's instructions (Beckman Coulter, A63880). Samples were sequenced on a HiSeq2500 HighOutput (Illumina) with a custom primer of sequence: 5'-CTTGTGGAAAGGACGAAACACCGGTAATTTCTACTCTTGTAGAT. The first nucleotide sequenced with the primer is the first nucleotide of the guide RNA, which will contain a mix of all four nucleotides, and thus staggered primers are not required to maintain diversity when using this approach. Reads were counted by alignment to a reference file of all possible guide RNAs present in the library. The read was then assigned to a condition (e.g. a well on the PCR plate) on the basis of the 8 nt index included in the P7 primer. Due to poor sequencing quality of the enCas12a and Cas12a tiling screens in both MELJUSO and HT29, and of the genome-wide screen of Humagne in MELJUSO, we allowed for a single nucleotide mismatch during sequence alignment in these five cases. The corresponding pDNA was deconvoluted in the same manner to allow for accurate internal comparison.

Screen analysis

Following deconvolution, the resulting matrix of read counts was first normalized to reads per million within each condition by the following formula: read per guide RNA / total reads per condition x 1e6. Reads per million was then log2-transformed by first adding one to all values, which is necessary in order to take the log of guides with zero reads. For each guide, the log2-fold-change from plasmid DNA (pDNA) was then calculated. All reported log2-fold-changes for dropout screens are relative to pDNA; for positive selection screens with small molecules, the log2-fold-change are calculated relative to the dropout arm (i.e. no small molecule treatment).

On-target modeling

To pick guides we filtered for genes that were significantly more active than non-essential control genes (one-tailed T-test Bonferroni corrected p-value < 1e-5).

Gradient boosting was done as previously with Cas9[21,22]. We trained a gradient boosted regression model using the Python library *scikit-learn*, with early stopping after 20 iterations of no improvement for a 10% validation split. Guides were featurized with position dependent and independent 1 and 2-mers, GC content, and melting temperature calculated with *biopython*.

Convolutional neural networks were built using the same architecture as Seq-DeepCpf1[8]. Networks were trained using the *keras* submodule of *tensorflow* for 200 epochs or until the model showed no improvement for 20 epochs on a 10% validation split.

In-silico mutagenesis

To understand the learned nucleotide features of the on-target models, we started with a random seed sequence, and then for each iteration we randomly changed one nucleotide from the previous step. We ensured each new sequence was unique. We iterated sequences over 30,000 iterations. Then we scored each sequence and took the difference between its score and the score of the sequence from the previous step, yielding 29,999 differences. Then, for each position and nucleotide, we calculated an average difference for the substitution. Average differences were plotted using the R library ggseqlogo.

CFD score calculation

The Cutting Frequency Determination (CFD) score for off-target activity was calculated as described previously[21]. The calculation uses the observed activity at single mismatch sites to predict cutting at multiple mismatch sites. For example, to calculate the CFD score of a double mismatch guide, with an rG-dA mismatch at position 7 and an rC-dT mismatch at position 10, multiply the measured activity for each single mismatch site, using the matrix presented in Figure 3c.

Triple knockout with Cas12a

A375 cells stably expressing enCas12a were transduced with 6 triple knockout arrays, 3 single knockout constructs, and one empty control vector. Two days after transduction, cells were selected with puromycin (1μg/mL), and selected on puromycin for 7 days. Cells were visualized by flow cytometry on the BDAccuri C6 Sampler. To prepare samples for visualization, cells were stained with FITC anti-human CD47 (Biolegend # 323106), PE anti-human CD63 (Biolegend #353004) and APC anti-human β2-microglobulin (Biolegend #316312) antibodies, diluted 1:100 in flow buffer (PBS, 2% FBS, 5μM EDTA), incubated for 30 min on ice, washed with flow buffer twice to remove residual antibody, and resuspended in flow buffer. Flow cytometry data were analyzed using FlowJo (v10). Gates were set such that ~1% of cells score as knockout in the control condition. Compensation was applied using single stained empty vector control and triple stained empty vector control cell populations. Zero, single, double, and triple knockout populations were quantified using boolean gating in FlowJo.

EGFP Competition assay

A375 cells previously engineered[48] to express EGFP and SpCas9 were mixed at a 50:50 ratio with A375 cells expressing enCas12a (pRDA_174) but with no fluorescent marker. This mixed population of cells was then transduced with 8 constructs: 3 single guides, 2 triple knockout arrays, 2 triple cutting controls arrays, and one empty vector control. Puromycin selection was added 1 day post transduction, and was maintained for 5 days. The fraction of EGFP-positive cells was monitored by flow cytometry (BD Accuri C6 Sampler) at every cell passage.

Calculating expected LFCs

For a construct with the ordered elements guide1, DR, guide2, we say

$$Avg.\ LFC^{expected}_{guide1,\ DR,\ guide2} = m*\left(Avg.\ LFC_{guide1,\ DR,\ controls} + Avg.\ LFC_{controls,\ DR,\ guide2}\right) + \beta,$$

where m and β are the fitted slope and intercept. The residual then is the difference between the observed and expected LFCs.

SynLet libraries data filtering

We used data from the Big Papi SynLet screen in A375 and OVCAR8 cells to compare with the multiplexed enCas12a system. Both libraries targeted the pairs *MAPK1/MAPK3, PARP1/PARP2, BCL2L1/MCL1, BCL2L1/BCL2L2, MAP2K1/MAP2K2,* and *BRCA1/PARP1.* We filtered both data sets to account for the heterogeneity of the library designs. After filtering we had 18 observations for each programmed pair (3 guides for each gene in both orientations), and each guide was paired with 15 cutting controls. In the enCas12a library, because we limited design to only TTTN PAMs to enable comparison to 2xNLS-Cas12a, we were left with fewer than 3 guides for *MAP2K1, MAP2K2,* and *BCL2L2* so we removed these genes from the comparison.

To mitigate off-target effects in the enCas12a library we removed guides that were predicted to cut in alternative protein-coding regions between 20% and 100% of the time (Tier I, Bins I and II in the GPP sgRNA design tool). We also excluded guides that cut in the first 5% or last 20% of the coding sequence of a gene. Then to match the number of guides in the Big Papi library we used enPAM+GB to pick the 3 or 15 best remaining guides for synthetic lethal query genes or control genes, respectively.

We filtered the Big Papi library for the same target pairs (3 guides per gene) as well as 15 control guides targeting the cell surface marker *CD81* (10 guides) and intronic regions of *HPRT1* (5 guides). Note that the original library design for Big Papi already included on and off-target filters.

Scoring some-by-some combinatorial screens

For each "anchor" guide and all of its "target" pairs, we fit the the model $anchor/target\ LFC = m*Median(target/control\ LFC) + \beta$. Where m and β are the fitted slope and intercept. We used olfactory receptor receptor controls for the apoptotic library. The residual for each target guide is then the difference from this line. Negative residuals from

this line indicate a synthetic lethal relationship, whereas positive residuals represent a buffering interaction. To get a gene level score, we first Z-score residuals within each anchor guide, and then Z-score globally for each gene pair.

Library design

The SynLet library of guides was designed without any on-target scoring scheme, selecting from guides with a TTTN PAM. For the apoptosis combinatorial library, guides were selected from those with a Seq-DeepCpf1 score of >25 for on-target activity, using TTTN PAMs.

To design the Humagne genome-wide enCas12a library, we first designed all possible guides targeting protein-coding genes with PAMs in the three tiers defined for enCas12a. The on-target activity score for each guide was then calculated using Seq-DeepCpf1_mod, as enPAM+GB did not yet exist; for deposit in Addgene, we will provide a version designed using enPAM+GB. Off-target sites of these guides across the genome were determined using the CFD algorithm. The top four guides for every gene were then picked using a heuristic that weights their on-target score, off-target matches, and then filtered for their target region along the protein (avoiding the first 5% and last 20% of the coding sequence). The first two guides for the gene were included on the same construct in Humagne set A and the next two in Humagne set B. We also included non-targeting controls and guides that target intergenic sites as targeting controls. Each set has 21,820 guides targeting 20,080 protein-coding genes and 1,740 controls.

DATA AVAILABILITY

The read counts for all screening data and subsequent analyses are provided as Supplementary Data and are currently being deposited with the Sequence Read Archive, SRA: SRP228317.

CODE AVAILABILITY

All custom code used for analysis and notebooks are available on GitHub: https://github.com/PeterDeWeirdt

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

REFERENCES

1. Doench JG Am I ready for CRISPR? A user's guide to genetic screens. Nat. Rev. Genet 19, 67–80 (2018). [PubMed: 29199283]

2. Han K et al. Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. Nat. Biotechnol 35, 463–474 (2017). [PubMed: 28319085]

3. Adamson B et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell 167, 1867–1882.e21 (2016). [PubMed: 27984733]

4. Hegde M, Strand C, Hanna RE & Doench JG Uncoupling of sgRNAs from their associated barcodes during PCR amplification of combinatorial CRISPR screens. PLoS One 13, e0197547 (2018). [PubMed: 29799876]

5. Hill AJ et al. On the design of CRISPR-based single-cell molecular screens. Nat. Methods 15, 271–274 (2018). [PubMed: 29457792]

6. Zetsche B et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759–771 (2015). [PubMed: 26422227]

7. Kim HK et al. In vivo high-throughput profiling of CRISPR-Cpf1 activity. Nat. Methods 14, 153–159 (2017). [PubMed: 27992409]

8. Kim HK et al. Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. Nat. Biotechnol 36, 239–241 (2018). [PubMed: 29431740]

9. Esther Tak Y et al. Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. Nat. Methods (2017) doi:10.1038/nmeth.4483.

10. Li X et al. Base editing with a Cpf1-cytidine deaminase fusion. Nat. Biotechnol 36, 324–327 (2018). [PubMed: 29553573]

11. Kleinstiver BP et al. Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. Nat. Biotechnol (2019) doi:10.1038/s41587-018-0011-0.

12. Tu M et al. A 'new lease of life': FnCpf1 possesses DNA cleavage activity for genome editing in human cells. Nucleic Acids Res. 45, 11295–11304 (2017). [PubMed: 28977650]

13. Tóth E et al. Mb- and FnCpf1 nucleases are active in mammalian cells: activities and PAM preferences of four wild-type Cpf1 nucleases and of their altered PAM specificity variants. Nucleic Acids Res. 46, 10272–10285 (2018). [PubMed: 30239882]

14. Stegmeier F, Hu G, Rickles RJ, Hannon GJ & Elledge SJ A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. Proc. Natl. Acad. Sci. U. S. A 102, 13212–13217 (2005). [PubMed: 16141338]

15. Zetsche B et al. Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. Nat. Biotechnol 35, 31–34 (2016). [PubMed: 27918548]

16. Campa CC, Weisbach NR, Santinha AJ, Incarnato D & Platt RJ Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. Nat. Methods (2019) doi:10.1038/s41592-019-0508-6.

17. Chow RD et al. In vivo profiling of metastatic double knockouts through CRISPR-Cpf1 screens. Nat. Methods (2019) doi:10.1038/s41592-019-0371-5.

18. Liu J et al. Pooled library screening with multiplexed Cpf1 library. Nat. Commun 10, 3144 (2019). [PubMed: 31316073]

19. Hart T, Brown KR, Sircoulomb F, Rottapel R & Moffat J Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. Mol. Syst. Biol 10, 733 (2014). [PubMed: 24987113]

20. Meyers RM et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat. Genet (2017) doi:10.1038/ng.3984.

21. Doench JG et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol 34, 184–191 (2016). [PubMed: 26780180]

22. Najm FJ et al. Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. Nat. Biotechnol 36, 179 (2017). [PubMed: 29251726]

23. Liu P et al. Enhanced Cas12a editing in mammalian cells and zebrafish. Nucleic Acids Res 47, 4169–4180 (2019). [PubMed: 30892626]

24. Kleinstiver BP et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nat. Biotechnol 34, 869–874 (2016). [PubMed: 27347757]

25. Kim D et al. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nat. Biotechnol 34, 863–868 (2016). [PubMed: 27272384]

26. Hsu PD et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol 31, 827–832 (2013). [PubMed: 23873081]

27. Perez AR et al. GuideScan software for improved single and paired CRISPR guide RNA design. Nat. Biotechnol 35, 347 (2017). [PubMed: 28263296]

28. Tycko J et al. Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. Nat. Commun 10, 4063 (2019). [PubMed: 31492858]

29. van Delft MF et al. The BH3 mimetic ABT-737 targets selective Bcl-2 proteins and efficiently induces apoptosis via Bak/Bax if Mcl-1 is neutralized. Cancer Cell 10, 389–399 (2006). [PubMed: 17097561]

30. Teng F et al. Enhanced mammalian genome editing by new Cas12a orthologs with optimized crRNA scaffolds. Genome Biol. 20, 15 (2019). [PubMed: 30717767]

31. Tsherniak A et al. Defining a Cancer Dependency Map. Cell 170, 564–576.e16 (2017). [PubMed: 28753430]

32. Moriarity BS et al. Simple and efficient methods for enrichment and isolation of endonuclease modified cells. PLoS One 9, e96114 (2014). [PubMed: 24798371]

33. Agudelo D et al. Marker-free coselection for CRISPR-driven genome editing in human cells. Nat. Methods 14, 615–620 (2017). [PubMed: 28417998]

34. Aguirre AJ et al. Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. Cancer Discov. 6, 914–929 (2016). [PubMed: 27260156]

35. Munoz DM et al. CRISPR Screens Provide a Comprehensive Assessment of Cancer Vulnerabilities but Generate False-Positive Hits for Highly Amplified Genomic Regions. Cancer Discov. 6, 900–913 (2016). [PubMed: 27260157]

36. Liu Y et al. Engineering cell signaling using tunable CRISPR-Cpf1-based transcription factors. Nat. Commun 8, 2095 (2017). [PubMed: 29235474]

37. Shen JP et al. Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. Nat. Methods 14, 573–576 (2017). [PubMed: 28319113]

38. DeWeirdt PC et al. Genetic screens in isogenic mammalian cell lines without single cell cloning. Nat. Commun 11, 752 (2020). [PubMed: 32029722]

39. Kim E et al. A network of human functional gene interactions from knockout fitness screens in cancer cells. Life Sci Alliance 2, (2019).

40. Pan J et al. Interrogation of Mammalian Protein Complex Structure, Function, and Membership Using Genome-Scale Fitness Screens. Cell Syst 6, 555–568.e7 (2018). [PubMed: 29778836]

41. Wainberg M et al. A genome-wide almanac of co-essential modules assigns function to uncharacterized genes. bioRxiv 827071 (2019) doi:10.1101/827071.

42. Hart T et al. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell 163, 1515–1526 (2015). [PubMed: 26627737]

43. Wang T et al. Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. Cell 168, 890–903.e15 (2017). [PubMed: 28162770]

44. Tzelepis K et al. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. Cell Rep. 17, 1193–1205 (2016). [PubMed: 27760321]

45. Hanna RE & Doench JG A case of mistaken identity. Nat. Biotechnol 36, 802–804 (2018). [PubMed: 30188534]

46. Horlbeck MA et al. Mapping the Genetic Landscape of Human Cells. Cell 174, 953–967.e22 (2018). [PubMed: 30033366]

47. Dede M, McLaughlin M, Kim E & Hart T Multiplex enCas12a screens show functional buffering by paralogs is systematically absent from genome-wide CRISPR/Cas9 knockout screens. bioRxiv 2020.05.18.102764 (2020) doi:10.1101/2020.05.18.102764.

48. Sanson KR et al. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. Nat. Commun 9, 5416 (2018). [PubMed: 30575746]
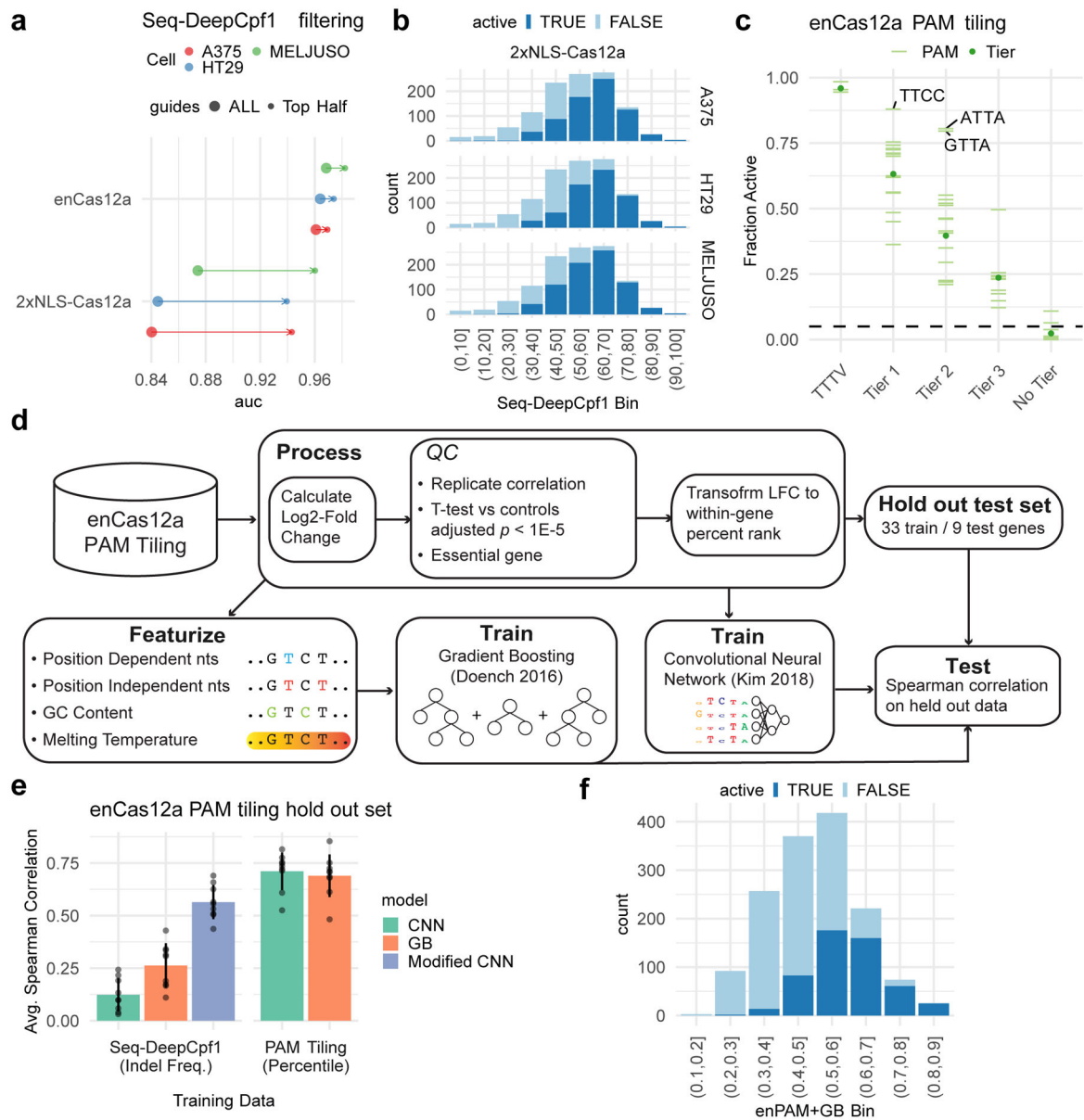
Figure 1.
Optimization of AsCas12a for pooled screens (**a**) Comparison of DNA cassettes necessary for dual knockout with Cas9 versus Cas12a. (**b**) Vector maps for Cas12a constructs. Point mutations for enCas12a are indicated. (**c**) Timeline for executing on-target library tiling screens. (**d**) ROC curves for guides targeting essential and cell surface genes for Cas12a and SpCas9, using viability data in A375 cells (n=1,146 and 2,468 for essential guides for Cas12a and SpCas9, respectively, and 153 and 673 for control guides). The area under the curve for each enzyme is noted in parentheses. (**e**) Same as (d) with 2 additional cell lines, HT29 and MELJUSO.

Figure 2.

On-target design rules for enCas12a (**a**) ROC-AUCs for Cas12a tiling screens improve when filtered for the top half of guides by Seq-DeepCpf1 score. ROC-AUCs were calculated using guides targeting essential genes as true positives (n=1148 for all, n=573 for top half) and cell surface genes as true negatives (n=153 for all, n=76 for top half). (**b**) Seq-DeepCpf1 improves the fraction of active guides for 2xNLS-Cas12a. Guide scores are binned into deciles.. (**c**) The fraction of active guides increases along predefined PAM tier classifications for screens with the enCas12a PAM tiling library. Points represent the cumulative activity for a given tier, lines represent the fraction active for a given PAM. A dashed line is drawn at 5% activity. (**d**) Pipeline for training and testing machine learning models with the enCas12a tiling data. (**e**) Models trained on the PAM tiling data outperform models trained on the Seq-DeepCpf1 indel frequency data. Bars represent the mean spearman correlation on n=9 hold
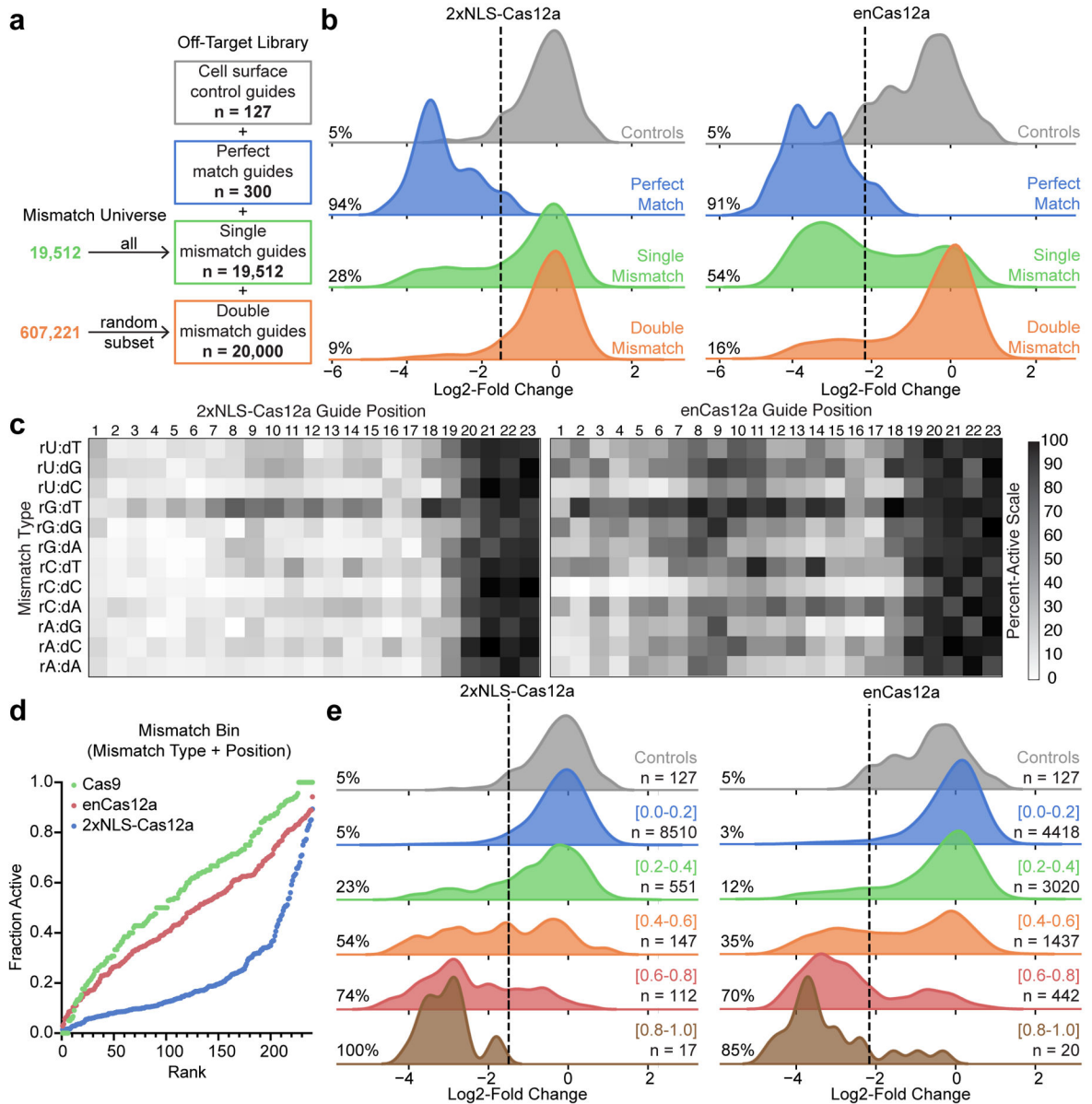
out genes from the PAM tiling data and line ranges represent the standard deviation. CNN denotes Convolutional Neural Network and GB denotes Gradient Boosted regression. (**f**) enPAM+GB improves the fraction of active guides for enCas12a. Guide scores are binned into deciles.

Figure 3.

Prediction of off-target activity for AsCas12a (**a**) Schematic depicting off-target library construction and guide selection. (**b**) Density plots showing activity of guides in dropout screens targeting essential genes with zero, one, and two mismatches. Line is displayed at the 5th percentile of guides targeting control genes(**c**) Heat map displaying the fraction of guides active for each mismatch type at a given position in the guide. Guide position is numbered such that position 1 is PAM proximal. The fraction of active guides is reported from essential guides in dropout assays, vemurafenib resistance genes in vemurafenib assays, and *HPRT1* guides in 6-thioguanine assays. (**d**) Comparison of off-target activity for Cas9, 2xNLS-Cas12a, and enCas12a. For each enzyme, the fraction of active guides at each guide position and nucleotide mismatch were ranked, and plotted in ascending order. (**e**) Density plots displaying measured activity of double mismatch guides targeting essential

genes binned by a prediction of activity using the Cutting Frequency Determination (CFD) score. CFD activity bin and number of guides in each bin is reported.
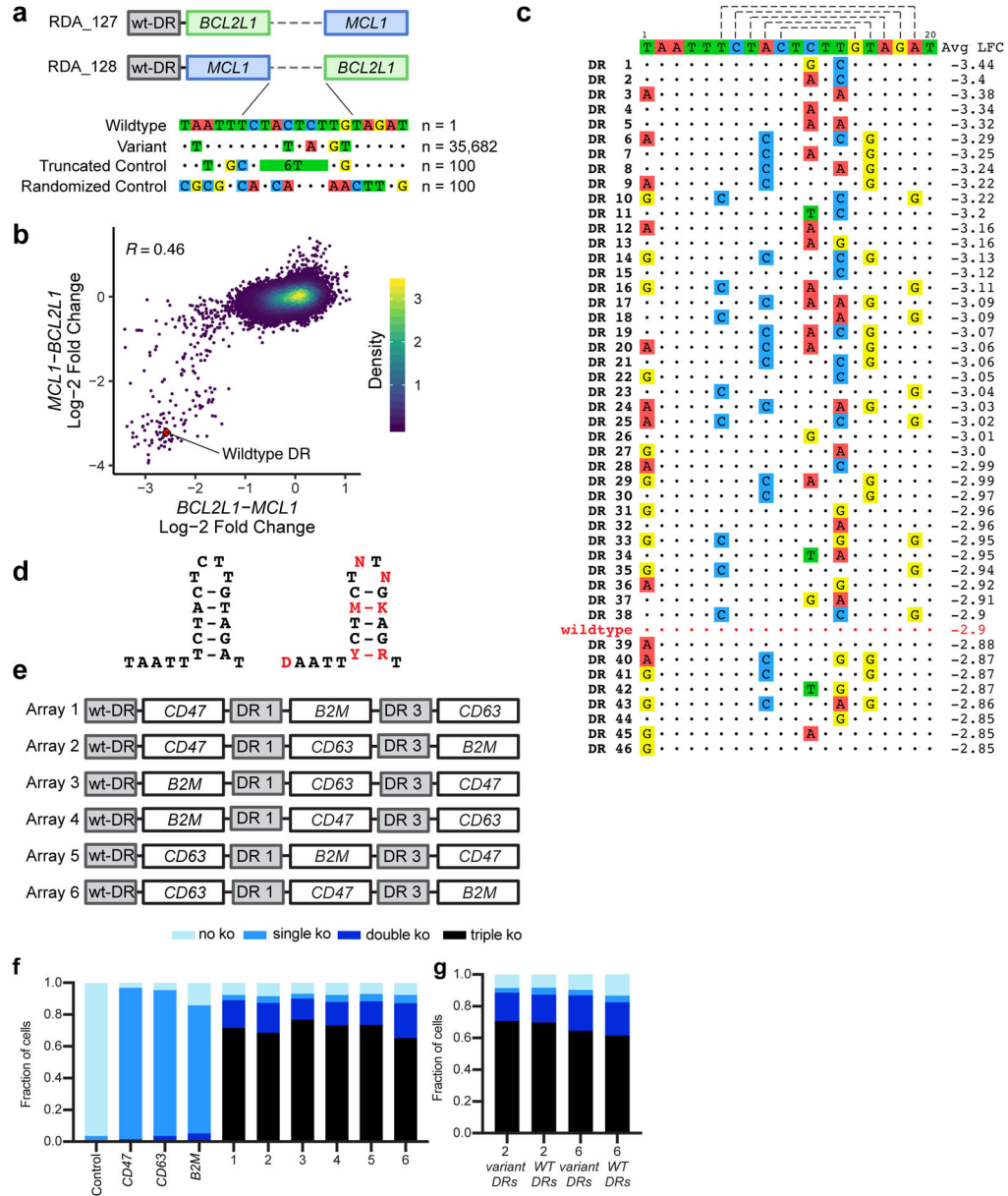
Figure 4.
Development of alternate direct repeat sequences for multiplexing with AsCas12a (**a**)
Schematic of experimental design. (**b**) Average log2-fold change for direct repeats with both
orientations of the *BCL2L1* and *MCL1* guides (n=35,883). Pearson correlation coefficient is
indicated. (**c**) Most active variant direct repeat sequences by average log2-fold change across
replicates and orientation. Nucleotide substitutions for top variants are shown below the
wildtype sequence (top line). (**d**) Wildtype direct repeat sequence (left) and a consensus
sequence for active variant direct repeats (right). Nucleotides in red denote positions that are
flexible for alternate sequences. (**e**) Schematic of 6 multiplexed arrays with guides targeting
CD47, B2M, and *CD63* used in the triple knockout experiment. (**f**) Fraction of cells with no,
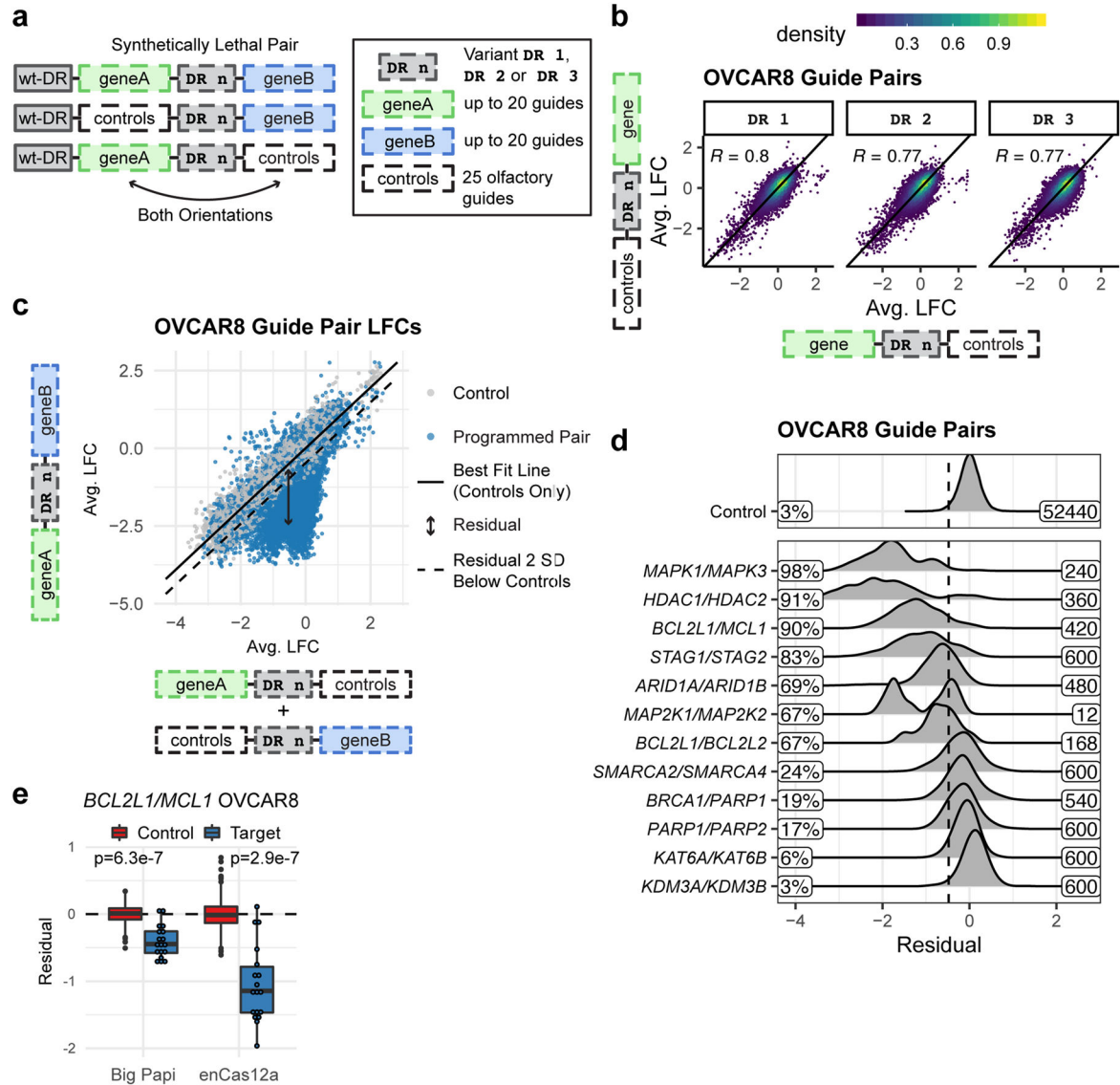one, two, or three genes knocked out, assayed by flow cytometry. (**g**) Comparison of

knockout with arrays 2 and 6 when guides are separated by variant, or all wild-type direct repeats.
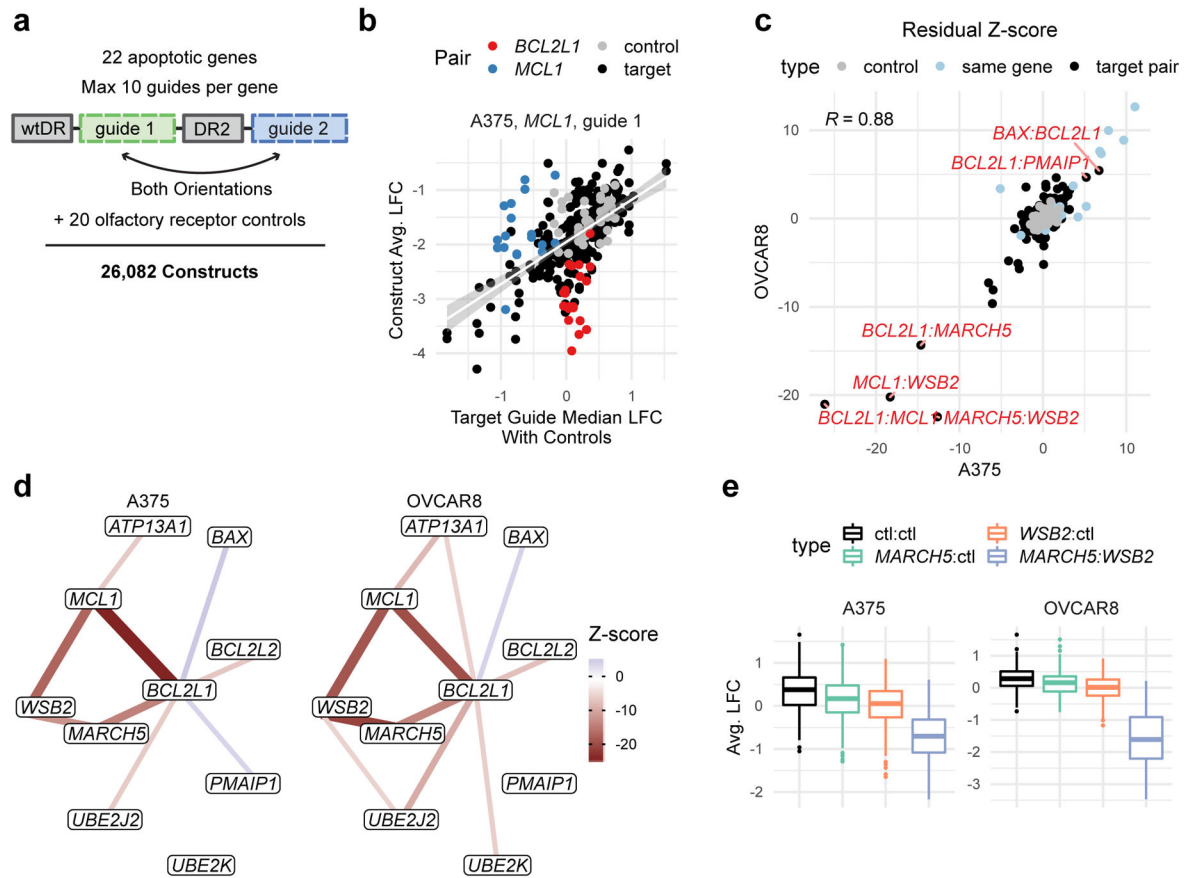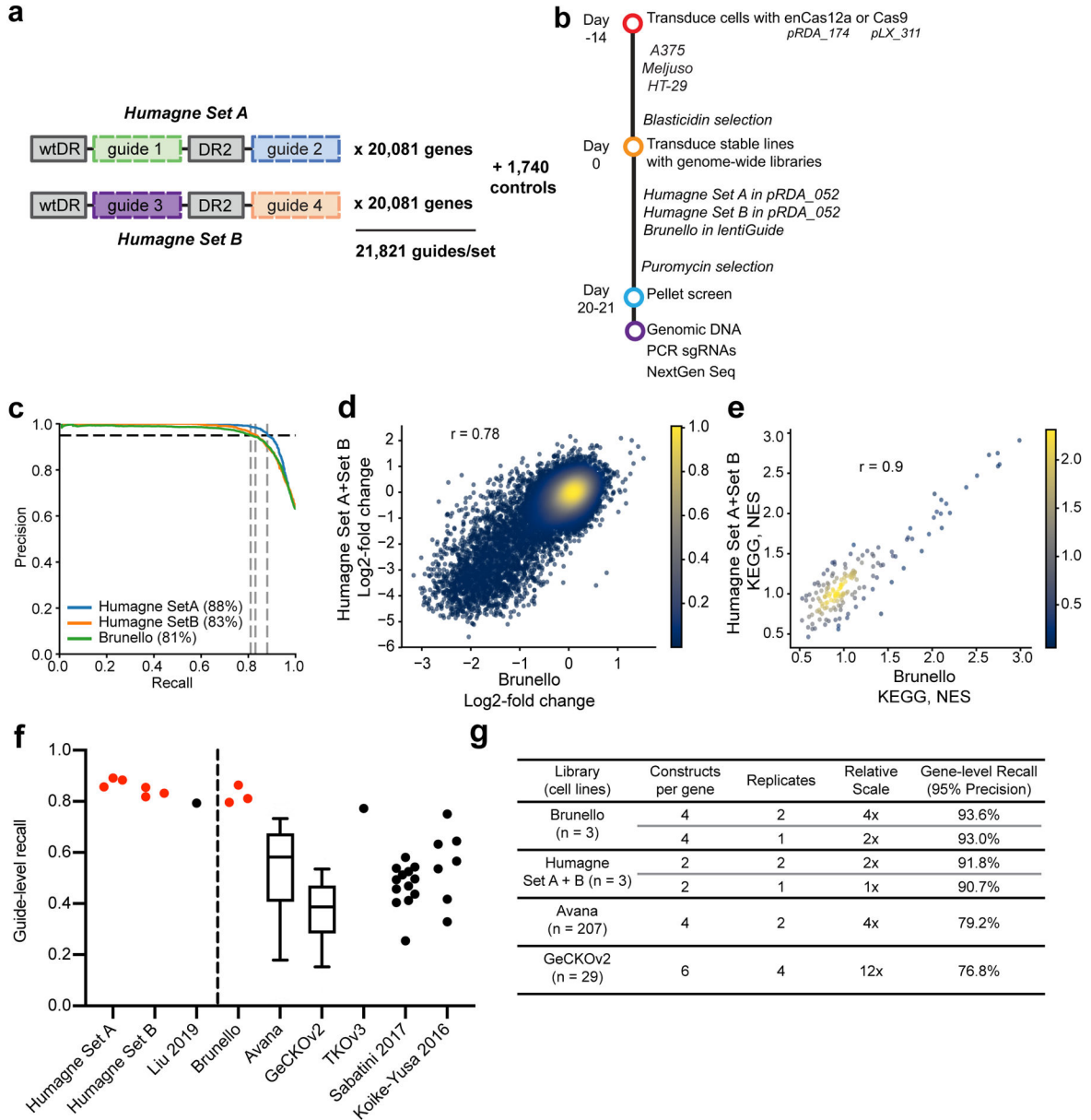
Figure 5.

Validation of AsCas12a performance with synthetic lethal gene pairs (**a**) Schematic of library design. Numbered direct repeats reference the same sequences as Figure 4. (**b**) Correlation between the average log2-fold change (LFC) of target guides in position 1 versus position 2 for all three DR variants screened in OVCAR8. Pearson correlation coefficient is indicated. n=8,533 constructs for each scatter. (**c**) Average LFC for guide pairs versus the sum of each guide paired with controls in OVCAR8. Control points represent guide pairs with one control guide and one target guide. Regression line fit with control points only. Dashed line represents a residual two standard deviations below the mean residual for control points. (**d**) Density of residuals for synthetic lethal guide pairs in OVCAR8 screening with enCas12a, filtered for the top half of guides for each gene based on the enPAM+GB score. Dashed line represents two standard deviations below the mean residual of controls. Percent of pairs with a residual to the left of the dashed line is included. Labeled on the right is the number of unique constructs in the distribution. (**e**) Comparison

of residuals for *BCL2L1/MCL1* in OVCAR8 across Cas platforms. Control constructs have one target guide (*BCL2L1* or *MCL1*) and one control guide (n=180), whereas target constructs contain a synthetic lethal guide pair (n=18). P-values were calculated using a one-sided t-test with the alternative hypothesis that the mean of the target population was less than the mean of controls. Boxes represent the 25th, 50th and 75th percentiles, whiskers show 10th and 90th percentiles.

Figure 6.
Combinatorial screen identifies a novel synthetic lethality in apoptotic genes (**a**) Schematic of library design. Numbered direct repeats reference the same sequences as in Figure 4. (**b**) Example linear fit with an *MCL1* guide in A375 cells. Each dot represents a guide paired with the *MCL1* guide that anchors the analysis (n=322). Shaded grey around the linear fit represents the standard error. (**c**) Gene pairs plotted by residual Z-score in OVCAR8 versus A375 cells. Negative scores represent synthetic lethal genes, and positive scores represent buffering genes. Pearson correlation coefficient is indicated. Select gene pairs are labeled. (**d**) Gene interaction network. Edges are drawn at an absolute Z-score of 5. Negative edges represent synthetic lethal genes, and positive edges represent buffering genes. Nodes are laid out using the stress-minimization algorithm, Kamada-Kawai. (**e**) Box plot visualization of *MARCH5 - WSB2* synthetic lethality. Boxes represent all guide pairs in a type (n=380 ctl:ctl, 400 *MARCH5*:ctl, 400 *WSB2*:ctl, and 200 *MARCH5*:*WSB2*), where 'ctl' denotes guides targeting olfactory receptor controls. Boxes represent the 25th, 50th and 75th percentiles, whiskers show 10th and 90th percentiles.

Figure 7.
Genome-wide libraries for enCas12a (**a**) Schematic of Humagne Set A and Set B. Each library contains one multiplexed array with two guides targeting the same gene. (**b**) Timeline by which genome-wide screens were executed. (**c**) Guide-level precision recall curves for Humagne Set A, Humagne Set B, and Brunello in A375 cells. (**d**) Gene-level correlation of Humagne Set A + Set B versus Brunello (n=18,952 genes). (**e**) Comparison of the GSEA normalized enrichment scores (NES) for KEGG genes sets for Humagne Set A + Set B versus Brunello (n=164 gene sets). Pearson correlation coefficient is indicated in (d) and (e). (**f**) Guide-level recall for selected Cas12a and Cas9 libraries. Points in red are A375, MELJUSO, and HT29 screens described in this work. Each dot represents a cell line. Due to sample size, GeCKOv2 (n=33) and Avana (n=340) data from the DepMap are presented as

box and whisker plots. Boxes represent the 25th, 50th and 75th percentiles, whiskers show 10th and 90th percentiles. (**g**) Recall of essential genes at 95% precision for various combinations of libraries and experimental replicates. Relative scale is calculated with reference to Humagne Set A + B screened with one replicate each.