Cell Genomics Transparent Peer Review Record

Modular vector assembly enables rapid assessment of emerging CRISPR technologies

Abby V. McGee[1], Yanjing V. Liu[1], Audrey L. Griffith[1], Zsofia M. Szegletes[1], Bronte Wen[1], Carolyn Kraus[2], Nathan W. Miller[1], Ryan J. Steger[1], Berta Escude Velasco[1], Justin A. Bosch[3], Jonathan D. Zirin[3], Raghuvir Viswanatha[3], Erik J. Sontheimer[2,4,5], Amy Goodale[1], Matthew A. Greene[1], Thomas M. Green[1], John G. Doench[1,6,*]

Summary

Initial submission: Received : Oct 25, 2023
Scientific editor: Judith Nicholson
First round of review: Number of reviewers: 2
Revision invited : Nov 14, 2023
Revision received : Dec 31, 2023

Second round of review: Number of reviewers: 2
Accepted : Feb 8, 2024

Data freely available: Yes

Code freely available: Yes

Referees' reports, first round of review

Reviewer #1: In the paper Modular Vector Assembly Enables Rapid Assessment Of Emerging CRISPR Technologies, Doench et. al introduce Fragmid- a toolkit composed of modular fragments that allows users to create customized CRISPR/Cas vectors for their research. The group has also developed a web portal that aids users in picking components for their choice application or Cas technology. The Fragmid system has been well thought out and contains numerous components which gives users many options in their design process. Importantly by offering and working to standardize CRISPR Cas components, it is hoped that vectors and new technologies can be quickly adapted and implemented into research. The articles experiments and figures mostly describe workflow and show applications of Fragmid.
Overall, this paper details a resource/ toolkit that will be of broad interest to the CRISPR/Cas9 and genome editing field, and is in line with the scope of Cell Genomics.

General comments:

1. On the Fragmid website, give the names if possible of the CBE and ABE cassettes (ie BE4max, ABE8e)
2. It is not described clearly how a researcher fully uses the Fragmid resource. A genbank file is created, but what next? How do you request the plasmid? This is the first question most readers will have.
3. Discuss whether the modules will be distributed on Addgene
4. In the online Fragmid User Manual, Module 3-4 Cas Protein Description is "Duh". Agreed, but not professional
5. Adding "cloning spacers" for N and C terminus module options (which contain a rare restriction site) would be nice to allow users to order a vector and then put in their own protein domains.
6. Adding significance values to graphs would help
7. Fragmid will be of most interest to those performing screens or for in vitro applications. We suggest emphasizing these techniques.

Specific:

Line 210: "28 targets were tested using 5 different enzyme-technologies" , not 28 gene technologies.
Line 211:" performing guides were assessed individually and in arrayed format and validated guides"
Line 222-226- run on sentence, hard to understand, should be broken up for readability.

Fig 1b: Calling the backbone lentiCRISPRv2 may lead to confusion b/c most people familiar with field associate it with "expresses Cas9"

Fig 2b: Why are there distinct populations that have pos LFC in Rep B, but neg LFC in rep A and vice versa? Strange.

Fig 3a: Add statistics to graph

Supp Fig 2: Would be neat to compare Cas9 and Cas12a CRISPRa in same experiment just to see if one works better.
I think it would help to clarify section C, "For the EnAsCas12a data, all four guides, each targeting a unique gene are expressed on the same array." In case someone thinks 4 guides targeting the same gene for each histogram

Reviewer #2: The manuscript "Modular vector assembly enables rapid assessment of emerging CRISPR technologies" by Doench et al. describes a Golden-Gate-based cloning kit that allows for the combinatorial assembly of various Cas9-based expression cassettes and associated gRNAs in different types of vectors. The authors show that this resource allows for rapid testing - even in a pooled approach - if a certain CRISPR technology (CRISPRko, CRISPRi, CRISPRa) will work in a certain non-model cell type and if yes, which Cas9 variant under which promoter will perform the best.

The created toolkit could become a very valuable resource for the genomics community but critical information is missing that disables users to get the full potential of this kit, especially the use of the positive controls for benchmarking.

Overall the manuscript is written in a very dense, jargon-rich way such that important information is lost in translation or never given. Especially the figure legends should give more experimental details (e.g. cell type), and number of replicates that describe the reported error.

1. Figure 2D: The table is unclear.
Does the second and third columns mean that only for CD47, CD55, and CD59 well-working guides could be identified and for the other genes not? Or that only those guides were further validated? Please clarify this table.

2. Missing performance metrics for positive controls.
I like the idea of providing positive controls such that future users can benchmark their protocols and libraries in an established cell line first. However critical information is missing to allow users to do this:

a. Supplementary data 1 provides the sequences of the validated positive controls, but no further performance metrics and/or associated assay protocols are given. But those are essential for future users to use these positive controls as a benchmark.
Standardized protocols, readouts, and performance metrics need to be added to Supplementary Data 1 for each guide: For example, for the CRISPRko, what is the percentage of successful KO's e.g. measured by NGS?

In Line 186 the authors state: After sample processing and sequencing, we calculated the fold change between the log-normalized read counts of guides (LFC) in the low and high expressing populations. Replicate LFCs were modestly correlated, with good enrichment of CD47-targeting guides (Figure 2b).

However, Figure 2b is very dense and doesn't highlight the chosen positive controls given in Supplementary Data 1. Further, what is "other" in the figure legends?

b. For all data depicted in Figure S2c, which version of the positive control gRNA has been tested in each panel, what is the cell type? Ideally, the fluorescence should be calibrated to an independent calibrant, and the fold change difference between +/- gRNA should be given, such that future users can use it as a reference to check if the positive control in their experiment behaves is the same way.

c. Line 182. When determining positive control guides: "We used an A375 (melanoma) cell line engineered to express SpCas9."
The authors need to specify the exact construct that were used in the initial screen as well as the constructs that were used for individual testing of the positive control guides. Those should be specified for all technologies CRISPRko, CRISPRi, and CRISPRa.

d. Figure 2C: The Schematic depicting GG-assembled CD47-targeting construct architecture is not clear. Where the gRNAs expressed from the same construct as the Cas9? Related to point 2b above, where the architecture of the Cas9 constructs is not clear.

e. Will the cloned positive control plasmids be provided via Addgene? That would be very useful for future users

3. A list of all parts included in the kit should be given that is stand-alone from the online tool.
The list should include all parts (promoters, Cas9 variants, etc.) plus their relevant metrics, such as part size and PAM sequence for each Cas9, promoter performance (high, medium, low for a given cell type), references for the repression and activation domains and where they have been shown to perform well, etc., reason for including them

4. the details of the Barcoding approach are not enough described to allows repetition. A list of the used barcodes should be given and a note should be given on how they were integrated into the GG scheme. How can users design such a library? By PCR? Figure S4a gives an idea but leaves the approach vague.

Authors' response to the first round of review

Reviewer #1

In the paper Modular Vector Assembly Enables Rapid Assessment Of Emerging CRISPR Technologies, Doench et. al introduce Fragmid- a toolkit composed of modular fragments that allows users to create customized CRISPR/Cas vectors for their research. The group has also developed a web portal that aids users in picking components for their choice application or Cas technology. The Fragmid system has been well thought out and contains

numerous components which gives users many options in their design process. Importantly by offering and working to standardize CRISPR Cas components, it is hoped that vectors and new technologies can be quickly adapted and implemented into research. The articles experiments and figures mostly describe workflow and show applications of Fragmid.

Overall, this paper details a resource/ toolkit that will be of broad interest to the CRISPR/Cas9 and genome editing field, and is in line with the scope of Cell Genomics.

General comments:
1. On the Fragmid website, give the names if possible of the CBE and ABE cassettes (ie BE4max, ABE8e)

In places where this is appropriate we have done so, although please note that nomenclature can be a bit complicated. For example, the term "BE4max" (Koblan et al., Nat. Biotech, 2018, PMID: 29813047) includes changes made to the codon usage of Cas9, not to the Apobec domain. From that paper:

Next we generated bis-bpNLS BE4 variants using eight codon usages: from IDT (used in BE4), GeneArt, Coller and co-workers, and GenScript. Every codon optimization method improved editing efficiency over IDT codon usage in HEK293T cells (Fig. 1e, Supplementary Fig. 2b). We also tested four chimeric codon-optimized BE4 variants that mixed different deaminase and Cas9 nickase codon usages (Supplementary Fig. 3a, 3b), but none outperformed the GenScript-only variant (BE4max), which induced 1.8-fold higher editing over bis-bpNLS BE4 with IDT codons (Fig. 1e).

We use the codon-optimized variant published by the Dow Lab (PMID: 29969439). Finally, note that the updated Supplementary Data 1 includes a spreadsheet that provides citations for all components, so that users can determine the primary source of the information and read further.

2. It is not described clearly how a researcher fully uses the Fragmid resource. A genbank file is created, but what next? How do you request the plasmid? This is the first question most readers will have.

Please see answer to next question.

3. Discuss whether the modules will be distributed on Addgene

Prospective users of this resource will start by ordering the kit from Addgene, which will consist of two 96 well plates of fragments, destination vectors, and positive controls. The user then cherry-picks the relevant constructs needed to assemble their desired plasmid. Addgene may, in the future, establish a service in which they assemble plasmids for users, but the economic viability of that model requires assessment on their part.

4. In the online Fragmid User Manual, Module 3-4 Cas Protein Description is "Duh". Agreed, but not professional

Yikes, we clearly did not intend for that comment to make it to the public! This 'User Manual' was written before we had the Fragmid web portal fully functional, as we beta-tested this resource with some internal collaborators, and we fully intend that this manuscript and the accompanying portal replace the information contained therein.

We have thus taken down the User Manual, and information about specific fragments is now provided both in the comments field on the web portal and in materials provided with this manuscript (Supplementary Data 1). Please note that the web portal includes a list of all available fragments for those who wish to browse the full collection, as a complement to Supplementary Data 1: https://portals.broadinstitute.org/gppx/fragmid/public/fragmentinfo

5. Adding "cloning spacers" for N and C terminus module options (which contain a rare restriction site) would be nice to allow users to order a vector and then put in their own protein domains.

No restriction site, no matter how rare, could be future-proofed against some other fragment ever containing that site, and thus we are quite hesitant to assume that anything other than BsmBI would ever be unique in an assembled plasmid (BsmBI is used for cloning in guide RNAs). Rather, for a user wishing to use a custom protein domain, we recommend synthesizing or PCR-cloning that domain into a vector that contains the appropriate BbsI sites, i.e., making a new fragment.

Note that, either way, there are two cloning steps. In the method suggested by the reviewer, one first assembles an intermediate vector via Fragmid and then clones in the desired domain via rare restriction sites to create the final vector. In our suggested approach, one first makes a new fragment vector and then performs Golden Gate cloning to make the final vector. The benefit of the latter approach is that this new fragment can be added to the collection and re-used in the future.

6. Adding significance values to graphs would help

Significance values have been added to all graphs where it is appropriate.

7. Fragmid will be of most interest to those performing screens or for in vitro applications. We suggest emphasizing these techniques.

AAV is often used in vivo, and thus we do not want to suggest this toolkit is only for in vitro applications or performing screens, although we acknowledge that we did not apply AAV to any in vivo setting here. Note that Fragmid was also used in live Drosophila. That said, the reviewer is right that the majority of the applications demonstrated here are directed towards screens conducted in cell culture models, but we do not want to suggest these are the only potential applications.

Specific:
Line 210: "28 targets were tested using 5 different enzyme-technologies" , not 28 gene technologies.

Good clarification, we have reworded this sentence: “Similar screens were then performed to identify positive controls for other technologies, with 28 unique combinations of target genes and Cas constructs assessed in this pooled format.”

Line 211:" performing guides were assessed individually and in arrayed format and validated guides"

We have rephrased this sentence to: “Top-performing guides were then individually validated (Figure 2d, Supplementary Figure 2c) and these positive control constructs were added to Fragmid (Supplementary Data 1).”

Line 222-226- run on sentence, hard to understand, should be broken up for readability.

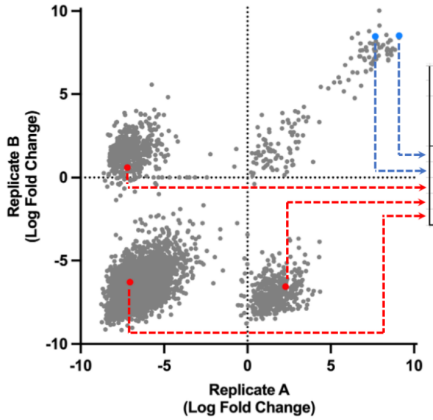
Yes, that was indeed a mouthful of a sentence, we have rephrased: The much larger cargo size led to low lentiviral titers and thus several modifications were incorporated to generate lentiCRISPRv2. Importantly, a different lentiviral backbone was employed: lentiCRISPRv1, lentiGuide, and pLKO all use the RSV promoter and SV40 polyA to express and polyadenylate the full lentiviral transcript, while lentiCRISPRv2 utilizes the CMV promoter and bGH polyA. We sought to assess the effects of these backbone differences on both lentiviral titer and activity.

Fig 1b: Calling the backbone lentiCRISPRv2 may lead to confusion b/c most people familiar with field associate it with "expresses Cas9"

This is a good point. We have modified this figure to remove this phrase and clarified the language in the text, as indicated in the response to the prior comment.

Fig 2b: Why are there distinct populations that have pos LFC in Rep B, but neg LFC in rep A and vice versa? Strange.

This is actually a fairly common occurrence in stringent positive selection screens, whether by flow cytometry as performed here or in viability screens, i.e., selecting for drug resistance. Below we plot replicate correlation – each point represents a guide – and highlight several points that show the underlying data for this phenomenon. First, the two blue points are the CD47 guides that were selected for inclusion in Fragmid. They are very abundant in the CD47 negative population in both replicates. We also show, in red, example data for control guides that score in one replicate but not the other, as well as one guide that scored in neither. Note that the blue CD47 guides are approximately 50-fold more abundant in the CD47 negative population than the false positive red control guides. If we were to sort substantially more cells, then the replicates would likely correlate better, but the top hits would not change, and in this experiment we are less interested in eliminating false positives than we were in identifying a few true positives, which clearly have more signal.



pDNA		Replicate A					Replicate B				
		CD47 pos	CD47 neg	CD47 pos	CD47 neg	neg - pos	CD47 pos	CD47 neg	CD47 pos	CD47 neg	neg - pos
Read Counts	Log2Norm per Million Reads	Read Counts		Log2Norm per Million Reads		Log Fold Change	Read Counts		Log2Norm per Million Reads		Log Fold Change
2442	7.46	189	8032	5.66	13.32	7.67	204	13894	5.82	14.30	8.48
2260	7.35	72	8523	4.31	13.41	9.10	129	9126	5.18	13.69	8.52
2105	7.24	557	0	7.20	0.00	-7.20	906	257	7.95	8.55	0.59
1902	7.10	363	367	6.59	8.87	2.29	343	0	6.56	0.00	-6.56
1025	6.22	498	0	7.04	0.00	-7.04	284	0	6.29	0.00	-6.29

Rebuttal Figure 1.

Please note that, also in response to comments from Reviewer 2, we have provided an alternative visualization of these data, provided here for convenience, which more directly illustrates the point that, as a class, the guides targeting CD47 are readily distinguishable:

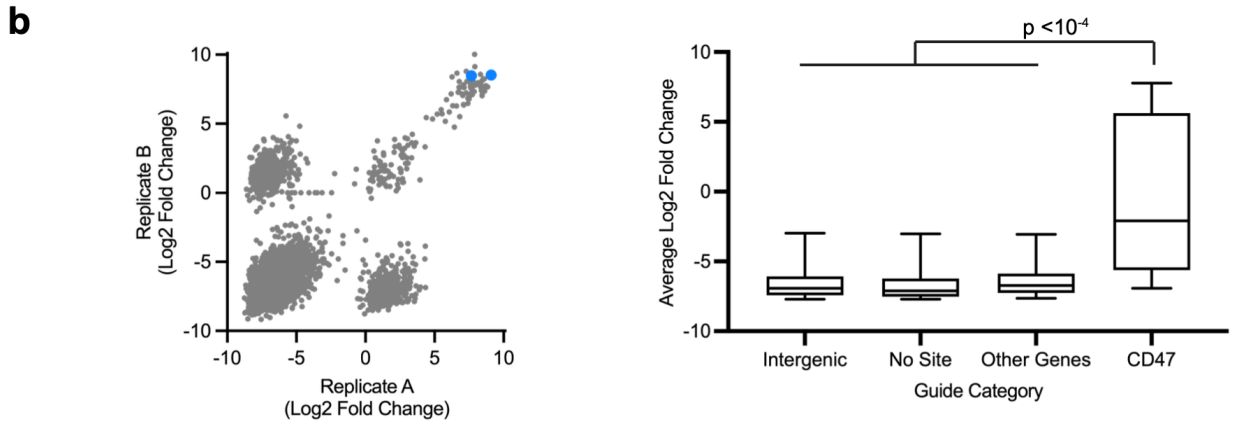


Figure 2b, updated. B) Left: LFC correlations (Pearson's R) of the CD47 flow-sorted samples for SpCas9 knockout comparing two replicates. LFCs were calculated by subtracting sorted samples from one another (CD47-negative - CD47-positive sorted population). The two guides selected for validation are indicated in blue. Right: Average LFC by guide category. Boxes indicate mean, 25th, and 75th percentiles, whiskers show 10th and 90th percentiles. Other Genes refers to all the guides in the library targeting a gene other than CD47. Two-tailed Mann-Whitney test used to calculate p-value.

Fig 3a: Add statistics to graph

Done!

Supp Fig 2: Would be neat to compare Cas9 and Cas12a CRISPRa in same experiment just to see if one works better.

We very much agree with the sentiment, but we do not believe this is in-scope for this manuscript. To make a conclusion about whether either Cas enzyme is fundamentally better for CRISPRa, one would really need to compare dozens, if not hundreds, of guides. We are indeed working on this (as a follow-up to our recent Cas12a CRISPRa paper, Griffith et al. 2023, PMID: 37719144), but this requires large-scale pooled screens to do it right, and that seems out-of-scope for this resource.

I think it would help to clarify section C, "For the EnAsCas12a data, all four guides, each targeting a unique gene are expressed on the same array." In case someone thinks 4 guides targeting the same gene for each histogram

We believe this comment is in reference to the legend of Supplementary Figure 2c. We have rephrased to "For each EnAsCas12a technology, data are shown from one array that expresses four guides, one targeting each gene."

Reviewer #2
The manuscript "Modular vector assembly enables rapid assessment of emerging CRISPR technologies" by Doench et al. describes a Golden-Gate-based cloning kit that allows for the combinatorial assembly of various Cas9-based expression cassettes and associated gRNAs in different types of vectors. The authors show that this resource allows for rapid testing - even in a pooled approach - if a certain CRISPR technology (CRISPRko, CRISPRi, CRISPRa) will work in a certain non-model cell type and if yes, which Cas9 variant under which promoter will perform the best.

The created toolkit could become a very valuable resource for the genomics community but critical information is missing that disables users to get the full potential of this kit, especially the use of the positive controls for benchmarking.

Overall the manuscript is written in a very dense, jargon-rich way such that important information is lost in translation or never given. Especially the figure legends should give more experimental details (e.g. cell type), and number of replicates that describe the reported error.

1. Figure 2D: The table is unclear. Does the second and third columns mean that only for CD47, CD55, and CD59 well-working guides could be identified and for the other genes not? Or that only those guides were further validated? Please clarify this table.

Cas enzyme	Activity	Gene	Validated Construct & Fragment ID
Cas9 [Sp]	**Knockout** A375 cells *pLX_311-Cas9*	CD47	pRDA_832, AA242 pRDA_833, AA243
		CD55	pRDA_836, AA246
		CD59	pRDA_834, AA244
	Interference A375 cells *pRDA_662*	CD47	pRDB_162, AA286
		CD55	pRDB_166, AA290
		CD81	pRDB_161, AA285
	Activation HT29 cells *pRDA_831*	CD47	pRDB_118, AA232
		CD274	pRDB_156, AA233
Cas12a [EnAs]	**Knockout** A375 cells *pRDA_174*	CD46	pRDA_936, AA324
		CD47	
		CD55	
		CD81	
	Activation HT29 cells *pRDA_816*	CD4	pRDB_131, AA435
		CD26	
		CD97	
		CD274	

To avoid confusion and provide clarity, we have streamlined this table to more directly show what users will ultimately care about, the positive control reagents identified in this study with their relevant identifiers (related to point 2e, below). We have added additional experimental detail to the table as well, per later comments by this reviewer, to make it clear what constructs and cell lines were used in the validation experiments. The updated table is shown here for convenience:

2. Missing performance metrics for positive controls. I like the idea of providing positive controls such that future users can benchmark their protocols and libraries in an established cell line first. However critical information is missing to allow users to do this:

a. Supplementary data 1 provides the sequences of the validated positive controls, but no further performance metrics and/or associated assay protocols are given. But those are essential for future users to use these positive controls as a benchmark. Standardized protocols, readouts, and performance metrics need to be added to Supplementary Data 1 for each guide: For example, for the CRISPRko, what is the percentage of successful KO's e.g. measured by NGS?

The performance of all of the positive control guides are demonstrated in Supplementary Figure 2c, illustrating the per-cell decrease or increase in protein levels as assessed by flow cytometry, and we have included more experimental details, such as in Figure 2d. Methods and the catalog numbers for antibodies are in the Key Resources Table. Finally, Supplementary Data 1 now contains maps for all the plasmids included in the Fragmid kit, which includes the sequences of the guides.

We chose not to quantitate guide activity by Next Generation Sequencing (NGS), as suggested by the reviewer, but rather by protein-level assessment, which is why we focused on proteins expressed on the cell surface. Flow cytometry is faster and less expensive than employing an entire NGS run dedicated to assessing one target site. Further, NGS can only be used to quantitate efficiency of knockout guides but would not be amenable to CRISPRa and CRISPRi.

In Line 186 the authors state: After sample processing and sequencing, we calculated the fold change between the log-normalized read counts of guides (LFC) in the low and high expressing populations. Replicate LFCs were modestly correlated, with good enrichment of CD47-targeting guides (Figure 2b).

However, Figure 2b is very dense and doesn't highlight the chosen positive controls given in Supplementary Data 1. Further, what is "other" in the figure legends?

In the figure legends, 'other' refers to guides in the library targeting other genes. For example, in the CD47 plot, 'other' includes guides that target CD55, CD59, etc. This has been clarified in the legend and we have rephrased to 'other genes'. Per the reviewer's suggestion, we highlight the performance of the guides that were ultimately validated. Additionally, to avoid the overplotting noted by the reviewer, we have also shown these data in box-and-whiskers format. The updated Figure 2b is provided here for convenience:

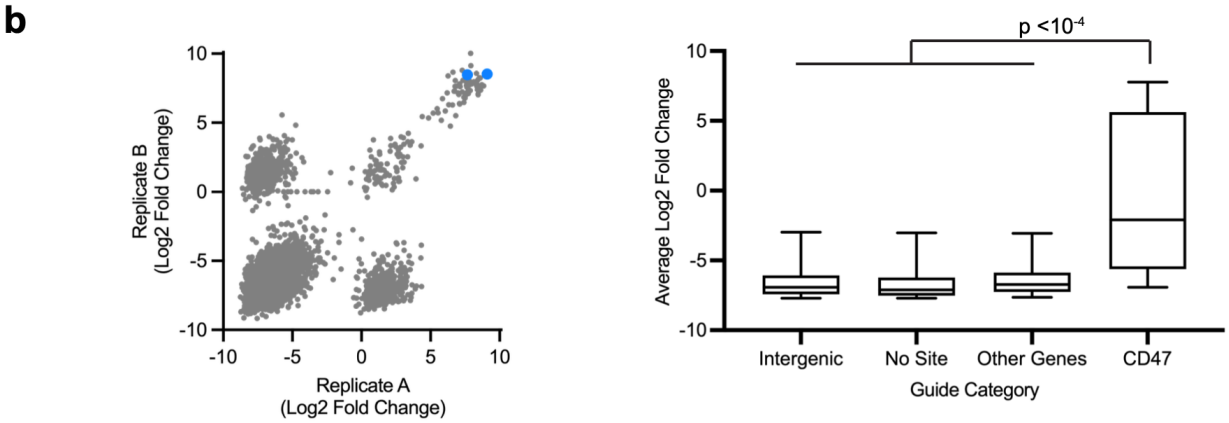
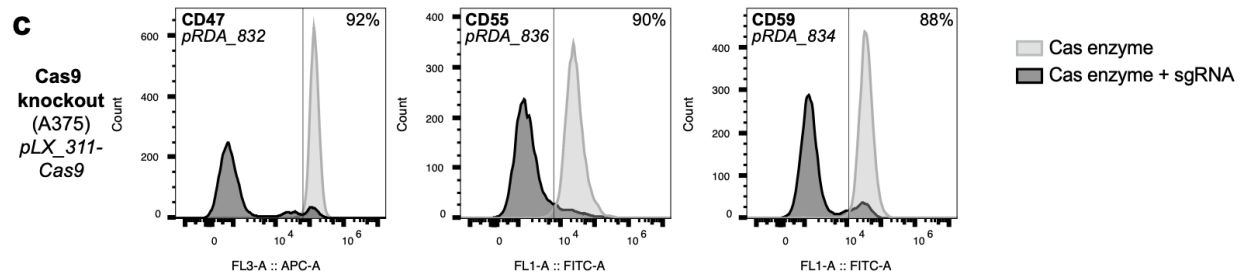


Figure 2b, updated. B) Left: LFC correlations (Pearson's R) of the CD47 flow-sorted samples for SpCas9 knockout comparing two replicates. LFCs were calculated by subtracting sorted samples from one another (CD47-negative - CD47-positive sorted population). The two guides selected for validation are indicated in blue. Right: Average LFC by guide category. Boxes indicate mean, 25th, and 75th percentiles, whiskers show 10th and 90th percentiles. Other Genes refers to all the guides in the library targeting a gene other than CD47. Two-tailed Mann-Whitney test used to calculate p-value.

b. For all data depicted in Figure S2c, which version of the positive control gRNA has been tested in each panel, what is the cell type? Ideally, the fluorescence should be calibrated to an independent calibrant, and the fold change difference between +/- gRNA should be given, such that future users can use it as a reference to check if the positive control in their experiment behaves is the same way.

The positive control guides shown in Figure S2c correspond to the versions and sequences listed in Supplementary Data 1. We have added cell line information for these data in the figure and main text. Rather than fold-change, we express performance as the percentage of modified cells - we have added percentages relative to the control population, which indicates the performance benchmark for others using these reagents. An example panel is provided here:



Updated Supplementary Figure 2c, example panel

c. Line 182. When determining positive control guides: "We used an A375 (melanoma) cell line engineered to express SpCas9." The authors need to specify the exact construct that were used in the initial screen as well as the constructs that were used for individual testing of the positive control guides. Those should be specified for all technologies CRISPRko, CRISPRi, and CRISPRa.

This information has been added to Figure 2d and Supplementary Figure 2c, and full maps are now provided in Supplementary Data 1.

d. Figure 2C: The Schematic depicting GG-assembled CD47-targeting construct architecture is not clear. Where the gRNAs expressed from the same construct as the Cas9? Related to point 2b above, where the architecture of the Cas9 constructs is not clear.

We have removed the simplified schematic and the exact constructs used are now indicated and maps provided.

e. Will the cloned positive control plasmids be provided via Addgene? That would be very useful for future users

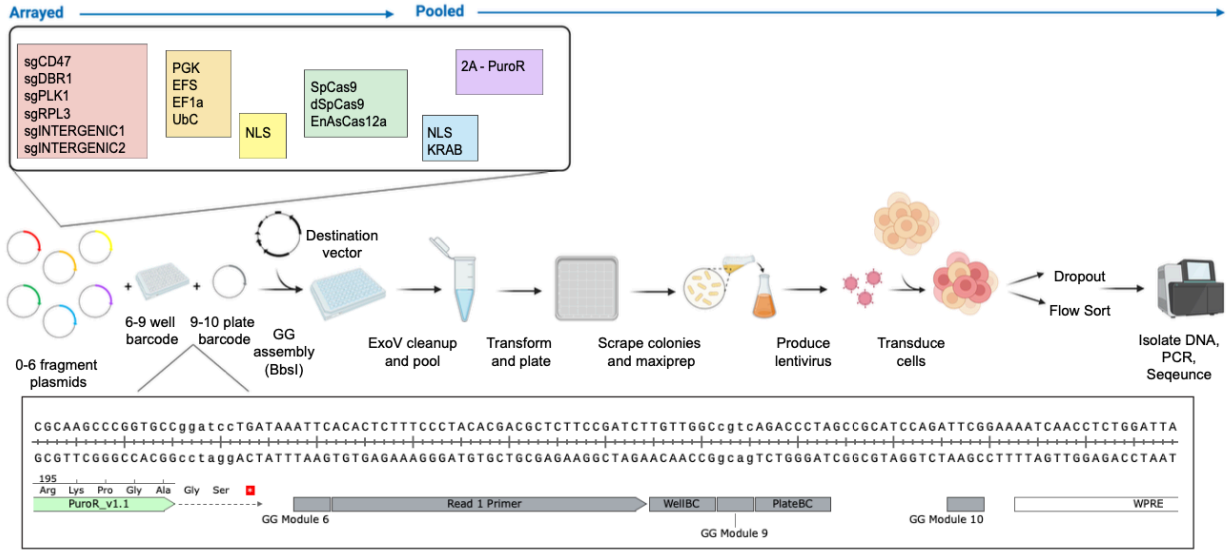
Yes, these are included in the Addgene submission.

3. A list of all parts included in the kit should be given that is stand-alone from the online tool. The list should include all parts (promoters, Cas9 variants, etc.) plus their relevant metrics, such as part size and PAM sequence for each Cas9, promoter performance (high, medium, low for a given cell type), references for the repression and activation domains and where they have been shown to perform well, etc., reason for including them

We have provided an annotated Fragmid Inventory file as part of Supplementary Data 1 that lists the different parts as well as relevant metrics and citations, as requested. We also provide Genbank maps for all constructs.

4. the details of the Barcoding approach are not enough described to allows repetition. A list of the used barcodes should be given and a note should be given on how they were integrated into the GG scheme. How can users design such a library? By PCR? Figure S4a gives an idea but leaves the approach vague.

We have updated Supplementary Figure 4a with a map of the assembled construct, which should assist in understanding the overall approach, shown here for convenience:



Supplementary Figure 4a, updated. A) Schematic of partially-pooled GG assembly, including fragments used and dual barcoding strategy. An example assembled vector is shown at the bottom, indicating the position of the Plate and Well Barcodes, as well as the Read 1 Primer for Illumina sequencing and the modules (overhangs) used for Golden Gate cloning.

We have also expanded Supplementary Dataset 3 to include the sequences of all the Plate and Well barcode sequences, as well as an example .gb file with an assembled construct. Please also note that there is a section of the Methods section dedicated to the creation of this pool, which we have elaborated on for further clarity.

Referees' reports, second round of review

Reviewer #1: All comments have been addressed to my satisfaction, and I recommend publication. This will be a valuable resource for the community and move the field forward faster.

Reviewer #2: Comments enter in this field will be shared with the author; your identity will remain anonymous.
The authors have addressed my previous comment adequately.
One comment; it might be useful for the field to assign CasID's to the available Cas proteins. This Cas nomenclature was recently suggested by leaders in the field:
https://academic.oup.com/nar/article/52/D1/D590/7331014?login=true

---

Authors' response to the second round of review

Reviewers had one remaining comment:

One comment; it might be useful for the field to assign CasID's to the available Cas proteins.
This Cas nomenclature was recently suggested by leaders in the field:
https://academic.oup.com/nar/article/52/D1/D590/7331014?login=true

We have added links to Caspedia in Supplementary Data 1.